



HAL
open science

Régulation de la maturation en 3' des pré-ARNm de gènes de susceptibilité aux cancers p53 et msh6

Adrien Decorsière

► To cite this version:

Adrien Decorsière. Régulation de la maturation en 3' des pré-ARNm de gènes de susceptibilité aux cancers p53 et msh6. Biologie moléculaire. Université Toulouse III Paul Sabatier (UT3 Paul Sabatier), 2011. Français. <NNT: >. <tel-02364811>

HAL Id: tel-02364811

<https://hal.science/tel-02364811v1>

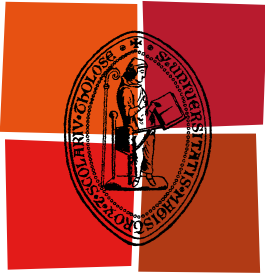
Submitted on 15 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



Université
de Toulouse

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse III Paul Sabatier (UT3 Paul Sabatier)

Discipline ou spécialité :

Biologie Moléculaire

Présentée et soutenue par :

Adrien Decorsière

le : vendredi 23 septembre 2011

Titre :

Régulation de la maturation en 3' des pré-ARNm de gènes de susceptibilité
aux cancers p53 et msh6

Ecole doctorale :

Biologie, Santé, Biotechnologies (BSB)

Unité de recherche :

INSERM 1037

Directeur(s) de Thèse :

Dr Stéphan Vagner et Dr Stefania Millevoi

Rapporteurs :

Dr Beatrice Eymin et Dr Lionel Minvielle-Sebastia

Membre(s) du jury :

Dr Martin Dutertre (Examinateur) et Pr Gilles Favre (Président du jury)

REMERCIEMENTS

Le travail présenté dans ce manuscrit a été réalisé au sein du Centre de lutte contre le cancer Claudius-Regaud, dans l'unité de recherche INSERM U563, dans l'équipe « Altérations de l'expression post-transcriptionnelle des gènes et cancers », dirigée par le Docteur Stephan Vagner.

Je tiens à adresser mes remerciements aux membres du jury :

- Au Professeur Gilles Favre, je vous remercie pour avoir accepté de présider mon jury de thèse mais également pour m'avoir accueilli au sein de votre laboratoire dans l'unité INSERM U563. J'ai ainsi bénéficié d'un environnement agréable et propice à la science.

- Au Docteur Beatrice Eymin, au Docteur Lionel Minvielle-Sebastia et au Docteur Martin Dutertre, je vous exprime toute ma gratitude pour l'intérêt que vous avez porté à ce travail et pour l'honneur que vous me faites en acceptant d'évaluer mes travaux.

Un grand merci au Docteur Stefania Millevoi. Mon encadrement n'a pas dû être simple. Tu as su me faire évoluer, en me poussant dans les moments de relâchement. Tu m'as fait partager ton expérience de la paillasse, tu m'as guidé dans mes réflexions et tu m'as appris la rigueur scientifique tout au long de ces quatre années. Sans cet entêtement à me faire progresser, je n'aurais sans doute pas pu réaliser ce travail. Grazie de tutto cuore !

Stéphan Vagner, je te remercie très sincèrement pour ta disponibilité. Ce travail n'aurait pu avoir lieu sans tes connaissances approfondies dans le domaine de la maturation en 3'. Tu m'as apporté ton aide scientifique tout en m'apprenant une qualité essentielle pour la suite : l'autonomie. Merci de ta confiance.

Merci à Christine Toulas pour la confiance accordée pour l'étude du gène MSH6 mais également pour l'étude de la séquence de p53. Merci aussi pour leur soutien technique à Pauline Escudie et Dominique Telly ainsi qu'au service d'oncogénétique.

J'adresse aussi mes remerciements à tous ceux qui ont participé directement à ces travaux de thèse. Notamment Anne Cayrel, Sandra Bernat et Vanessa Ligonnet dont le soutien technique fut une aide précieuse. Olivier Sordet Anne-Françoise Tilkin et Anne Pradine, pour leurs conseils scientifiques et techniques. Vos conseils m'ont permis d'avancer lorsque j'étais coincé...

REMERCIEMENTS

Un grand merci à mes collègues et amis:

- Piou (Pierre Zindy) le marathonien du dimanche qui n'arrive pas à exploiter les chinois.
- Gégé (Gérard Maillot) le Ge qu'est papa, vive les poulies électriques.
- Mamie (Julia Cherier) pour ces éclats de rire très discrets et ces expressions « top moumoutte ».
- Popo (Sophie Poles) 3 pommes qui mangent des cartes de crédits. Attention il y a un chauve derrière toi.
- Boulet (Emilie Bousquet) la chauve souris, j'espère que tu digères bien le chorizo.
- le blond (Nicolas Bery) arrête de draguer les serveuses, sans toi ces années n'auraient pas eu le même parfum...
- le beau gosse (Rémi Gence) j'aurais aimé voir Audrey un peu plus souvent, Rémi sans famille !
- Val (Valérie Schmitt), arrête de te cogner la tête et à bientôt en Suisse.
- Audrey (Audrey Delmas). Tu parles vraiment trop fort mais t'assures grave le Y.
- Raphaël Wapiti forever
- Julien maintenant que tu vas être papa fini de rentrer en scred les bouteilles en boîte.
- Florian, Slave, pense à repayer un balaie au commissariat de Castres.
- El Hadj ça y est les gars je suis plus étudiant.
- Pierre, il est marron.
- Amélie bourito oui joli cœur.
- Fauret !, Fauret !, Fauret !, Fauret !, Fauret ! Fauret ! Fauret ! Ahhhhh... Je t'aime Julia, mais on ne parle pas la bouche pleine.
- Antonin, je t'achèterai le prochain album de Renaud. Et au fait tu préfères qui, ta femme ou ton gosse ???

Que de moments partagés, de fous rires, de soirées arrosées, de discussions sans fin ni tête... Merci pour nos week-ends, ski et randonnées. Ce fut un réel plaisir de passer ces années avec vous. Vous êtes de véritables amis et j'espère vous voir bientôt à Genève!!

Je remercie tous les membres du laboratoire de l'U563 .Loubna Mhamdi, tu siffles vraiment trop bien... Enfin juste les Compères. Magali Lacroix-Triki « l'âne à patte », j'ai deux, trois TMI à faire. Anne Cammas, au prochain congrès je te garde une place devant. Samantha Beck « quand y a des nuages il fait plus froid », Margaux Roques (bouyaka bouyaka), Isabelle Lajoie-Mazenc, Anthony Lemarie, Stéphanie Cabantous, Jean-Charles Faye, Aurélien Olichon, Anne Casanova, Patrick Chinestra, Nicolas Delcourt, Kenza Mamouni, Claire Marty, Claire Médale, Julien Mazière, Nicolas Meye, Alexis Peyret, Christine Pich, Iotefa Teiti, Elise Meunier, Olivia Lanvin, Judith Martinez Gala, Cyril Inard, Sandrine Mazoyer, Mickael Paillasse, Gregory Segala, Philippe Demedina, Sylvie Montferran, Nelly, Delphine, Guillaume Sarrabayrousse, Marine Goffinet, Bruno Canguilhem, Nicolas Skuli, je vous remercie tous pour les services rendus, les conseils scientifiques et pour m'avoir apporté un environnement stimulant et joyeux.

REMERCIEMENTS

Merci aussi à Marie-Ange, Jeanine Léon, Lourdès et Joce qui m'ont permis de travailler dans un cadre agréable.

Un immense merci à ma famille sans qui rien n'aurait été possible.
A ma mère et à ma soeur Cécile (la mongoloïde), merci pour votre soutien sans faille, vos encouragements toujours efficaces, vos oreilles attentives et vos conseils.
A mon père pour ces discussions scientifiques parfois houleuses mais constructives.
A Nonna, Jack, René, Vero, Titi, Olivier et les bouts de choux Eva, Elouen et Allan, ainsi que tout les membres de ma famille, votre présence est réconfortante et m'est indispensable. Merci.

Enfin, je tiens à dédier « **notre** » thèse à **Sandra** !
Que te dire sans être cucu, les mots me manquent. Je sais que ce ne fût pas toujours très simple de me supporter, tant au boulot qu'à la maison. Merci de m'avoir soutenu et accompagné dans cette aventure. Tu m'apportes ce dont je rêve depuis longtemps, du fond de mon coeur, merci.

SOMMAIRE

SOMMAIRE.....	2
TABLE DES FIGURES	5
TABLE DES TABLEAUX.....	5
ABREVIATIONS	7
INTRODUCTION.....	11
I Généralités sur la maturation en 3'	12
II Mécanisme de la maturation en 3'	16
II-1 La maturation en 3' des ARNm contenant une queue polyadénylée.....	16
II-1-1 Les éléments de séquence permettant la polyadénylation	16
II-1-1-1 Le signal de polyadénylation.....	16
II-1-1-2 Le « Downstream sequence element »	17
II-1-1-3 Le site de clivage.....	18
II-1-2 La machinerie de maturation en 3'	18
II-1-3 Mode d'action de la maturation en 3'	20
II-2 Maturation en 3' des pré-ARNm d'histones	24
III Interaction de la maturation en 3' avec les étapes de l'expression des gènes ..	26
III-1 Interaction entre la maturation en 3' et la transcription	26
III-2 Interaction entre la maturation en 3' et la coiffe	28
III-3 Interaction entre la maturation en 3' et l'épissage	29
III-4 Interaction entre la maturation en 3' et l'export nucléaire des ARNm.....	31
III-5 Stabilisation de l'ARNm par la queue poly(A).....	31
III-6 Interaction entre la queue polyadénylée et la traduction.....	32

IV Régulation de la maturation en 3'	34
IV-1 Eléments de Séquences régulateurs de la maturation en 3'	34
IV-1-1 Le « upstream sequence element »	34
IV-1-2 Le « Downstream sequence element auxiliaire ».....	35
IV-1-3 Autres éléments de séquences.....	37
IV-2 Les facteurs de régulation de la maturation en 3'	38
IV-3 Mécanismes de régulation de la maturation en 3'	42
IV-4 Exemples de régulations physiologiques de la maturation en 3'	45
IV-4-1 La différenciation cellulaire.....	45
IV-4-1-1 Exemple de l'immunoglobuline M.....	45
IV-4-1-2 Exemple de la calcitonine.....	47
IV-4-3 Les dommages à l'ADN.....	50
V Dérégulation de la maturation en 3' associée à des pathologies	53
V-1 Les maladies sanguines	53
V-1-1 Les thalassémies	53
V-1-2 Les thrombophilies.....	54
V-1-3 L'hyperéosinophilie idiopathique.....	56
V 2 Les infections virales	57
V-2-1 Le virus de l'immunodéficience humaine de type 1	57
V-2-2 L'Herpès simplex virus 1	57
V-2-3 L'influenza A	58
V 3 Les cancers	59
V-3-1 Polymorphisme de la N-acetyltransferase 1	59
V-3-2 Activation de la PAP	59
V-3-3 Raccourcissement de la taille des 3'UTR.....	60
RESULTATS	63
I Rôle essentiel de l'interaction entre hnRNP H/F et un G-quadruplexe pour maintenir la maturation en 3' du pré-ARNm de p53 lors de dommages à l'ADN	64
I-1 Présentation et objectifs de l'étude	64
I-2 Contexte de l'étude	67
I 2-1 La famille des hnRNP H/F	67
I 2-2 Les G-quadruplexes.....	69
I-3 Stratégie expérimentale	72

I-3-1 Etude de l'effet des dommages à l'ADN sur la maturation en 3' du pré-ARNm de p53	72
I-3-2 Caractérisation des séquences régulant la maturation en 3' du pré-ARNm de p53.....	72
I-3-3 Etude de la structuration en G-quadruplexe des séquences localisées en aval du site de clivage du pré-ARNm de p53	73
I-3-4 Rôle du G-quadruplexe de p53 sur la résistance à l'inhibition de la maturation en 3' du pré-ARNm lors de dommages à l'ADN	74
I-3-5 Etude de l'association des protéines hnRNP H/F avec le G-quadruplexe de p53.....	74
I-3-6 Rôle des protéines hnRNP H/F dans l'accumulation et la fonction de p53 lors de dommages à l'ADN.....	75
I-4 Article I	76
I-5 Discussions et perspectives	101
I-5-1 Implication des hnRNP H/F dans l'apoptose.....	101
I-5-2 Implication des hnRNP H/F dans la résistance à l'inhibition de la maturation en 3' lors de dommages à l'ADN	102
I-5-3 Les G-quadruplexes : cible thérapeutique.....	102
I-5-4 Généralisation du mécanisme de résistance à l'inhibition de la maturation en 3' lors de dommages à l'ADN	104
II Une duplication de 20 nucléotides retrouvée dans des familles atteintes du syndrome de Lynch diminue l'efficacité de maturation en 3' du pré-ARNm de MSH6	106
II-1 Présentation et objectifs de l'étude	106
II-2 Stratégie expérimentale.....	108
II-2-1 Identification de mutations dans la séquence correspondant au 3' UTR de MSH6.....	108
II-2-2 Etude de l'implication de la duplication dans le 3'UTR sur l'expression de MSH6.....	108
II-2-3 Etude du mécanisme impliqué dans la régulation de l'expression de MSH6.....	109
II-3 Article II :	110
II-4 Discussions perspectives.....	132
DISCUSSION ET CONCLUSIONS	134
La maturation en 3' des pré-ARNm : Importance dans le contrôle de l'expression des gènes et implications dans les pathologies	136
REFERENCES.....	140
ANNEXES	165

TABLE DES FIGURES

Figure 1: Les différentes étapes de l'expression des gènes.....	13
Figure 2 : Maturation en 3' alternative	15
Figure 3 : Machinerie de la maturation en 3' des ARNm contenant une queue polyadénylée	21
Figure 4 : Modèle d'allongement de la queue poly(A).....	23
Figure 5 : Machinerie de la maturation en 3' des pré-ARNm d'histones.....	25
Figure 6 : Couplage entre la maturation en 3' et l'épissage	30
Figure 7 : Un réseau complexe d'interactions entre les différentes étapes de l'expression des gène.....	33
Figure 8 : Représentation schématique d'un G-quadruplexe	35
Figure 9 : Schéma représentant les différents signaux de polyadénylation humaine.....	37
Figure 10 : Polyadénylation alternative du pré-ARNm des chaînes lourdes de l'IgM.....	45
Figure 11 : Polyadénylation alternative du pré-ARNm de la CT/CGRP	47
Figure 12 : Régulation de la maturation en 3' de l'exon 4 du pré-ARNm de CT/CGRP en fonction du type cellulaire.....	49
Figure 13 : Inhibition de la maturation en 3' lors de dommages à l'ADN	51
Figure 14 : Exemples de G-quadruplexe intra et intermoléculaire.....	69

TABLE DES TABLEAUX

Tableau 1: Efficacité de maturation en 3' associée aux différents signaux de poly(A).....	17
Tableau 2: Facteurs de maturation en 3' et fonctions associées	19
Tableau 3: Protéines impliquées dans la régulation de la maturation en 3' chez les mammifères	38
Tableau 4: Exemples de pathologies associées au polymorphisme dans les séquences entourant le site de polyadénylation susceptibles d'inhiber la maturation en 3'	136

ABREVIATIONS

ADL3 :	Adénovirus L3
ADN :	Acide DesoxyriboNucléique
ARE-BP :	AU-rich Element sequence Binding Protein
ARN :	Acide RiboNucléique
ARNi :	ARN interférant
ARNm :	Acide RiboNucléique messenger
ATM :	Ataxia Telangiectasia Mutated
ATP :	Adenosine tri-phosphate
ATR :	Ataxia Telangiectasia Related
BARD1 :	BRCA1-associated RING domain protein 1
BCL2L1 :	B-Cell Leukemia/Lymphoma
BPV :	Bovin PapillomaVirus
BRCA1 :	BReast CAncer 1
BRCT ;	BRCA1 C-terminal repeats
CBC :	Cap Binding Complex
CBP20 :	Cap Binding Protein 20
CBP80 :	Cap Binding Protein 80
Cdc :	Cell division cycle
CDK9 :	Cyclin-Dependent Kinase 9
CFIm :	Cleavage Factor I
CFIIm :	Cleavage Factor II
CHK2 :	CHecKpoint 2
Clp1 :	Cleavage and polyadenylation I subunit
CSR1 :	Candida Suppressor of Rok1
CstF :	Cleavage stimulation Factor
CT/CGRP :	Calcitonin/Calcitonin Gene-Related Peptide
CTD Pol II :	Carboxy Terminale Domain of RNA polymerase II
DNA-PK :	DNA-dependent Protein Kinase
DSE :	Downstream Sequence Element
DSE aux :	Downstream Sequence Element auxiliaire
eIF :	eukaryotic Initiation Factor
FG γ -H2 :	FibrinoGen gamma Haplotype 2)
FGF2 :	Fibroblast Growth Factor 2
ELAV :	Embryonic Lethal Abnormal Vision
ELL2 :	Elongation factor, RNA polymerase II, 2
EST :	Expressed Sequence Tag
F2 :	coagulation factor II
FBP :	FUSE binding protein
FMRP :	Fragile X Mental Retardation Proteine
FUSE :	Far UpStream Element
GAPDH :	GlycerAldehyde 3-Phosphate DeHydrogenase
Gfap :	Glial Fibrillary Acidic Protein

ABBREVIATIONS

HCT :	Human Colorectal
HDE :	Histone Downstream Element)
HeLa :	Henrietta Lacks
HIV-1 :	Human Immunodeficiency Virus de type 1
hnRNA :	heterogeneous nuclear RNAs
hnRNP :	heterogeneous nuclear RiboNucléoProtéique
HRAS :	Harvey Rat sarcoma Viral oncogene homolog
HSF :	heat shock factor
HSV-1 :	Herpès Simplex Virus 1
Ig :	Immunoglobuline
IGF2BP1 :	Insulin-like growth factor 2 mRNA binding protein 1
INTS6 :	Integrator complex subunit 6
IRBIT :	IP(3)R binding protein released with inositol 1,4,5-trisphosphate
IRES :	Internal Ribosomal Entry Site
ISE :	Intron Splice Enhancers)
MAPK :	Mitogen-Activated Protein Kinase
MC1R :	Melanocortine 1 recepteur
MCF-7 :	Michigan Cancer Foundation - 7
MDM2 :	Murine Double Minute 2
MLH1 :	MutL Homolog 1
MMR :	MisMatch Repair
mRBP :	mRNA Binding Protein
MSH :	MutS Homolog
MST2 :	Mammalian STE20-like kinase 2
NAT1 :	N-AcetylTransferase 1
ND :	Non Déterminé
NF- κ B :	Nuclear Factor-kappa B
NLS :	Nuclear Localization Signal
NMD :	Nonsense-Mediated Decay
NS1 :	Non Structurale 1
ORF :	Open Reading Frame
PABP :	Poly(A) Binding Protein
PAP :	Poly(A) Polymérase
PARP :	Poly ADP Ribose Polymerase
Pcf11 :	Polyadenylation and cleavage factor subunit
PCR :	Polymerase Chain Reaction
PDGF-A :	Platelet-derived growth factor subunit A
PDGFR α :	Platelet-Derived Growth Factor Receptor, alpha
PIP2 α :	Phosphatidylinositol-5-Phosphate 4-Kinase, type II, alpha
Pol II :	ARN polymérase II.
PP1 :	Pyrophosphate
PRD :	Proline Rich Domain
PSF :	PTB-associated splicing factor
PSM2 :	Postmeiotic Segregation increased 2
PTB :	Poly-pyrimidine Tract Binding protein
RGG :	domaines riches en arginines et glycines,
RING :	Really Interesting New Gene
RNase :	RiboNucléase
RNGTT :	RNA Guanylyltransferase and 5'-phosphatase
RNMT :	RNA (guanine-7-) methyltransferase

ABREVIATIONS

RPL26 :	Ribosomal Protein L26
RREB-1 :	Ras Responsive Element Binding protein-1
RRM :	RNA Recognition Motif
RS :	arginine serine
RT :	Reverse Transcription
SAGE :	Serial Analysis of Gene Expression)
SHMT1 :	Serine Hydroxymethyltransferase 1
SLBP :	Stem Loop Binding Protein)
SnRNP :	Small nuclear RiboNuceoProtein
SRC :	SaRComa
SRp :	arginine serine protein
SV40 L :	Simian Virus 40 Large
Tat :	Trans-Activator of Transcription
TBP :	TATA box Binding Protein
TERT :	TelomErase Reverse Transcriptase
TFIIB :	Transcription factor II B
TFIIH :	Transcription Factor II H
TIAR :	T-cell-restricted Intracellular Antigen 1 Related protein
TMPyP4 :	porphyrin 5,10,15,20 Tetra(N-Methyl-4-Pyridyl)Porphin
U2AF :	U2 SnRNP Auxiliary Factor
USE :	Upstream Sequence Element
UTR :	UnTranslated Region
UV :	UltraViolet
VEGF :	Vascular endothelial growth factor
WDR33 :	WD repeat domain 33
ZFP100 :	Zinc Finger Protéin 100

INTRODUCTION

I Généralités sur la maturation en 3'

Chez les eucaryotes supérieurs, l'expression d'un gène débute par la transcription par l'ARN polymérase II d'un ARN immature appelé pré-ARN messager (pré-ARNm) ou, historiquement, hnRNA (« heterogeneous nuclear RNAs »), du fait de leur taille hétérogène et de leur localisation nucléaire. Ce pré-ARNm est mûri dans le noyau en 3 étapes co-transcriptionnelles. L'ARN néo-transcrit est déphosphorylé dans sa partie 5' par une triphosphatase, une guanine peut alors être transférée sur l'extrémité 5' du pré-ARNm. Cette guanine est méthylée par la guanine-7-méthyltransférase pour former la structure m⁷G(5')ppp(5')N (N représentant un des quatre nucléotides), appelée coiffe (Shatkin, 1976). Le pré-ARNm est alors épissé, ce qui permet l'excision des introns grâce à l'assemblage d'un complexe ribonucléoprotéique appelé « spliceosome » (Newman and Nagai, 2010). La maturation de l'extrémité 3' (aussi appelée polyadénylation) comprend le clivage dans la partie 3' terminale du pré-ARNm et la polyadénylation proprement dite (addition d'une queue poly(A) contenant 150 à 300 résidus d'Adénosines). L'ARNm mature est alors exporté vers le cytoplasme où il est traduit.

Le contrôle dynamique de l'expression des gènes permet de générer une diversité protéique et fonctionnelle nécessaire à la survie des organismes et à leur adaptation. Jusqu'aux années 1980, l'idée prévalait que la majorité des régulations de l'expression des gènes se faisait à l'étape transcriptionnelle. Plus récemment, d'autres niveaux de régulation ont été mis en évidence au cours d'évènements post-transcriptionnels. En effet, la maturation, l'export nucléaire, la localisation, la stabilité et la traduction des ARNm, sont des étapes post-transcriptionnelles pouvant réguler l'expression des protéines au niveau quantitatif et qualitatif. Une des étapes clé de cette régulation post-transcriptionnelle est la maturation en 3' des pré-ARNm.

La polyadénylation s'effectue principalement dans le noyau, bien que dans certains types cellulaires, une polyadénylation cytoplasmique ait été mise en évidence. Cette polyadénylation cytoplasmique permet d'allonger la queue poly(A)

d'environ 80-150 nucléotides à partir d'ARNm « raccourcis » (moins de 20 résidus d'Adénosines), réactivant la traduction de ce messenger dans des conditions spécifiques. Ceci se produit en particulier dans le développement des cellules germinales, au cours de l'embryogenèse précoce, et dans les sites post-synaptiques des cellules nerveuses lors de l'apprentissage (Bramham and Wells, 2007; Colegrove-Otero et al., 2005; Kimble and Crittenden, 2007; Radford et al., 2008; Villalba et al., 2011).

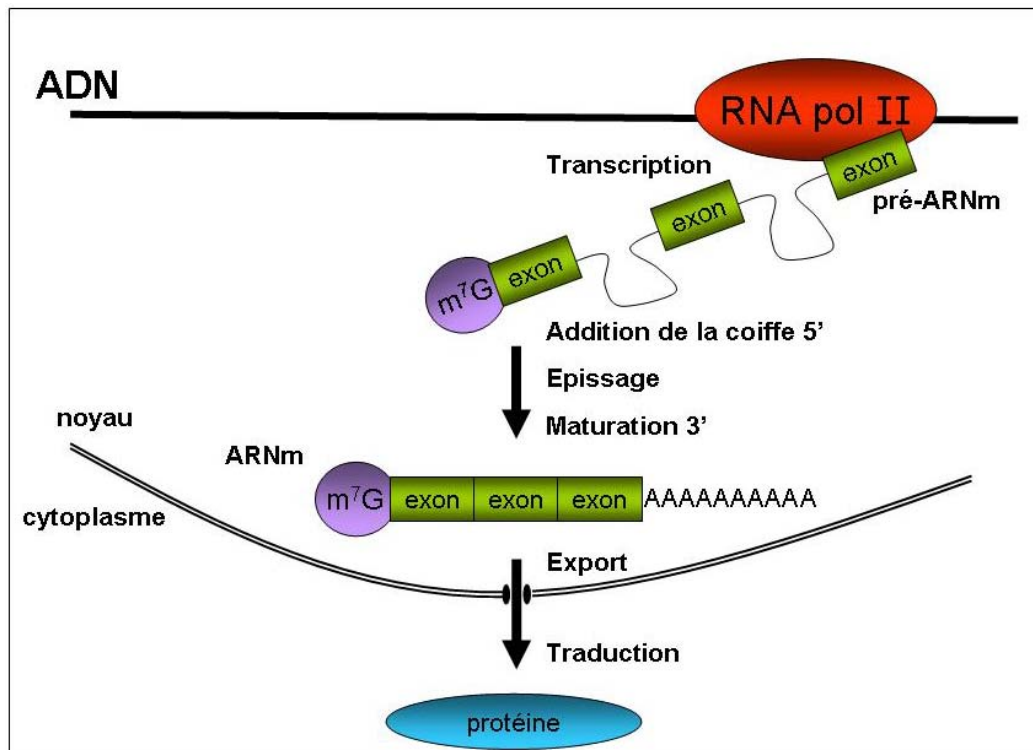


Figure 1: Les différentes étapes de l'expression des gènes.

Représentation schématique de l'expression des gènes. Le pré-ARNm est transcrit par l'ARN polymérase II, il est mûré (incluant l'ajout de la coiffe, l'épissage est la maturation en 3') dans le noyau pour former un ARNm mûré qui peut alors être exporté dans le cytoplasme où il est traduit en protéine. Les introns du pré-ARNm sont représentés en trait fin ; m⁷G correspond à la coiffe ; ARN Pol II correspond à l'ARN polymérase II.

La maturation en 3' est une étape cruciale chez les eucaryotes et l'inhibition de cette étape a des effets dramatiques sur la viabilité des cellules, leur croissance ou leur développement (Takagaki and Manley, 1998; Zhao and Manley, 1996). En effet, la maturation en 3' est essentielle pour la terminaison de la transcription, elle stimule l'épissage et la queue poly(A) permet l'export de l'ARNm, sa stabilisation et

INTRODUCTION

augmente sa traduction. De plus, le choix alternatif du site de la maturation en 3' permet d'accroître la diversité du transcriptome et du protéome.

Selon des prédictions *in silico*, la polyadénylation alternative concerne chez les mammifères plus de 50% des gènes (Tian et al., 2005; Yan and Marr, 2005). Il existe trois types de polyadénylation alternative qui sont définies en fonction de l'organisation des sites d'épissage et des signaux de maturation en 3' (Yan and Marr, 2005) (figure 2) :

- La première catégorie de maturation en 3' alternative correspond à la présence de plusieurs signaux de polyadénylation disposés en tandem sur le dernier exon en 3'. Ce type de polyadénylation alternative génère des ARNm avec des 3'UTR de différentes tailles entraînant différentes régulations post-transcriptionnelles (transport, stabilité, localisation, traduction...).
- La seconde classe correspond à des exons 3' terminaux alternatifs mutuellement exclusifs qui sont en compétition pour leur utilisation dans un ARNm. Ce type de polyadénylation alternative est couplé à un épissage alternatif et génère des ARNm contenant des 3'UTR différents et codant des protéines qui diffèrent dans leur région C-terminale. Leur sélection peut être régulée par la reconnaissance des sites 3' d'épissage et/ou des signaux de maturation en 3'.
- La troisième catégorie est également couplée à un épissage alternatif entraînant des signaux de maturation en 3' mutuellement exclusifs, mais un des exons possède un site 5' d'épissage interne. Cet exon appelé exon composite interne 3' terminal peut être, comme son nom l'indique, utilisé en tant qu'exon interne quand son site 5' d'épissage est reconnu, ou comme exon terminal quand son signal de polyadénylation est sélectionné. La régulation de l'utilisation de ces exons fait donc intervenir une compétition entre le site 5' d'épissage interne et le signal de polyadénylation.

La force des signaux de polyadénylation joue un rôle important dans la production de l'extrémité 3' des ARNm. En effet, la sélection de l'exon le plus proximal va conduire au clivage du pré-ARNm et donc prévenir l'utilisation des exons situés en aval. Il a été montré par analyses bioinformatiques d'EST (Tian et al., 2005), confirmées expérimentalement sur des signaux de polyadénylation disposés en tandem (Moucadel et al., 2007), que les signaux de polyadénylation distaux sont plus

INTRODUCTION

forts que les signaux les plus proximaux. Cette organisation implique que l'utilisation des exons 3' terminaux alternatifs soit contrôlée à la fois par les signaux d'épissage et de polyadénylation.

Le choix du site de polyadénylation se fait en fonction des conditions cellulaires telles que la croissance, la position du cycle, la différenciation, etc (Ji et al., 2009).

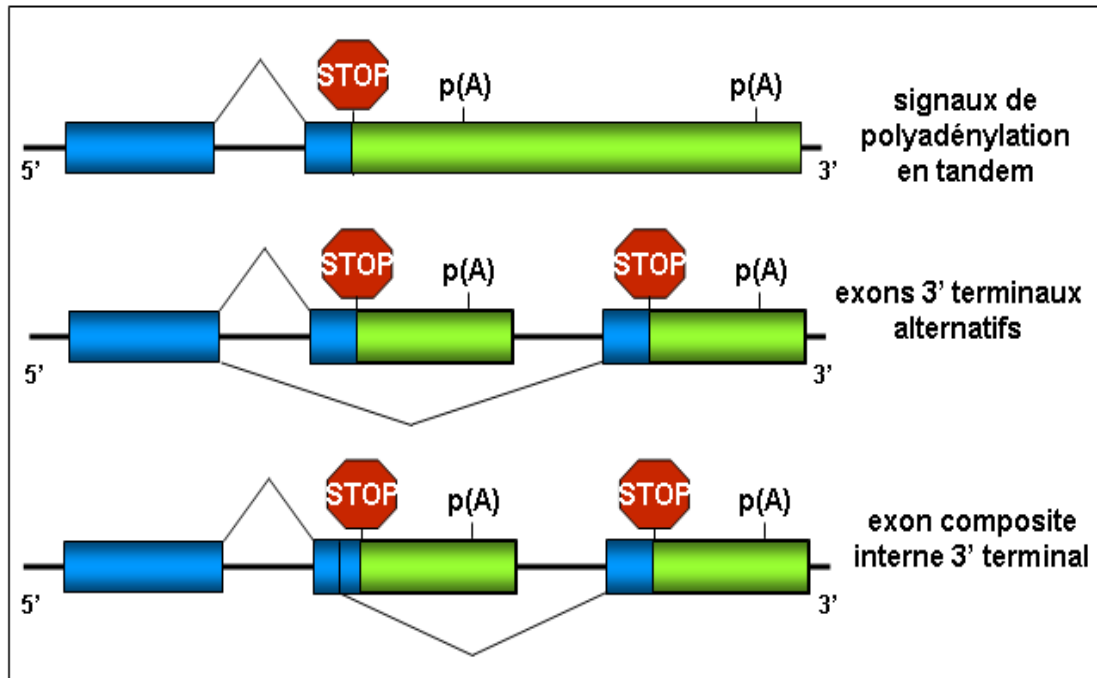


Figure 2 : Maturation en 3' alternative.

Représentation schématique des différentes formes de polyadénylation alternative. Les introns du pré-ARNm sont représentés en trait fin, les exons sont représentés en rectangle coloré. Le STOP correspond au stop de traduction, en bleu la partie 3' de l'ORF et en vert, les 3'UTR. p(A) représente le site de polyadénylation.

II Mécanisme de la maturation en 3'

II-1 La maturation en 3' des ARNm contenant une queue polyadénylée

La plupart des pré-ARNm eucaryotes, à l'exception des pré-ARN d'histones, sont polyadénylés. Bien qu'il existe une relative conservation de la maturation en 3' chez tous les eucaryotes, nous nous sommes plus particulièrement intéressés aux mécanismes conduisant à la synthèse d'ARNm polyadénylés chez les Métazoaires. Ces mécanismes font intervenir différents éléments de séquences constituant le signal de polyadénylation ainsi qu'un complexe enzymatique de maturation en 3'.

II-1-1 Les éléments de séquence permettant la polyadénylation

La maturation en 3' se fait suite à la reconnaissance de différentes séquences parmi lesquelles le signal de polyadénylation et le DSE (« Downstream Sequence Element ») qui permettent la polyadénylation en aval du site de clivage.

II-1-1-1 Le signal de polyadénylation

Le signal de polyadénylation appelé également hexamère ou hexanucléotide est constitué de la séquence AAUAAA. Il a été découvert par la comparaison des séquences précédant le site de clivage dans différents ARNm (Proudfoot and Brownlee, 1976). Depuis cette séquence a été retrouvée dans de très nombreux ARNm de cellules animales (Wahle and Kuhn, 1997). Cette séquence située entre 10 et 30 nucléotides en amont du site de clivage est essentielle pour la maturation en 3' (Manley, 1988). Néanmoins, il existe d'autres signaux de polyadénylation (Beaudoing et al., 2000; Tian et al., 2005) (Tableau 1) tel que le AUUAAA qui est le plus représenté après le signal canonique. Ces autres hexamères entraînent une plus faible efficacité de clivage et de polyadénylation des pré-ARNm les possédant (Nunes et al., 2010; Sheets et al., 1990).

Tableau 1: Efficacité de maturation en 3' associée aux différents signaux de poly(A).

	Fréquence (%)	Polyadénylation (%)	Clivage (%)
AAUAAA	58.2	100	100
AUUAAA	14.9	77 ± 4.7	66 ± 6.4
UAUAAA	3.2	17 ± 3.0	ND
AGUAAA	2.7	29 ± 8.1	30 ± 7.2
AAUAUA	1.7	10 ± 2.3	ND
AAUACA	1.2	11 ± 2.3	ND
CAUAAA	1.3	18 ± 6.4	28 ± 1.4
GAUAAA	1.3	11 ± 1.0	30 ± 2.8
UUUAAA	1.2	ND	ND
AAGAAA	1.1	6 ± 1.0	ND
AAUGAA	0.8	4.3 ± 0.6	ND
ACUAAA	0.6	11 ± 6.0	11 ± 1.8
AAUAGA	0.7	3.3 ± 1.5	2 ± 2.8

La fréquence des différents signaux de polyadénylation a été déterminée à partir plus de 22 000 gènes humains. Le pourcentage d'efficacité de polyadénylation et de clivage est calculé à partir d'une efficacité maximale pour le signal de poly(A) AAUAAA. ND : non déterminé. Ce tableau s'inspire de (Beaudoing et al., 2000) pour la fréquence et de (Sheets et al., 1990) pour les efficacités de clivage et de polyadénylation.

II-1-1-2 Le « Downstream sequence element »

Le DSE (« Downstream Sequence Element ») a été identifié par des délétions successives en aval du site de clivage sur des plasmides contenant les séquences de polyadénylation de l'antigène de surface du virus de l'hépatite B. Ces différentes délétions réduisent l'efficacité de maturation en 3' (Simonsen and Levinson, 1983). Cet autre élément est localisé à plus de 30 nucléotides en aval du site de clivage. Il n'est plus présent sur l'ARN clivé et n'est donc pas impliqué dans l'étape de la polyadénylation. Le DSE est moins conservé que l'hexanucléotide. Il peut être soit riche en uraciles (UUUUU) (Chou et al., 1994), soit riche en guanines et uraciles YGUGUUY (Y=pyrimidine) (McLauchlan et al., 1985). Ces DSE ne sont pas forcément présents sur un pré-ARNm, mais ils peuvent également être présents en plusieurs exemplaires (McDevitt et al., 1986; Sittler et al., 1994). De plus, une mutation ponctuelle du DSE n'a que peu d'effet sur la maturation en 3'. En revanche des délétions plus larges modifient la position du site de clivage (Mason et al., 1986) ainsi que l'efficacité de clivage (Gil and Proudfoot, 1987).

II-1-1-3 Le site de clivage

Le site de clivage se situe généralement entre 10 et 30 nucléotides en aval de l'hexamère et à 30 nucléotides, ou plus, en amont du DSE. Le site de clivage n'est pas strictement conservé. L'ordre de préférence pour le nucléotide du site de clivage est A>U>C>>G (Chen et al., 1995). Chez les vertébrés, 70% des sites de clivage sont localisés en 3' d'un résidu adénosine (Sheets et al., 1990). De ce fait, le premier résidu d'adénosine de la queue poly(A) est généralement un nucléotide transcrit (Moore et al., 1986). Le nucléotide précédant le site de clivage est une cytosine dans 59% des cas (Sheets et al., 1990). Le site de clivage est donc le plus souvent constitué du dinucléotide CA (Zhao et al., 1999) et le clivage a généralement lieu après le résidu adénosine (Sheets et al., 1990).

II-1-2 La machinerie de maturation en 3'

La machinerie de maturation en 3' requiert un large complexe composé de facteurs multi-protéiques. Chez les Métazoaires, CPSF (« Cleavage and Polyadenylation Specificity Factor »), CstF (Cleavage stimulation Factor), CFIm et CFIIIm (« Cleavage Factor I et II »), la PAP (Poly(A) Polymérase), PABP2 (« Poly(A) Binding Protein ») ainsi que la CTD Pol II (partie carboxy-terminale de la RNA polymérase II) sont indispensables à la réaction de clivage en 3' réalisée avec des facteurs purifiés *in vitro*. Alors que seuls, CPSF, PAP et PABP2 sont nécessaires à une polyadénylation *in vitro* (Wahle, 1991).

Le rôle des différentes protéines constituant ces facteurs de maturation en 3' est détaillé dans le tableau 2.

Tableau 2: Facteurs de maturation en 3' et fonctions associées.

Facteur de maturation	Etapas	Nom de la protéine	Interaction	Mécanisme	Références
CPSF	Clivage Poly(A)	CPSF-160	CstF-77, CTD pol II, PAP, hFip1	Reconnaît la séquence AAUAAA Stabilise le complexe de maturation en 3'	(Murthy and Manley, 1992, 1995) (Jenny et al., 1994)
		CPSF-100	CPSF-73		(Dominski et al., 2005b; Mandel et al., 2006; Ryan et al., 2004)
		CPSF-73	CPSF-100,	Endoribonucléase responsable du clivage en 3'	
		CPSF-30	hFip1	Reconnaît les séquences riches en U	(Barabino et al., 1997)
		hFip1	PAP, CPSF-160, CPSF-30, CstF-77	Stabilise le complexe de maturation en 3'	(Kaufmann et al., 2004)
CstF	Clivage	CstF-77	CPSF-160, CstF-64, CstF-50, CTD pol II	Stabilise le complexe de maturation en 3'	(Bai et al., 2007; McCracken et al., 1997)
		CstF-64	CstF-77, Symplektine	Reconnaît les séquences riches en G/U	(Takagaki and Manley, 1997)
		CstF-50	CstF-77, CTD pol II	Stabilise le complexe de maturation en 3'	(Takagaki and Manley, 2000)
CFIm	Clivage Poly(A)	CFIm-25	PAP, CFIm-68	Aide à reconnaître la séquence AAUAAA	(Ruegsegger et al., 1998)
		CFIm-68/59/72	CFIm-25	Stabilise le complexe de maturation en 3'	(Ruegsegger et al., 1996)
CFIIm	Clivage	hPcf11	CTD pol II, hClp1	Dissocie les facteurs de maturation en 3' de la pol II	(Hollingworth et al., 2006; Meinhart and Cramer, 2004; Noble et al., 2005; Zhang et al., 2005)
		hClp1	CFIm, CPSF, hPcf11	Permet une interaction entre CPSF et CFIm et recrute de l'ATP	(de Vries et al., 2000)

INTRODUCTION

CTD pol II	Clivage	CTD pol II	hPcf11, CPSF-160, CstF-50, CstF-77	Permet de recruter plusieurs facteurs de maturation en 3'	(Higgs et al., 1983; McCracken et al., 1997; Meinhart and Cramer, 2004; Ryan et al., 2002)
PAP	Clivage Poly(A)	PAP	hFip1, CFIm, CPSF-160, PABP2	Polymérise la queue poly(A) sur le pré-ARNm clivé	(Balbo and Bohm, 2007; Martin et al., 2004; Wahle, 1991)
PABP2	Poly(A)	PABP2	CPSF, PAP	Permet une polyadénylation processive	(Kuhn et al., 2009)
Symplekine	Clivage	Symplekine	CstF-64, CstF-77, CPSF	Permet une interaction entre CPSF et CstF	(Ruepp et al., 2011; Takagaki and Manley, 2000)
WDR33	Clivage	WDR33	CPSF-73	Non défini	(Shi et al., 2009)
PP1	Poly(A)	PP1		Non défini	(Shi et al., 2009)

Ce tableau s'inspire de (Mandel et al., 2008).

II-1-3 Mode d'action de la maturation en 3'

Il existe une coopérativité de fixation des différents facteurs de la machinerie de maturation en 3' qui permet une stabilisation globale du complexe sur le pré-ARNm. L'ordre réel de l'arrivée de chaque facteur sur le pré-ARNm n'est donc pas clairement établi. Néanmoins, le complexe de maturation en 3' est recruté à proximité du site de clivage grâce à son interaction avec la partie CTD pol II. Les facteurs CFIm et CPSF sont tous deux associés précocement sur le pré-ARNm (Rappsilber et al., 2002; Venkataraman et al., 2005). Le signal de polyadénylation est reconnu par CPSF, suggérant que ce facteur serait le premier à reconnaître le site de maturation en 3' (Murthy and Manley, 1992, 1995). Toutefois, la haute affinité de CFIm pour l'ARN et sa capacité à accélérer l'assemblage de la machinerie de maturation en 3' indique qu'il serait recruté de manière précoce (Ruegsegger et al., 1998). L'assemblage serait ensuite rapidement complété par la fixation de CstF sur le pré-ARNm (Ruegsegger et al., 1998), et de CFIm sur le complexe de maturation en 3'.

INTRODUCTION

CPSF-160 reconnaît le AAUAAA avec l'aide de CPSF-30 et probablement CPSF-100 alors que CstF-64 reconnaît le DSE (Murthy and Manley, 1995; Ruegsegger et al., 1998). Les interactions entre CPSF-160, CstF-77 et le pré-ARNm sont stabilisées par les autres sous-unités de ces facteurs (Jenny et al., 1994; Kaufmann et al., 2004) et par le recrutement de la symplektine (Ruepp et al., 2011; Takagaki and Manley, 2000). Enfin, CPSF permet le recrutement de la PAP (Kaufmann et al., 2004). C'est l'interaction entre CPSF et CstF qui définit la position du site de clivage (Zhao et al., 1999) mais la réaction de clivage nécessite le recrutement des autres facteurs de polyadénylation définis précédemment (tableau 2). Le complexe ainsi formé est positionné de part et d'autre du site de clivage et permet à CPSF73 de catalyser le clivage endonucléotidique du pré-ARNm (Mandel et al., 2006; Ryan et al., 2004). Après le clivage, le complexe de maturation en 3' se dissocie.

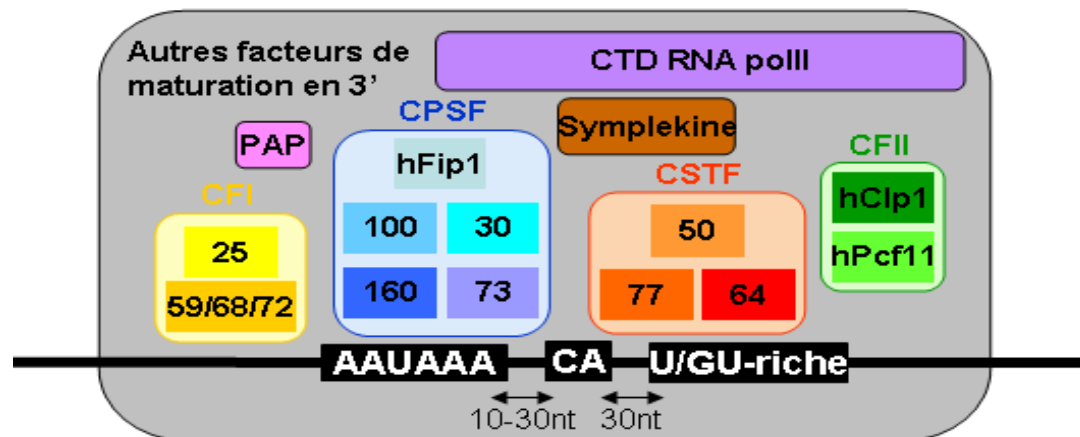


Figure 3 : Machinerie de la maturation en 3' des ARNm contenant une queue polyadénylée.
 Les éléments *in cis* sont représentés en rectangle noir, les facteurs de maturation en 3' sont représentés en couleur ou en gris. La position des différents facteurs de maturation en 3' prend en compte leurs interactions avec le pré-ARNm et dans la mesure du possible les interactions protéiques. Cette figure s'inspire de (Millevoi and Vagner, 2010)

Suite au clivage endonucléolytique, un remaniement du complexe est effectué s'accompagnant du départ des facteurs CstF, CFI_m et CFI_l. Seul CPSF associé à la PAP reste lié à la séquence AAUAAA en amont de l'extrémité 3' OH libre et permet une polyadénylation distributive entraînant l'ajout d'une dizaine de résidus

INTRODUCTION

d'Adénosines sur lesquels la PABP2 vient se fixer (Bienroth et al., 1993). La polyadénylation peut alors se faire de manière processive, une boucle de poly(A) recouverte de PABP2, est alors formée jusqu'à atteindre une taille de 250 résidus d'Adénosines. Lorsque la queue poly(A) atteint cette longueur, la structure est déstabilisée (Wahle, 1995) et conduit à la rupture de la liaison entre CPSF et la PAP, PABP2 ne pouvant plus stimuler l'interaction entre les deux facteurs. La fin de cette interaction entraîne l'arrêt de la polyadénylation et la dissociation du complexe CPSF, PAP avec l'ARN (Kuhn et al., 2009; Wahle, 1995).

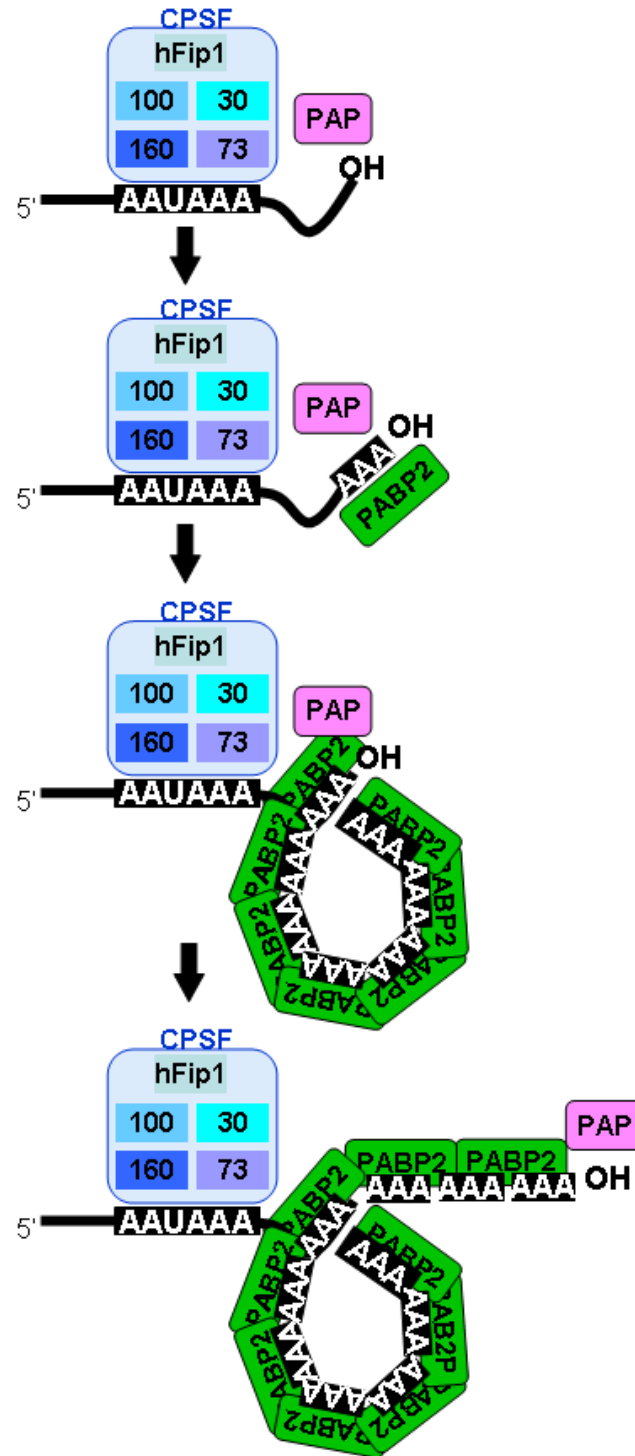


Figure 4 : Modèle d'allongement de la queue poly(A).

Les éléments en *cis* sont représentés en rectangle noir, les facteurs de polyadénylation sont représentés en couleur. CPSF lié au signal de polyadénylation AAUAAA recrute la PAP. La PABP2 rejoint le complexe une fois que la queue poly(A) a atteint une longueur d'environ 12 nucléotides. D'autres molécules PABP2 couvrent la queue poly(A) en cours d'élongation et forme une structure en boucle, nécessaire au maintien du complexe CPSF-PAP-PABP2 indispensable à l'élongation de la queue poly(A). Lorsque la queue poly(A) atteint une longueur critique d'environ 250 résidus, l'interaction entre CPSF-PAP n'est plus maintenue entraînant la dissociation du complexe CPSF, PAP avec l'ARN et l'arrêt de la polyadénylation.

II-2 Maturation en 3' des pré-ARNm d'histones

Les ARNm d'histones, contrairement aux autres ARNm ne sont pas polyadénylés (Dominski and Marzluff, 2007; Marzluff et al., 2008). Cependant, la simple insertion d'un intron à l'intérieur d'un gène d'histone (normalement dépourvu d'intron) transforme l'ARNm de l'histone en ARNm polyadénylé (Pandey et al., 1990). Les éléments de séquence de la maturation en 3' des histones se composent d'un site de clivage en 3' (préférentiellement CA) situé entre une séquence structurée en tige-boucle et un élément riche en purine appelé HDE (Histone Downstream Element) (Birchmeier et al., 1984; Mowry and Steitz, 1987). La tige-boucle est localisée 100 nucléotides en aval du codon stop. Sa séquence hautement conservée, est composée d'une tige de 6 bases appariées et d'une boucle de 4 nucléotides. L'HDE est moins conservé que la tige-boucle même si il existe une séquence consensus de 9 nucléotides (CAAGAAAGA) (Birnstiel et al., 1985; Mowry and Steitz, 1987).

La maturation en 3' des pré-ARNm d'histones commence par l'interaction sur la structure en tige boucle de la protéine SLBP (« Stem Loop Binding Protein ») contenant un domaine de liaison aux ARN. Cette interaction favorise la liaison entre l'HDE et la snRNP U7. SLBP interagit avec la protéine à doigt de zinc de 100 kDa ZFP100 (« Zinc Finger Protéin 100 ») (Dominski et al., 2002) qui recrute la sous unité Lsm11 de la snRNP U7 (Azzouz et al., 2005; Wagner and Marzluff, 2006). Ce complexe protéique permet de recruter CPSF-73, CPSF-100, la symplekin et la protéine pro-apoptotique FLASH (Dominski et al., 2005a; Kolev and Steitz, 2005; Wagner et al., 2007; Yang et al., 2009). La formation du complexe de maturation en 3' des histones entraîne l'activation de l'activité endonucléase de CPSF-73 permettant le clivage du pré-ARNm et la dissociation du complexe (Dominski et al., 2005a). Le produit de clivage se terminant par la tige-boucle associée à SLBP et suivie d'une séquence ACCCA correspondant à un ARNm mature d'histone. Le produit en aval contenant l'HDE est rapidement dégradé permettant la libération de la snRNP U7 qui peut alors être recyclée (Walther et al., 1998). Comme pour la maturation en 3' de pré-ARNm polyadénylés, la terminaison de la transcription des gènes des histones est dépendante de la maturation en 3' (Chodchoy et al., 1991).

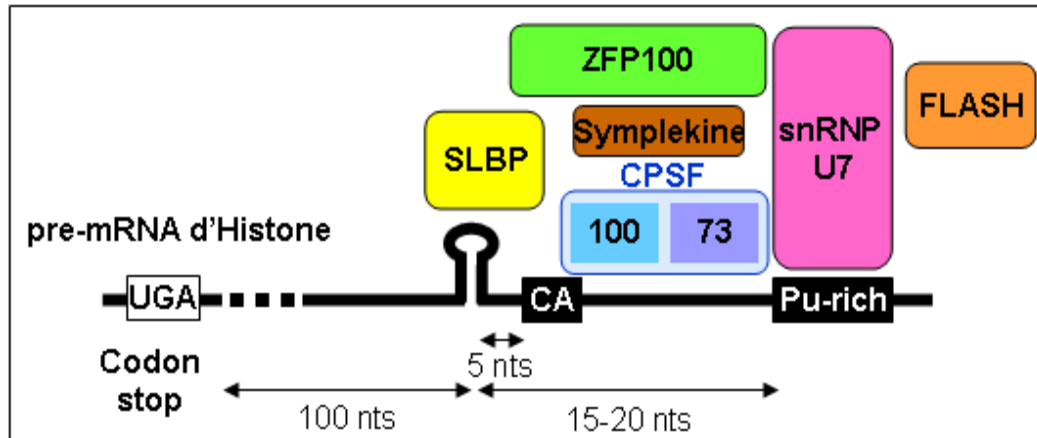


Figure 5 : Machinerie de la maturation en 3' des pré-ARNm d'histones.

Les éléments *in cis* sont représentés en rectangle noir, le codon stop est représenté en rectangle blanc et les facteurs *in trans* de maturation en 3' représentés en couleur. La position des différents facteurs de maturation en 3' prend en compte leurs interactions avec le pré-ARNm et dans la mesure du possible les interactions protéiques. Cette figure s'inspire de (Millevoi and Vagner, 2010)

III Interaction de la maturation en 3' avec les étapes de l'expression des gènes

Il existe un réseau complexe d'interactions fonctionnelles et physiques, entre les différentes étapes de l'expression des gènes (Bentley, 1999; Cramer et al., 2001; Hirose and Manley, 2000; Maniatis and Reed, 2002; Proudfoot, 2000). Bien que *in vitro* ces différentes étapes peuvent être reproduites indépendamment, *in vivo* ces réactions sont reliées par de nombreuses interconnexions (Maniatis and Reed, 2002; Neugebauer and Roth, 1997).

III-1 Interaction entre la maturation en 3' et la transcription

Le lien entre la maturation en 3' et la transcription est médié par la région C-terminale CTD (« C-terminal domain ») de la pol II. En effet, le CTD de la pol II est un facteur de maturation en 3' (Ryan et al., 2002). Une ARN pol II tronquée au niveau du CTD conduit à un blocage de l'épissage et de la maturation en 3' sans pour autant affecter la transcription (McCracken et al., 1997). Le CTD se compose d'un motif répété 52 fois chez l'homme ($^1\text{YSPTSPS}^7$)_n très conservé au cours de l'évolution. Lors du recrutement de la pol II sur un promoteur, ce motif du CTD n'est pas phosphorylé. Très tôt après l'initiation de la transcription, la sérine en position 5 de ce motif est phosphorylée par le facteur de transcription TFIIF (« Transcription Factor II H ») permettant de recruter plusieurs facteurs indispensables à l'ajout de la coiffe et à l'épissage (Das et al., 2007; Kameoka et al., 2004). Une fois que l'élongation de la transcription est initiée, la sérine en position 2 est phosphorylée à son tour par la kinase Cdk9 (« Cyclin-dependent kinase 9 ») (Komarnitsky et al., 2000) favorisant la fixation de facteurs d'épissage (Das et al., 2007). Durant l'élongation, la sérine 5 est déphosphorylée par la phosphatase Ssu72 (Hausmann et al., 2005). La phosphorylation unique de la sérine 2 serait alors favorable au recrutement de plusieurs facteurs de maturation en 3' notamment CPSF et CstF (Ahn et al., 2004; Gudipati et al., 2008). Ces études montrent l'importance de la

INTRODUCTION

phosphorylation et de la déphosphorylation du CTD de la pol II pour le recrutement des facteurs participant à la maturation des pré-ARNm. Le CTD permet un recrutement de différents composants des machineries de maturation à proximité du transcrit naissant (Listerman et al., 2006; Phatnani and Greenleaf, 2006).

Il a également été montré que le facteur associé à la pol II, Cdc73 (« Cell division cycle 73 ») interagit avec le complexe CPSF-CstF (Rozenblatt-Rosen et al., 2009). Cette association est nécessaire lors de la maturation en 3' *in vitro* et entraîne une augmentation de la maturation en 3' du pré-ARNm de INTS6 (« Integrator complex subunit 6 ») *in vivo*.

Un autre mécanisme de stimulation de la maturation en 3' par la pol II a également été caractérisé. En effet, plusieurs études menées par l'équipe de Proudfoot montrent que pour certains gènes, comme la β -globine (Enriquez-Harris et al., 1991), la β -actine (Gromak et al., 2006) ou le complément C2 (Yonaha and Proudfoot, 1999, 2000), la pol II marque une pause dans la région 3' en aval des pré-ARNm transcrits. Cette pause de la pol II semble jouer un rôle dans la terminaison de la transcription et stimule la maturation en 3'.

De manière plus générale, un autre type d'interaction entre la maturation en 3' et la transcription a été décrit. Une étude récente a montré que des mutations du signal de polyadénylation, AAUAAA, entraîne la diminution de l'initiation de la transcription sur le même gène (Mapendano et al., 2010). Les mutations du signal de polyadénylation inhibent la terminaison de la transcription entraînant une séquestration des facteurs d'initiation de la transcription TBP (« TATA box Binding Protein »), TFIIB (« Transcription Factor II B ») et CDK9 (« Cyclin-Dependent Kinase ») à l'origine de cette diminution d'initiation de la transcription (Mapendano et al., 2010).

Deux modèles généraux ont été proposés pour expliquer le lien entre la maturation en 3' et la transcription (Buratowski, 2005). Le premier, appelé « anti-terminator », propose que la reconnaissance des signaux de polyadénylation par la machinerie de maturation en 3' entraîne un changement conformationnel qui déstabilise les interactions entre les facteurs de polyadénylation (Logan et al., 1987). Le deuxième modèle, appelée « torpedo », requiert l'évènement de clivage en 3' du pré-ARNm qui génère deux fragments d'ARN. L'ARN néotranscrit en aval possédant

une extrémité 5' phosphate est rapidement dégradé par la 5'-3' ARN exonucléase Xrn2 ce qui entraîne la terminaison de la transcription et la libération de la pol II de la matrice d'ADN (Connelly and Manley, 1988; West et al., 2004).

L'un ou l'autre de ces différents processus peut être prédominant en fonction des gènes. En effet, certaines études montrent qu'une terminaison de la transcription est possible en absence de clivage en 3' (Osheim et al., 2002; Tran et al., 2001a). Néanmoins, la déplétion de Xrn2 provoque des défauts de terminaison de la transcription (West et al., 2004). De plus, la séquence du pré-ARNm β -globine en aval du site de polyadénylation peut se replier dans une structure en tête de marteau semblable au pré-ARNr (ribosomique) capable d'auto-clivage (Teixeira et al., 2004). Après l'auto-clivage du pré-ARNm de la β -globine, Xrn2 dégrade le brin néotranscrit et permet la terminaison de la transcription.

Il semble probable que les modèles « anti-terminator » et « torpedo » collaborent pour permettre la terminaison de la transcription. Il a été proposé un modèle hybride ou unifié qui inclut certains aspects de ces deux modes « anti-terminator » et « torpedo » (Luo and Bentley, 2004). Un complexe composé de la pol II, des facteurs de maturation en 3' et Xrn2, se forme de manière coopérative aux abords du signal de polyadénylation. Ce complexe subirait plusieurs transitions au niveau du site de clivage : un changement conformationnel du site catalytique de la pol II qui diminue sa processivité ainsi qu'une dégradation médiée par Xrn2 de l'ARN naissant. La dissociation du complexe serait alors facilitée par les facteurs de terminaison (Luo and Bentley, 2004; Luo et al., 2006).

III-2 Interaction entre la maturation en 3' et la coiffe

La coiffe m⁷G(5')ppp(5')N en 5' d'un pré-ARNm améliore son efficacité de clivage en 3' *in vitro* alors que l'ajout de la queue poly(A) n'est pas affectée par la présence de la coiffe (Cooke and Alwine, 1996; Gilmartin et al., 1988; Hart et al., 1985; Lewis and Izaurralde, 1997). Cette stimulation passe par le CBC (« cap binding complex ») formé des protéines CBP80 et CBP20 (« Cap Binding Protein »). En effet, la stimulation de la maturation en 3' par la coiffe est inhibée par la déplétion du CBC. Le mécanisme exact n'est pas encore clairement établi mais il a été montré que le CBC interagit avec la machinerie de polyadénylation et augmente sa stabilité

(Flaherty et al., 1997). Plus récemment, il a été montré que les facteurs d'ajout de la coiffe RNGTT (« RNA Guanylyltransferase and 5'-phosphatase ») et RNMT (RNA (guanine-7-) methyltransferase) localisés dans la région 5' terminale des gènes peuvent également être présents dans la région 3' terminale et entraîner une pause de la pol II qui stimule la maturation en 3' (Glover-Cutter et al., 2008)

III-3 Interaction entre la maturation en 3' et l'épissage

Les interactions entre l'épissage et la maturation en 3' ont été mises en évidence à l'aide d'ARN chimériques dérivés de virus qui ont permis de montrer que la présence de signaux de polyadénylation stimule l'épissage du dernier intron (Cooke et al., 1999; Niwa and Berget, 1991; Scott and Imperiale, 1996). Les bases moléculaires de ces interactions entre épissage du dernier intron et maturation en 3' sont principalement dues à des interactions entre les facteurs d'épissage et les facteurs de maturation en 3' (Kyburz et al., 2006; Millevoi et al., 2006; Vagner et al., 2000b).

En effet, l'adressage de la région C-terminale de la PAP sur un exon stimule l'épissage de l'intron situé en amont (Vagner et al., 2000b). Cette stimulation est due à une interaction entre le domaine RS situé dans la région amino-terminale de U2AF65 et le domaine carboxy-terminal de la PAP (Vagner et al., 2000b). De plus, l'adressage de U2AF65 en amont d'un signal de polyadénylation favorise le clivage en 3' d'un pré-ARNm (Millevoi et al., 2002) alors que des mutations de la région riche en pyrimidines du site 3' d'épissage, qui diminuent la fixation de U2AF65, réduisent l'efficacité du clivage (Millevoi et al., 2002). Le domaine RS de U2AF65 interagit également avec CFIm-59 (Millevoi et al., 2006). L'adressage de U2AF65 en amont d'un signal de polyadénylation facilite le recrutement de CFIm stimulant ainsi le clivage (Millevoi et al., 2006). Il a donc été proposé que l'interaction entre U2AF65 et la PAP a lieu quand le signal de polyadénylation est fort, aidant ainsi à la reconnaissance d'un site 3' d'épissage faible. L'interaction entre U2AF65 et CFIm favorise la reconnaissance d'un signal polyadénylation faible par un site 3' d'épissage fort (Millevoi et al., 2006).

Un mécanisme similaire se produit avec la snRNP U2 qui lie le point de branchement du dernier intron et interagit avec CPSF. En effet, la purification du complexe CPSF à partir de thymus de veau a permis d'identifier plusieurs sous-unités

de la snRNP U2 (Kyburz et al., 2006). *In vitro*, la présence de CPSF est nécessaire pour un épissage efficace d'exons 3' terminaux et la digestion du snRNA U2 conduit à une diminution du clivage d'un pré-ARNm modèle (Kyburz et al., 2006)

De plus, il a été montré que la partie C-terminale de la PAP interagit avec U1A et U170K (composants de la snRNP U1). L'ajout sur des extraits nucléaires, d'un peptide nommé PRD (« PAP Regulatory Domain »), correspondant à la région de U1A interagissant avec la PAP, dissocie les réactions d'épissage et de polyadénylation (Gunderson et al., 1998; Gunderson et al., 1997; Moore and Sharp, 1985; Vagner et al., 2000a). La fixation du domaine PRD inhibe l'activité de polyadénylation de la PAP (Ko and Gunderson, 2002). La protéine U170K co-immunoprécipite également avec le facteur CFIm (Awasthi and Alwine, 2003). Une autre étude montre au contraire, que le blocage de la snRNP U1 inhibe la polyadénylation (Moore and Sharp, 1985). L'influence de l'interaction entre la snRNP U1 et la machinerie de maturation en 3' n'est donc pas clairement établie.

Un autre exemple de couplage entre la maturation en 3' et l'épissage fait intervenir le facteur SRm160. Ce facteur coactivateur de l'épissage se fixe sur les éléments ESE (« Exonic Splicing Enhancers ») *via* son domaine amino-terminal PWI (« proline tryptophane isoleucine ») (Blencowe et al., 1998). SRm160 interagit avec CPSF et CstF50 pour activer le clivage en 3' des pré-ARNm (McCracken et al., 2002; McCracken et al., 2003; Szymczyna et al., 2003).

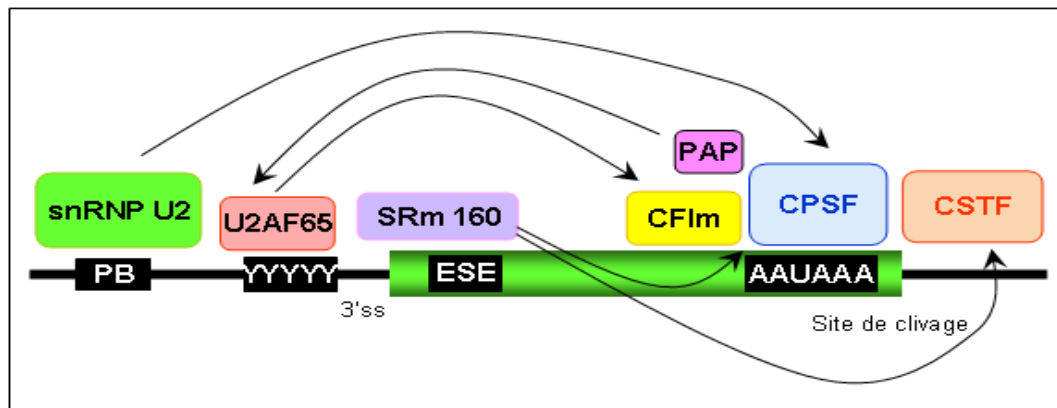


Figure 6 : Couplage entre la maturation en 3' et l'épissage.

Les éléments en *cis* sont représentés en rectangle noir, PB représente le point de branchement, Y représente des bases pyrimidiques, ESE représente un « Exonic Splicing Enhancer » et le trait vert représente le dernier intron. Les facteurs en *trans* sont représentés en couleur. La position des différents facteurs en *trans* prend en compte leurs interactions avec les éléments de séquence du pré-ARNm. Les flèches noires indiquent une interaction stimulatrice.

Il existe donc de nombreuses interactions entre l'épissage et la maturation en 3'. De plus, l'épissage alternatif peut être déterminant dans le choix du dernier exon et donc du site de maturation en 3' (Edwards-Gilbert et al., 1997).

III-4 Interaction entre la maturation en 3' et l'export nucléaire des ARNm

La polyadénylation est indispensable pour le transport de l'ARNm du noyau vers le cytoplasme. En effet, il a été montré que la substitution des signaux de polyadénylation par le site de clivage d'un ARN ribosomique, produisant un ARNm clivé mais non polyadénylé, entraîne une diminution de la quantité d'ARNm cytoplasmique par rapport à la quantité d'ARNm nucléaire. Cependant, l'ajout d'une suite d'adénines transcrites en amont du site de clivage de l'ARN ribosomique ne permet pas de rétablir l'export nucléaire de l'ARNm polyadénylé (Huang and Carmichael, 1996). Cette expérience montre le rôle essentiel de l'étape de maturation en 3', dans le transport de l'ARNm du noyau vers le cytoplasme. Il a également été montré que la mutation de séquences *cis* ou de facteurs *trans* activateurs de la maturation en 3' bloque l'export des ARNm vers le cytoplasme (Brodsky and Silver, 2000; Hammell et al., 2002; Hilleren et al., 2001). Enfin, les transcrits générés par l'ARN polymérase T7 ne sont exportés que s'ils sont correctement maturés en 3' (Dower and Rosbash, 2002). Ceci indique que la maturation en 3' est nécessaire à l'export, et que les facteurs d'export peuvent être recrutés sur des ARNm de manière indépendante de la pol II (Dower and Rosbash, 2002).

III-5 Stabilisation de l'ARNm par la queue poly(A)

La maturation en 3' permet de stabiliser l'ARNm dégradé par des ARN exonucléases. Les extrémités 3' et 5' doivent donc être protégées. La coiffe de l'extrémité 5' protège de l'action des ARN exonucléases 5'-3' alors que la queue poly(A) permet de protéger de l'action des ARN exonucléases 3'-5' (Wilusz et al., 2001). La queue poly(A) est reconnue par la PABP ce qui prévient la dégradation de l'ARN (Ford et al., 1997; Wickens et al., 1997). De plus, il a été montré que seul le

recrutement de la PABP sans queue poly(A) est suffisant pour protéger l'extrémité 3' de l'ARNm de la dégradation (Coller et al., 1998). La queue polyadénylée permet donc, par l'intermédiaire des PABP, de protéger l'ARNm de la dégradation.

Chez les mammifères, il existe plusieurs déadénylases qui en dégradant la queue poly(A) permettent de dégrader de manière rapide l'ARNm. Ces déadénylases peuvent soit dégrader l'ARNm total, soit dégrader des ARNm spécifiques ou simplement de réduire la taille de la queue poly(A) (Zhang et al., 2010). Elles sont impliquées dans de nombreux processus tel que la réponse aux dommages à l'ADN (Cevher et al., 2010), le système NMD (« Nonsense-Mediated Decay ») (Garneau et al., 2007), la dégradation des ARNm par les microRNA (Wakiyama et al., 2007) et dans le contrôle de la différenciation cellulaire (Kimble and Crittenden, 2007).

III-6 Interaction entre la queue polyadénylée et la traduction

La reconnaissance de la queue poly(A) par la PABP permet l'initiation de la traduction et favorise le recrutement des ribosomes (Derry et al., 2006; Gallie, 1998; Kahvejian et al., 2005). La PABP interagit avec le facteur d'initiation de la traduction eIF4G lié à la coiffe, entraînant la circularisation de l'ARNm. Cette circularisation est essentielle à la traduction (Wakiyama et al., 2000) et la liaison de eIF4G avec la PABP stimule l'interaction entre eIF4G et la coiffe, augmentant le recrutement de la sous unité ribosomique 60s ainsi que le recyclage des ribosomes. (Kahvejian et al., 2001).

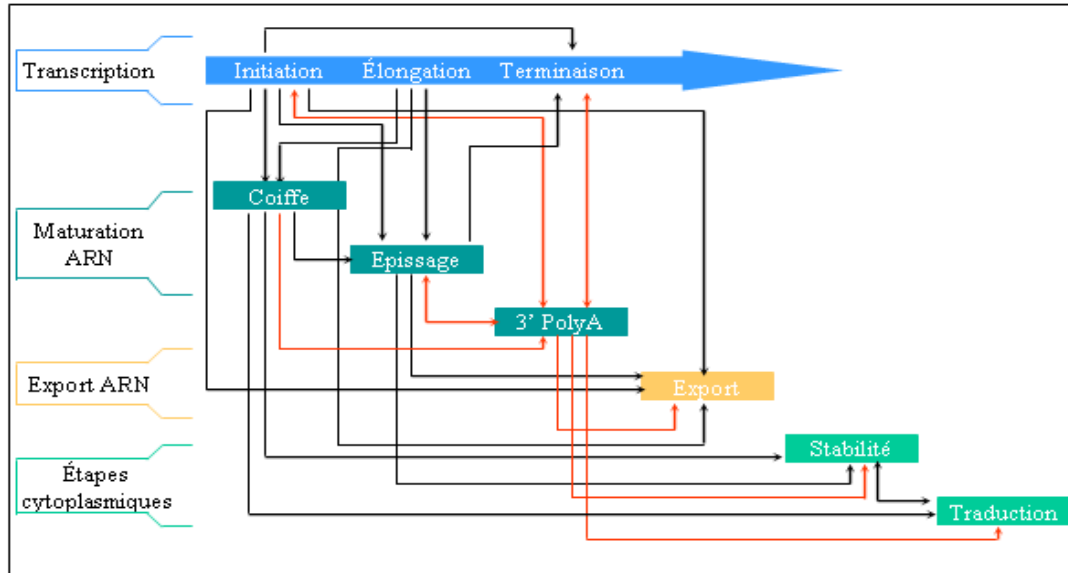


Figure 7 : Un réseau complexe d'interactions entre les différentes étapes de l'expression des gène.

Les flèches noires indiquent un couplage fonctionnel et/ou physique entre 2 étapes de l'expression des gènes. Les flèches rouges indiquent un couplage fonctionnel et/ou physique entre la maturation en 3' et une autre étape de l'expression des gènes. Cette figure s'inspire de (Maniatis and Reed, 2002).

IV Régulation de la maturation en 3'

La régulation de la maturation en 3' est un processus complexe capable de réguler l'expression des ARNm au niveau qualitatif et quantitatif. Ce processus complexe fait intervenir de nombreux facteurs de régulation interagissant avec différents éléments de séquences régulateurs.

IV-1 Eléments de Séquences régulateurs de la maturation en 3'

IV-1-1 Le « upstream sequence element »

Il existe 2 types d'éléments de séquence localisés en amont du signal de polyadénylation appelés USE (« Upstream Sequence Element »).

Les premiers types d'USE sont riches en uraciles bien qu'il n'y ait pas de séquence consensus. Ces USE stimulent la maturation en 3' et ont été identifiés au départ sur des séquences virales tel que SV40 L (« Simian Virus 40 large ») (Carswell and Alwine, 1989; Schek et al., 1992), ADL3 (Adénovirus L3) (Prescott and Falck-Pedersen, 1994) ou HIV-1 (« Human Immunodeficiency Virus de type 1 ») (Cherrington and Ganem, 1992; DeZazzo et al., 1991; Valsamakis et al., 1991). Depuis, ces USE ont été identifiés sur des gènes humains tels que le complément C2 (Moreira et al., 1995), le collagène (Natalizio et al., 2002), la lamine B2 (Brackenridge and Proudfoot, 2000) ou la prothrombine F2 (Danckwardt et al., 2007). Ces USE stimulent la maturation en 3' en recrutant différents facteurs capables de stimuler la maturation en 3' (Danckwardt et al., 2011; Moreira et al., 1998).

Les autres types d'USE sont formés de la séquence UGUA répétée. Ils ont été décrits dans plusieurs pré-ARNm pour stimuler la maturation en 3' (Brown and Gilmartin, 2003). La régulation de la maturation en 3' par ces USE peut se faire par le recrutement de manière directe de CFIm qui stimule à son tour le recrutement de CPSF (Venkataraman et al., 2005).

IV-1-2 Le « Downstream sequence element auxiliaire »

Les DSE aux (« Downstream Sequence Element auxiliaire ») stimulateurs de la maturation en 3', sont riches en guanines et sont localisés entre 1 et 200 nucléotides en aval du site de clivage (Zarudnaya et al., 2003), mais peuvent être localisés plus en aval comme sur le gène du MC1R (« Melanocortine 1 récepteur »). En effet, pour ce gène, une séquence riche en guanines localisée à plus de 440 nucléotides en aval du site de clivage permet de stimuler la maturation en 3' (Dalziel et al., 2007). Ces DSE aux sont présents dans un tiers des pré-ARNm (Arhin et al., 2002) et sont susceptibles de se structurer en G-quadruplexe (Zarudnaya et al., 2003).

Les G-quadruplexes sont des empilements hydrophobes de plusieurs G-quartets. Les G-quartets correspondent à l'association de 4 guanines dans un plan cyclique par des liaisons de type « hoogsteen ». Ces guanines sont à la fois donneurs et accepteurs de 2 liaisons hydrogènes et délimitent une zone centrale électro-négative qui recrute un cation stabilisant l'interaction (figure 8). La taille du cation et le niveau d'interaction avec les guanines expliquent les différences de stabilité ($K^+ > Rb^+ > Na^+ > Cs^+ > Li^+$) des G-quadruplexes (Sundquist and Klug, 1989).

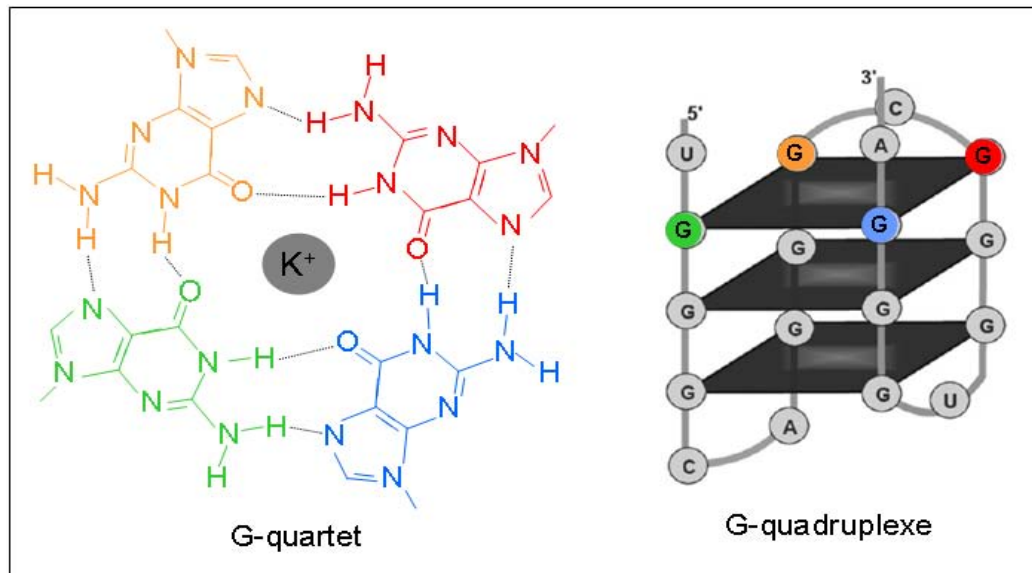


Figure 8 : Représentation schématique d'un G-quadruplexe.

G-quartet en vue transversale. 4 guanines sont associées par une liaison de type « hoogsteen » dans un plan cyclique. La présence de K^+ au centre de la structure permet de la stabiliser. Un G-quadruplexe correspond à l'empilement de plusieurs G-quartets. Cette figure s'inspire de (Kostadinov et al., 2006). Les guanines du G-quartet sont représentées en couleurs.

INTRODUCTION

Les G-quadruplexes sont principalement décrits dans l'ADN au niveau des télomères (Oganesian and Bryan, 2007; Parkinson et al., 2002; Wang and Patel, 1993), des régions de commutation des gènes des chaînes lourdes des immunoglobulines (Sen and Gilbert, 1990), des régions répétées de l'ADNr (Hanakahi et al., 1999) et des promoteurs (Brooks et al., 2010; Eddy and Maizels, 2006). Ces structures sont également présentes sur des ARNm où elles sont principalement décrites pour réguler la traduction (Beaudoin and Perreault, 2010; Bonnal et al., 2003; Halder et al., 2009; Kumari et al., 2007)

Il a été proposé que les DSE aux G-riches permettent de stimuler la maturation en 3' de différentes manières (Chen and Wilusz, 1998).

Premièrement, les DSE aux riches en guanines pourraient maintenir les signaux de polyadénylation dans un état non structuré entraînant une meilleure efficacité d'assemblage du complexe de maturation en 3'. En effet, un G-quadruplexe pourrait maintenir le DSE ou l'hexanucléotide dans une conformation accessible aux facteurs de maturation en 3' (Zarudnaya et al., 2003).

Deuxièmement, les DSE aux pourraient, en formant un G-quadruplexe, empêcher CstF de coulisser en aval du pré-ARNm et donc de limiter la zone d'interaction de CstF proche du signal de polyadénylation (Chen and Wilusz, 1998). En effet, la stimulation de la maturation en 3' par la séquence DSE aux du gène viral ADL3 peut être remplacée par une séquence capable de former un pseudoknot (structure d'acide nucléique contenant au moins deux structures en tige-boucle dont la moitié d'une tige est intercalée entre les deux moitiés d'une autre tige).

Troisièmement les DSE aux riches en guanines pourraient recruter des protéines de régulation de la maturation en 3' et notamment la famille des hnRNP H/F qui stimule la maturation en 3' (Bagga et al., 1998; Bagga et al., 1995).

Enfin, il a été décrit que les séquences riches en guanines en aval du signal de polyadénylation entraîne l'arrêt de la pol II permettant alors de stimuler la polyadénylation (Gromak et al., 2006; Yonaha and Proudfoot, 2000). Les G-quadruplexes qui peuvent également se structurer sur l'ADN pourraient être à l'origine de ce blocage de la pol II.

IV-1-3 Autres éléments de séquences

De nombreux éléments de séquence humains ont été identifiés par différents programmes informatiques (Hu et al., 2005; Legendre et al., 2005). Ces signaux de polyadénylation ont été identifiés en comparant différentes séquences humaines encadrant le site de maturation en 3'. Ces autres éléments sont conservés chez la plante et la levure et pourraient avoir un rôle dans la régulation de la maturation en 3' (Hu et al., 2005).

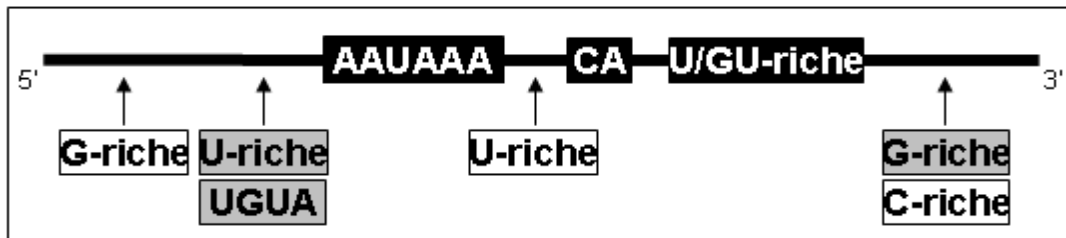


Figure 9 : Schéma représentant les différents signaux de polyadénylation humain.

Les signaux canoniques de la maturation en 3' sont représentés par des rectangles noirs, les signaux de stimulation de la polyadénylation sont représentés en rectangles gris et en rectangles blancs les signaux de polyadénylation identifiés par programme bioinformatique. Cette figure s'inspire de (Hu et al., 2005).

IV-2 Les facteurs de régulation de la maturation en 3'

Il existe un nombre croissant de protéines identifiées comme régulatrices de la maturation en 3'. Ces protéines sont pour l'essentiel des protéines de liaison aux ARN ou mRBP (« mRNA Binding Protein ») et interagissent avec les éléments de séquences régulateurs et la machinerie de polyadénylation. La plupart de ces facteurs protéiques sont également impliqués dans d'autres étapes de régulation de l'expression des gènes tel que la transcription, l'épissage ou l'export d'ARNm vers le cytoplasme.

Tableau 3: Protéines impliquées dans la régulation de la maturation en 3' chez les mammifères. Ce tableau s'inspire de (Millevoi and Vagner, 2010).

Facteur	Role	Inter-action	ARN cible	Séquence reconnue	Reference
14-3-3ε	-	PAP	La plupart des pre-ARNm (c-mos)		(Kim et al., 2003)
9G8	+	CFIm	HIV-1	Site de liaison consensus	(Valente et al., 2009)
BARD1	-	CstF50 pol II	Mécanisme global (SV40L et AdL3)		(Kleiman and Manley, 2001)
Cdc73	+	CPSF CstF	Pré-ARNm spécifiques (sous unité 6 du complexe d'integration)		(Rozenblatt-Rosen et al., 2009)
CSR1	-	PAP	Mécanisme global (β-actin et GAPDH)		(Zhu et al., 2009)
eIF3f	-	CFIm	HIV-1		(Valente et al., 2009)
ELL2	+	Cstf64 pol II	Immunoglobulin heavy chain locus	Promoteur	(Martincic et al., 2009)
FBP2 FBP3	-		Pré-ARNm spécifiques (Prothrombine)	USE	(Danckwardt et al., 2011)
hnRNP F	-	CstF64	IgM	DSE	(Veraldi et al., 2001)
hnRNP H	+	CstF64	Mécanisme global (SV40L, melanocortine 1 récepteur et et autre)	DSE aux	(Alkan et al., 2006; Dalziel et al., 2007)

INTRODUCTION

			Rous Sarcoma Virus	NRS (« Upstream negative regulator of splicing »)	(Wilusz and Beemon, 2006)	
hnRNP I (PTB)	+	CstF64	Cyclooxygenase-2	USE	(Guang et al., 2005)	
			Complement C2	USE	(Moreira et al., 1998)	
			Gfap (« Glial Fibrillary Acidic Protein »)			(Blechingberg et al., 2007)
			« Calcitonin/Calcitonin Gene-Related Peptide» (CT/CGRP)	Sites d'épissage et éléments riches en pyrimidines localisés en aval du site de clivage		(Lou et al., 1999)
			Pré-ARNm spécifiques (Prothrombine F2)	USE		(Danckwardt et al., 2007)
	-	CstF64	Pré-ARNm spécifiques (β -globin, l' α -globin et le Complement C2)	DSE		(Castelo-Branco et al., 2004)
		α -Tropomyosin	Elément riches en pyrimidines localisé en amont du site de clivage		(Le Sommer et al., 2005)	
hnRNP L	+		Herpes simplex virus thymidine kinase	Stimulateur localisé en amont du site de clivage	(Guang et al., 2005)	
			Pré-ARNm spécifiques (N-acylsphingosine amidohydrolase acid ceramidase 1)		(Hung et al., 2008)	
HSF1	+	CstF64 symplekine	Influence les protéines de réponse au choc thermique		(Xing et al., 2004)	
Hu	-	CstF64 CPSF160	Pré-ARNm spécifiques (CT/CGRP)	Séquences riches en U proche du AAUAAA	(Zhu et al., 2007)	
IRBIT	-	PAP hFip1	Pré-ARNm spécifiques (SV40L)	AAUAAA	(Kiefer et al., 2009)	
Nova	+/-		Transcrits du cerveau de souris	Motif YCAY dans le 3'-UTR	(Licatalosi et al., 2008)	
P54	+		COX-2 (Cyclooxygenase-2)	USE	(Hall-Pogar et al., 2007)	
			Mécanisme global (SV40L)		(Liang and Lutz, 2006)	
PIP2 α	+	PAP	Transcrits de réponse au stress		(Mellman et al., 2008)	

INTRODUCTION

PSF	+		Cyclooxygenase-2 (COX-2)	USE	(Hall-Pogar et al., 2007)
			Mécanisme global (Gal5HIV2dsx)		(Rosonina et al., 2005)
			SV40L		(Liang and Lutz, 2006)
Sam68	+		HIV-1		(McLaren et al., 2004)
Slm1, Slm2	+		HIV-1		(McLaren and Cochrane, 2009)
Smc1	+	CPSF	Pré-ARNm spécifiques (Xiro1)		(Collart et al., 2009)
SR proteins	+		Gfap		(Blechingberg et al., 2007)
		CFIm	Rous Sarcoma Virus	NRS	(Maciolek and McNally, 2007)
SRm160	+	CPSF	Mécanisme global (β -globine)		(McCracken et al., 2002)
SRp20	+	CstF64	CT/CGRP	Sites d'épissage et éléments riches en pyrimidines localisés en aval du site de clivage	(Lou et al., 1996; Lou et al., 1998)
SRp75	-	PAP	Pré-ARNm spécifiques (AdL3)	Site artificiel de liaison	(Ko and Gunderson, 2002)
U1 snRNP	+	CstF64	CT/CGRP	Sites d'épissage et éléments riches en pyrimidines localisés en aval du site de clivage	(Lou et al., 1998)
	-	PAP	HIV-1	Site 5' d'épissage en aval du site de clivage	(Ashe et al., 1997)
			BPV (« Bovin Papillomavirus »)	Site 5' d'épissage en amont du site de clivage	(Vagner et al., 2000a)
U1A snRNP	+	CstF64	SV40L	USE	(Lutz and Alwine, 1994)
			Mécanisme global (SV40L)	USE	(Lutz et al., 1996)
	-	CstF64	IgM	Proche du DSE	(Phillips et al., 2004)

INTRODUCTION

		PAP	Cyclooxygenase-2	USE	(Hall-Pogar et al., 2007)
			U1A	Structure en tige boucle en amont du site de clivage	(Gunderson et al., 1994; Gunderson et al., 1997)
U2 snRNP	+	CPSF	Mécanisme global (SV40L)	Point de branchement en amont du site 3' d'épissage	(Kyburz et al., 2006)
U2AF35 U2AF65	+		Pré-ARNm spécifiques (Prothrombine F2)	USE	(Danckwardt et al., 2007)
U2AF65	+	CFIm	Mécanisme global (β -globine et AdL3)	Pyrimidines adjacentes en amont du site 3' d'épissage	(Millevoi et al., 2002; Millevoi et al., 2006)
	-		Pré-ARNm spécifiques (albumin)	Poly(A)-limiting element (PLE)	(Gu and Schoenberg, 2003)
		PAP	Pré-ARNm spécifiques (AdL3)	Site artificiel de liaison	(Ko and Gunderson, 2002)
τ cstf64			Transcrits des testicules de souris		(Dass et al., 2007)

IV-3 Mécanismes de régulation de la maturation en 3'

La maturation en 3' est une étape post-transcriptionnelle complexe régulée de différentes manières. Une revue récente de l'équipe (Millevoi and Vagner, 2010) a proposé différents mécanismes de régulation de la maturation en 3' :

-La compétition pour la reconnaissance du signal de polyadénylation :

Nous avons vu précédemment que la reconnaissance du signal de polyadénylation par CPSF et CstF permet l'assemblage du complexe de maturation en 3'. Un des mécanismes de répression de la maturation en 3' consiste à inhiber cette reconnaissance du signal de polyadénylation en recrutant des protéines de liaison aux ARN qui interagissent avec ces mêmes séquences. Dans de nombreux cas, des facteurs de régulation bloquent l'interaction entre CstF et le DSE riche en U ou GU. Par exemple, il a été montré que la PTB (« poly-pyrimidine tract binding protein ») interagit avec le DSE du pré-ARNm du Complément C2 ou de la β -globine bloquant alors l'interaction de CstF avec le pré-ARNm et inhibe la maturation en 3'. (Castelo-Branco et al., 2004).

-L'inhibition de la PAP par interaction avec des protéines de liaison aux ARN : Un autre mécanisme de répression de la maturation en 3' consiste au blocage de l'addition de la queue polyadénylée. Des protéines de liaison aux ARN se fixent à proximité du signal de polyadénylation et interagissent avec la PAP entraînant alors l'inhibition de son activité de polyadénylation. Nous pouvons par exemple citer, l'autorégulation de la sous unité U1A de la U1snRNP. Cette protéine se lie à son propre pré-ARNm de manière spécifique et interagit avec la partie C-terminale de la PAP bloquant l'addition de la queue poly(A) (Boelens et al., 1993; Gunderson et al., 1994; Gunderson et al., 1997; Varani et al., 2000).

-La compétition cinétique avec les autres étapes de la synthèse des ARNm : Ce mécanisme fait intervenir des protéines ou des éléments de

INTRODUCTION

séquence qui, en modifiant la vitesse d'épissage ou d'élongation de la transcription, affecte l'efficacité de maturation en 3'. Inversement, des facteurs modifiant l'efficacité de maturation 3' peuvent influencer l'épissage ou l'élongation de la transcription. Notamment, le régulateur de la maturation des pré-ARNm spécifiques des neurones de Drosophile, ELAV (« Embryonic Lethal Abnormal Vision ») interagit avec le signal de polyadénylation du gène *ewg* (« erect wing »). ELAV ralentit le recrutement de CFIm, de CFIIm ou de la PAP favorisant l'épissage de cette région contenant le signal de polyadénylation (Soller and White, 2003).

-Le recrutement des facteurs de maturation en 3' : Un mécanisme de stimulation de la maturation en 3' fait intervenir des facteurs qui reconnaissent des éléments de séquence du pré-ARNm et recrutent la machinerie de maturation en 3' de manière directe ou indirecte. On peut citer PTB qui reconnaît les USE de plusieurs pré-ARNm et stimule la maturation en 3' (Hall-Pogar et al., 2007; Moreira et al., 1998) en recrutant un autre facteur : hnRNP H (Millevoi et al., 2009) qui à son tour recrute CstF (Bagga et al., 1998) ou la PAP (Millevoi et al., 2009).

-La redistribution des facteurs de maturation en 3' dans le cytoplasme : Il existe une régulation de la répartition sub-cellulaire de la machinerie de maturation en 3'. En effet, certains facteurs inhibent l'efficacité de la maturation en 3' en redistribuant au niveau du cytoplasme la machinerie de polyadénylation. C'est notamment le cas de la protéine de liaison à la phosphoserine 14-3-3 ϵ qui interagit de manière directe avec la partie C-terminale de la PAP, lorsqu'elle est phosphorylée, entraînant une relocalisation cytoplasmique de la PAP ainsi qu'une inhibition de l'activité de polyadénylation (Kim et al., 2003).

-La redistribution des facteurs de maturation en 3' dans un complexe nucléaire : A l'inverse, d'autres protéines redistribuent les facteurs de maturation en 3' dans des complexes protéiques différents, à l'intérieur du noyau. Cette redistribution peut stimuler ou inhiber l'efficacité de maturation en 3'. Par exemple, dans des cellules soumises à un choc thermique le facteur

INTRODUCTION

de transcription HSF1 (« heat shock transcription factor 1 ») forme un complexe avec la Symplekine et CstF 64 augmentant ainsi l'efficacité de polyadénylation de l'ARNm du pré-ARNm Hsp70 (« heat shock protein ») (Xing et al., 2004)

-Modification post-traductionnelle des facteurs et des régulateurs de la maturation en 3' : Des modifications post-traductionnelles contribuent à modifier l'activité, la répartition nucléo-cytoplasmique, la capacité à interagir avec la machinerie de maturation en 3' et la stabilité des facteurs et des régulateurs de la polyadénylation. Ces modifications peuvent donc moduler l'efficacité de maturation en 3' de différentes manières. Notamment, Il a été montré que la sumoylation de la PAP augmente sa localisation nucléaire et sa stabilité mais diminue son activité enzymatique (Vethantham et al., 2008).

Ces différents mécanismes sont représentés par de nombreux exemples (Millevoi and Vagner, 2010) et permettent d'englober l'ensemble des mécanismes de régulation de la maturation en 3'.

IV-4 Exemples de régulations physiologiques de la maturation en 3'

IV-4-1 La différenciation cellulaire

IV-4-1-1 Exemple de l'immunoglobuline M

Le gène des chaînes lourdes de l'IgM (immunoglobuline M) contient deux sites de polyadénylation alternatifs contenus sur deux exons distincts. Un exon composite interne proximal, appelé C μ 4 et un exon distal M2.

La maturation en 3' de ce pré-ARNm évolue au cours de la différenciation des lymphocytes B. En effet, l'exon C μ 4 est utilisé en tant qu'exon interne dans les lymphocytes immatures, alors qu'il est l'exon 3' terminal dans les cellules B matures. Lorsque l'exon C μ 4S est utilisé comme exon 3' terminal, cela entraîne la formation d'un ARNm qui code pour la forme sécrétée des IgM alors que l'ARNm contenant l'exon M2 terminal code pour un récepteur membranaire. La polyadénylation alternative de ce pré-ARNm permet donc de contrôler la sécrétion ou l'expression membranaire de la protéine qui en résulte.

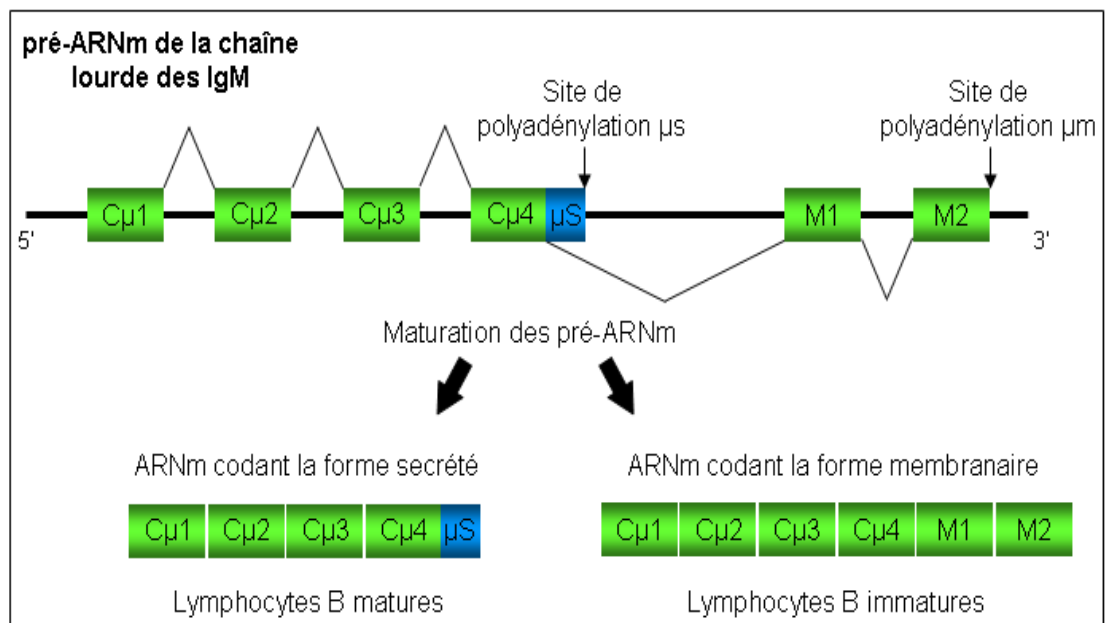


Figure 10 : Polyadénylation alternative du pré-ARNm des chaînes lourdes de l'IgM.

Les exons sont représentés en rectangle de couleur et les introns en trait fin. La maturation en 3' alternative du pré-ARNm de la chaîne lourde des IgM entraîne, en fonction de la différenciation des lymphocytes B, la formation de deux ARNm différents codant pour des protéines sécrétées ou membranaires. Cette figure s'inspire de (Phillips et al., 2001)

INTRODUCTION

De nombreuses études montrent qu'une régulation complexe s'exerce au niveau du site de polyadénylation μ S de l'exon C μ 4S des chaînes lourdes des IgM. Les faibles signaux de polyadénylation μ S sont sensibles à la concentration du facteur CstF64. En effet, la surexpression de ce facteur peut activer l'utilisation du site de polyadénylation μ S dans des cellules en culture (Takagaki and Manley, 1998; Takagaki et al., 1996). La concentration de CstF64 augmente au cours de la différenciation des lymphocytes B, mais cette variation n'est pas corrélée à la modification de la maturation en 3' du pré-ARNm des IgM (Martincic et al., 1998). En fait, le choix du site de polyadénylation μ S résulterait d'une modification de l'affinité de la liaison du facteur CstF64 plus que d'une modification de la concentration intracellulaire de ce facteur (Edwalds-Gilbert and Milcarek, 1995).

La sélection du signal μ S pourrait également être contrôlée par la sous unité U1A de la snRNP U1. Plusieurs expériences ont montré que U1A se fixe sur trois séquences régulatrices situées en amont du signal μ S et inhibe la réaction de polyadénylation d'un pré-ARNm clivé au site μ S, en affectant l'activité de la PAP (Phillips et al., 2001). Un autre site de liaison de la protéine U1A a été identifié en aval du signal μ S, entre deux DSE riches en guanines et en uraciles. L'interaction de la protéine U1A avec ce site, diminue la liaison du facteur CstF64 au niveau de ces DSE, et inhibe la réaction de clivage au site μ S (Phillips et al., 2004). D'autres expériences ont au contraire montré que des mutations des séquences reconnues par U1A entourant l'hexanucléotide AAUAAA du site μ S, ne modifient pas la maturation en 3' du gène des chaînes lourdes de l'IgM (Peterson et al., 2006). De plus, la régulation de l'utilisation du signal μ s n'est pas modifiée dans des cellules surexprimant le facteur U1A (Peterson et al., 2006).

Enfin, il a été proposé que la hnRNP F (« Heterogeneous nuclear ribonucleoproteins ») dont la quantité varie lors de la différenciation cellulaire des lymphocytes B, permet de réguler la maturation en 3' des chaînes lourdes de l'IgM. HnRNP F reconnaît les DSE en aval du signal μ S et bloque le recrutement de CstF-64 inhibant la réaction de clivage sur ce site de polyadénylation (Veraldi et al., 2001).

IV-4-1-2 Exemple de la calcitonine

Le pré-ARNm de la CT/CGRP (« calcitonin/calcitonin gene-related peptide ») contient deux sites de polyadénylation alternatifs sur deux exons distincts. La sélection du dernier exon se fait en fonction du type cellulaire. En effet, les lymphocytes T de la thyroïde utilisent l'exon 4 comme exon 3' terminal. En revanche, dans les cellules neuronales, cet exon est excisé, l'exon 6 est alors utilisé comme exon 3' terminal. Un ARNm ayant pour exon terminal l'exon 4 code pour la calcitonine alors que lorsque le dernier exon est l'exon 6, l'ARNm code pour le CGRP.

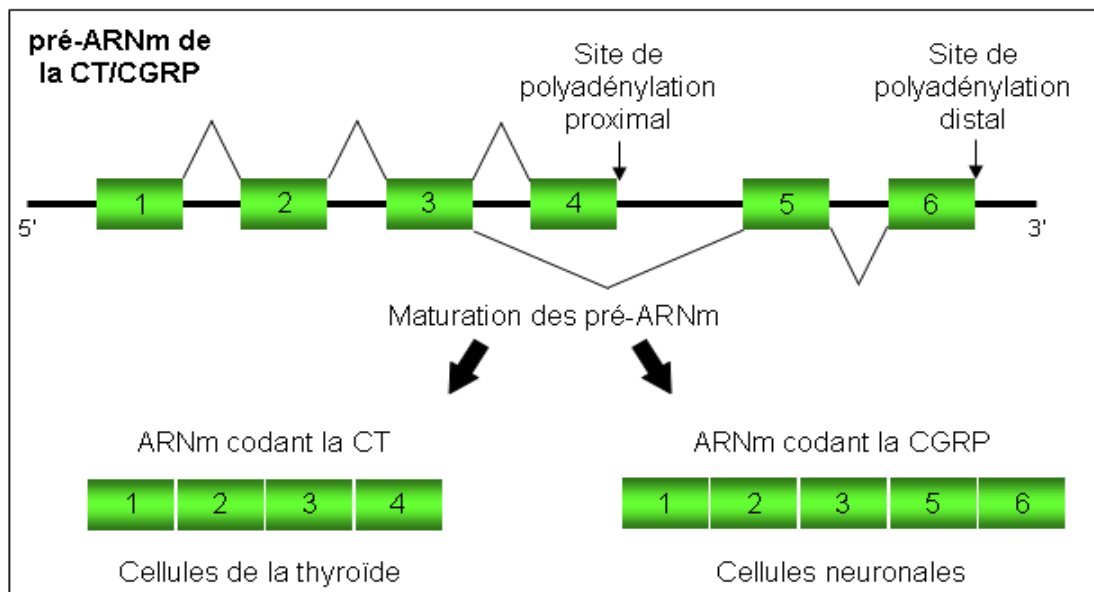


Figure 11 : Polyadénylation alternative du pré-ARNm de la CT/CGRP.

Les exons sont représentés en rectangle vert et les introns sont en trait fin. La maturation en 3' alternative du pré-ARNm de la CT/CGRP entraîne, en fonction du type cellulaire, la formation de deux ARNm différents codant pour la CT ou la CGRP.

L'exon 4 est délimité par un site 3' d'épissage et des signaux de polyadénylation faibles. L'inclusion de cet exon entraînant une maturation en 3' proximale est dépendante de la reconnaissance d'un élément intronique activateur ISE (« Intron Splice Enhancers »). Cet ISE localisé en aval du site de clivage de l'exon 4, dans l'intron 4, est constitué d'un pseudo-site 5' d'épissage encadré par une séquence riche en pyrimidines en amont et une répétition d'uraciles en aval (Lou et al., 1995). La mutation d'un de ces trois éléments entraîne l'exclusion de l'exon 4. Des expériences de clivage en 3' *in vitro*, montrent que le pseudo-site 5' d'épissage et la

INTRODUCTION

région riche en pyrimidines stimulent le clivage en 3' de l'exon 4. D'autre part, la répétition d'uraciles contrôlerait la reconnaissance du site 3' d'épissage localisé dans l'intron 3 (Lou et al., 1996; Zhu et al., 2003). Une autre séquence régulatrice permettant l'utilisation ce site 3' d'épissage a également été caractérisée dans l'exon 4. Ces séquences activatrices sont reconnues par un grand nombre de facteurs qui stimulent la rétention de l'exon 4 et la maturation en 3'.

La protéine SRp20 (protéine 20 riche en sérine et en arginine) interagit *in vitro* avec la séquence riche en pyrimidines située dans l'intron 4. La surexpression de SRp20 favorise la polyadénylation au niveau de l'exon 4 en recrutant le complexe de maturation en 3'. La surexpression d'une protéine SRp20 tronquée ne possédant plus de domaine RS (riche en arginines et en sérines) cause une baisse du recrutement du facteur CstF64 au niveau du signal de polyadénylation de l'exon 4 (Lou et al., 1998).

La PTB interagit avec différentes séquences localisées en aval du signal de polyadénylation de l'exon 4 dont la région riche en pyrimidines. Il a été montré que le recrutement de PTB empêche la fixation du facteur d'épissage U2AF (« U2 auxiliary factor ») ce qui favorise la polyadénylation au niveau de l'exon 4 (Lou et al., 1996; Lou et al., 1999). De plus, PTB peut stimuler la maturation en 3' en interagissant avec le site de polyadénylation (Lou et al., 1999).

La protéine TIAR (« T-cell-restricted Intracellular Antigen1-Related protein ») interagit avec la séquence riche en uraciles et stimule la fixation de la snRNP U6 sur le pseudo-site 5' d'épissage. Inversement, le recrutement de la snRNP U6 sur le pseudo-site 5' d'épissage active l'interaction entre TIAR et la séquence riche en uraciles. Ces deux facteurs activent la maturation en 3' au niveau du site proximal (Zhu et al., 2003). La snRNP U1 et le facteur ASF/SF2 se fixent également sur le pseudo-site 5' d'épissage. La snRNP U1 participerait à l'activation du clivage en 3' de l'exon 4 (Lou et al., 1996), et le rôle de ASF/SF2 (alternative splicing factor/splicing factor 2) n'est pas clairement établi.

Les protéines hTraβ2 (« human transformer β2 ») et la SRp55 interagissent avec la séquence activatrice localisée dans l'exon 4 et favorisent la reconnaissance du site 3' d'épissage (Tran et al., 2003; Tran and Roesser, 2003). Le recrutement de ces facteurs permet l'utilisation de l'exon 4 dans les cellules non neuronales.

INTRODUCTION

D'autres protéines au contraire inhibent l'épissage de l'exon 4 et favorisent une maturation en 3' au niveau de l'exon 6. Ainsi, il a été montré que la déplétion des protéines Fox (« Forkhead box ») 1 et 2 dans des cellules, favorise l'utilisation de l'exon 4. Ces protéines interagissent *in vitro* avec les motifs UGCAUG présents en aval du site 3' d'épissage dans l'intron 3d et dans l'exon 4. La fixation des protéines Fox sur cet élément régulateur bipartite, empêche le recrutement du facteur U2AF 65 au niveau du site 3' d'épissage (Zhou et al., 2007; Zhou and Lou, 2008).

Les protéines Hu, spécifiques des cellules neuronales, entrent en compétition avec la protéine TIAR pour la fixation sur la séquence intronique riche en uraciles de l'intron 4 (Zhu et al., 2006). L'expression de dominants négatifs des protéines Hu dans des cellules, conduit à l'inclusion de l'exon 4. Le saut de l'exon 4 dans les cellules neuronales est donc dû à la compétition de fixation entre les facteurs activateurs et les protéines Fox et Hu.

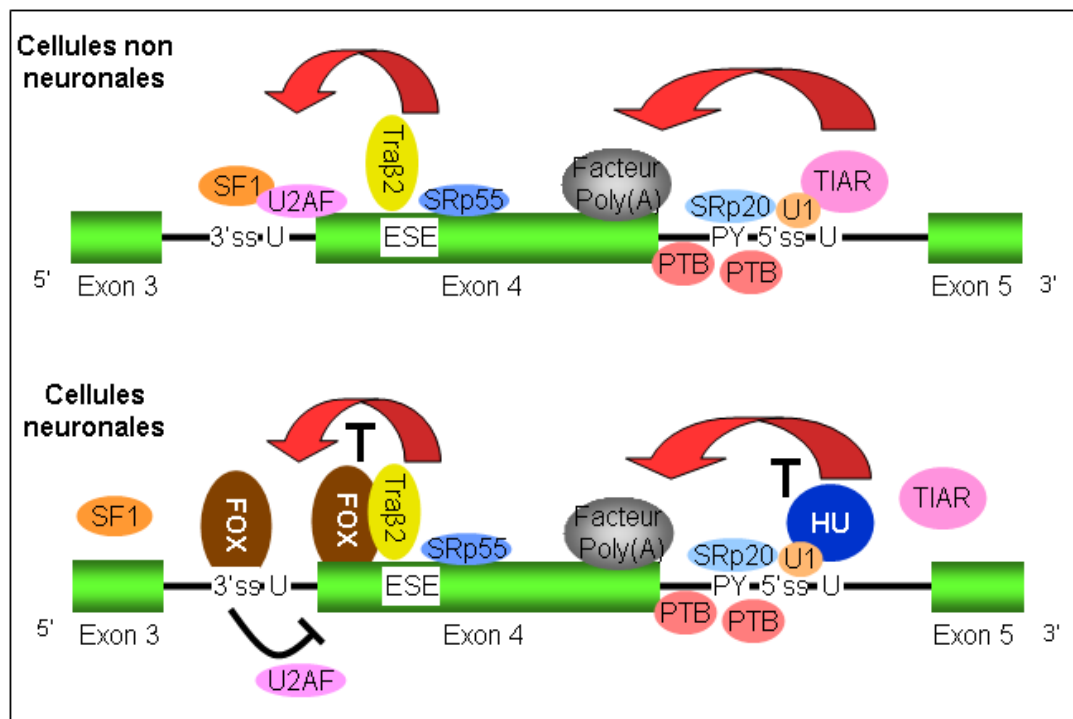


Figure 12 : Régulation de la maturation en 3' de l'exon 4 du pré-ARNm de CT/CGRP en fonction du type cellulaire.

Dans les cellules non neuronales, de nombreux facteurs *trans* permettent l'inclusion de l'exon 4 de CT/CGRP. Dans les cellules neuronales, l'exclusion de l'exon 4 se fait par les protéines Fox et Hu. PY correspond à la région riche en pyrimidines, U1 représente la snRNP U1 et Facteur Poly(A) représente la machinerie de maturation en 3'. D'après (Zhou et al., 2007)

IV-4-3 Les dommages à l'ADN

Lors de dommages à l'ADN, de nombreux mécanismes de réponse à ce stress se mettent en place, parmi lesquels une inhibition globale de la transcription et de la traduction entraînant une diminution de l'expression des gènes. Il a récemment été montré que la maturation en 3' participe de manière importante à la répression de l'expression génique lors de ce stress. En effet, les protéines impliquées dans la réponse aux dommages à l'ADN, BRCA1 (« Breast cancer 1 ») et son partenaire, la protéine BARD1 (« BRCA1-associated RING domain protein 1 ») (Jin et al., 1997; Scully et al., 1997) co-immunoprécipitent avec CstF-64 suggérant un lien entre la maturation en 3' et les dommages à l'ADN (Kleiman and Manley, 1999). De plus, l'ajout de protéine BARD1 recombinante dans des extraits nucléaires inhibent la réaction *in vitro* de clivage en 3' (Kleiman and Manley, 1999). La formation du complexe BARD1/BRCA1/CstF nécessite un lien entre les domaines ankyrine et BRCT (« BRCA1 C-terminal repeats ») de BARD1. Cette interaction entre ces deux domaines de BARD1 permet la reconnaissance du domaine WD-40 de CstF-50 entraînant une inhibition de la maturation en 3' (Edwards et al., 2008).

Lors de dommages à l'ADN, BRCA1 et BARD1 sont phosphorylées. La protéine kinase ATM (« Ataxia Telangiectasia Mutated ») phosphoryle BRCA1 (Thomas et al., 1997), alors que la phosphorylation de BARD1 se fait par la kinase CHK2 (« Checkpoint 2 ») et nécessite la présence de BRCA1 (Kim et al., 2006). La phosphorylation de BARD1 en réponse aux dommages à l'ADN est indispensable à la formation du complexe BARD1/BRCA1/CstF et à l'inhibition de la maturation en 3' (Kim et al., 2006; Kleiman and Manley, 2001). Le pré-ARNm non clivé est alors dégradé par l'exosome (Mirkin et al., 2008). Plus récemment, il a été montré que la protéine P53 co-immunoprécipite avec le complexe BARD1/BRCA1/CstF. De plus, une mutation de p53 associée à des tumeurs atténue l'interaction entre BARD1 et CstF et diminue l'inhibition de la maturation en 3' due aux dommages à l'ADN (Nazeer et al., 2011).

Il a également été montré que lors de dommages à l'ADN, BARD1 et BRCA1 interagissent avec la forme phosphorylée sur la serine 2 et la serine 5 de la CTD pol II qui intervient lors de l'élongation de la transcription (Kleiman et al., 2005). Cette interaction entraîne l'ubiquitination de la pol II et permet d'initier sa dégradation par le protéasome (Kleiman et al., 2005).

Enfin, le facteur de maturation en 3' CstF joue un rôle important dans la réponse aux dommages à l'ADN. La diminution de son expression diminue la viabilité des cellules traitées aux UV, réduit l'ubiquitination de la pol II et diminue la réparation des dommages à l'ADN (Mirkin et al., 2008). De plus BARD1, CstF et la pol II sont localisés sur les sites de réparation de l'ADN (Mirkin et al., 2008).

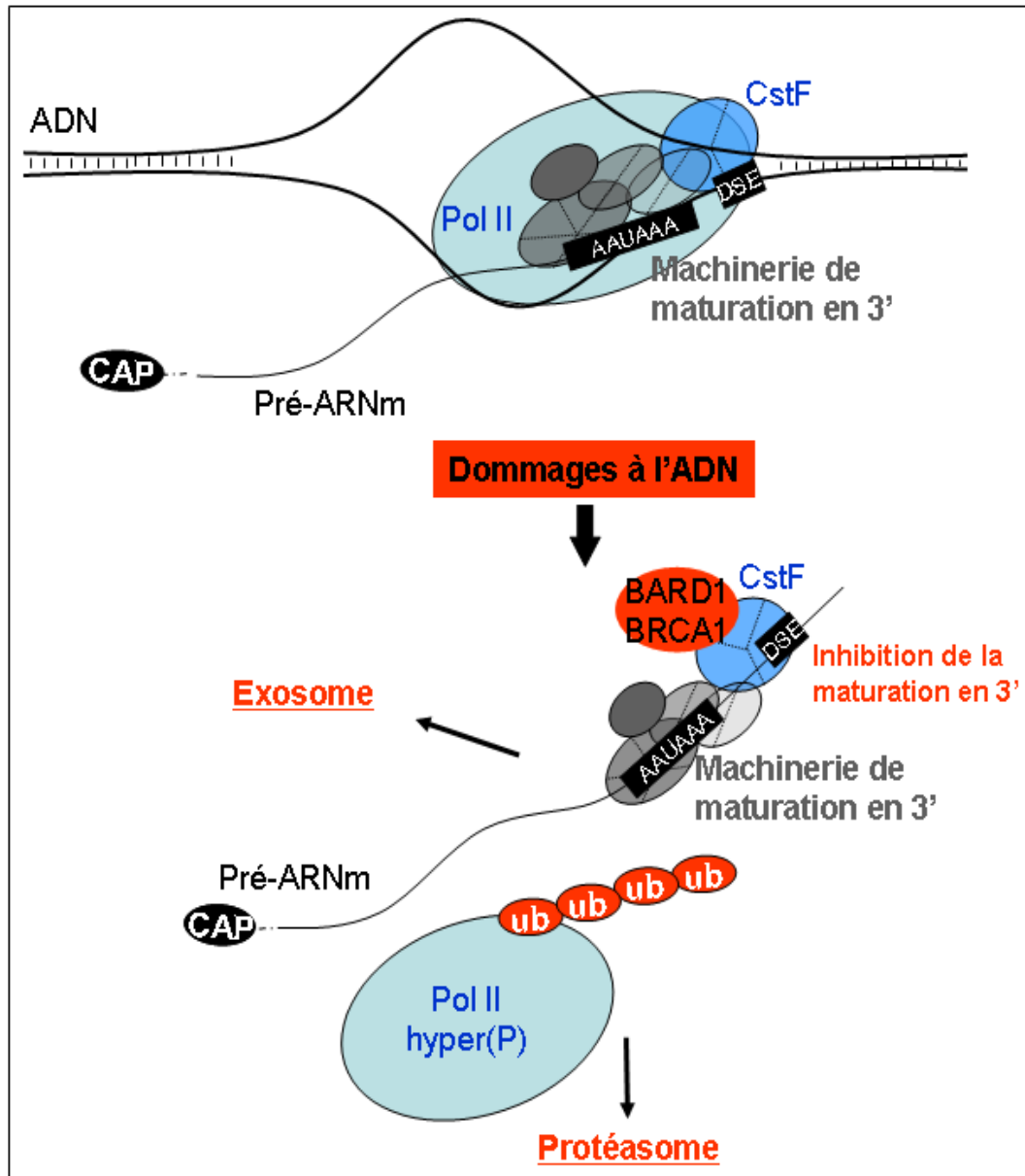


Figure 13 : Inhibition de la maturation en 3' lors de dommages à l'ADN.

Lors de dommages à l'ADN, le complexe BRCA1/BARD1 interagit avec CstF et inhibe la maturation en 3'. Le pré-ARNm non clivé est alors dégradé par l'exosome et la pol II est ubiquitinée et dégradée par le protéasome. Pol II hyper(P) représente la RNA polymérase II phosphorylée sur la serine 2 et 5, ub représente l'ubiquitine et CAP représente la coiffe. D'après (Mirkin et al., 2008).

INTRODUCTION

En conclusion, lors de dommages à l'ADN, BRCA1 et BARD1 sont phosphorylées permettant ainsi la formation d'un complexe BARD1/BRCA1/CstF qui interagit avec la pol II. La formation de ce complexe entraîne une inhibition globale de la maturation en 3' ainsi que la dégradation des pré-ARNm et de la pol II phosphorylée.

V Dérégulation de la maturation en 3' associée à des pathologies

V-1 Les maladies sanguines

V-1-1 Les thalassémies

Les thalassémies sont des maladies génétiques causant des formes d'anémies associées à une hémoglobinopathie (déficience dans la synthèse de l'hémoglobine). Il existe 2 groupes de thalassémies : les alpha-thalassémies se caractérisent par un déficit de synthèse de l' α -globine et les bêta-thalassémies se caractérisent par un déficit de synthèse de la β -globine.

Ces pathologies peuvent résulter d'un défaut de maturation en 3' des globines. Plusieurs études ayant analysé différentes mutations de patients atteints de thalassémie ont révélé, dans certains cas, des altérations de l'hexamère AAUAAA dans le gène de l' α -globine (Harteveld et al., 1994; Higgs et al., 1983) et dans celui de la β -globine (Jankovic et al., 1990; Losekoot et al., 1991; Orkin et al., 1985; Rund et al., 1992; van Solinge et al., 1996). Ces mutations sont variées et touchent différents nucléotides du signal de polyadénylation. Par exemple la séquence AATAAA de l' α -globine peut être mutée en AATAAG (Higgs et al., 1983) ou délétée de ses deux derniers nucléotides pour devenir AAUA (Harteveld et al., 1994). Plus récemment, nous avons contribué à montrer qu'une mutation (C→G) dans la région correspondant au 3'UTR du gène de β -globine, six nucléotides en aval du codon stop entraîne une diminution de l'efficacité de maturation en 3' responsable de bêta-thalassémies (Millevoi et al., 2009). En effet, cette mutation altère la liaison entre PTB et le pré-ARNm de la β -globine. Lorsque PTB interagit avec le pré-ARNm de la β -globine, elle stimule la maturation en 3' en recrutant hnRNP H qui à son tour recrute CstF (Bagga et al., 1998) ou la PAP (Millevoi et al., 2009). En présence de la mutation, PTB ne liant plus le pré-ARNm de la β -globine ne joue plus son rôle de stimulateur de la maturation en 3'.

De manière générale, ces mutations entraînent dans tout les cas une diminution de la maturation en 3' des globines à l'origine de la thalassémie.

V-1-2 Les thrombophilies

La thrombophilie est une augmentation de la vitesse de coagulation du sang entraînant un risque accru de thrombose. Une thrombose consiste en la formation d'un thrombus ou caillot sanguin obturant un vaisseau sanguin. Des dérégulations de la maturation en 3' de plusieurs gènes peuvent être à l'origine de cette pathologie.

En effet, le gène $FG\gamma$ -H2 (« Fibrinogen gamma Haplotype 2 ») possède deux sites de polyadénylation sur deux exons distincts entraînant la formation de 2 transcrits : le transcrit utilisant le site de polyadénylation proximal est appelé γ' et le transcrit utilisant le site de polyadénylation distal est appelé γA . Un ratio de la quantité d'ARNm de la forme γ' sur la forme γA faible est révélateur d'une thrombophilie. Des SNP (« single nucleotide polymorphism ») (9615C>T et 10034C>T) au niveau des DSE de ces deux sites de polyadénylation ont été identifiés (Uitte de Willige et al., 2007). La SNP 9615C>T semble diminuer la quantité de transcrit de la forme γ' alors que la SNP 10034C>T du DSE de la forme γA augmente la quantité de transcrit de cette forme. Il semblerait donc que ces SNP régulent la maturation en 3' de $FG\gamma$ -H2 et dans les 2 cas entraînent une diminution du ration γ' sur γA (Uitte de Willige et al., 2005; Uitte de Willige et al., 2007)

Cette pathologie peut aussi être la conséquence d'une dérégulation de la maturation en 3' du pré-ARNm de la prothrombine aussi appelé F2 (« coagulation factor II »). Il est connu que la maturation en 3' de F2 est stimulée lors de stress inflammatoire. En condition normale, les ARE-BP (« AU-rich Element sequence Binding Protein ») FBP2 et FBP3 (« far upstream element (FUSE) binding protein 2 et 3 ») lient l'USE de la prothrombine F2 et empêche le recrutement de U2AF ou de PTB stimulant alors la maturation en 3'. Lors de stress inflammatoire, la MAPK p38 phosphoryle FBP2 et FBP3 empêchant leur interaction avec l'USE. Cette phosphorylation permet à U2AF et PTB de lier l'USE et donc de stimuler la maturation en 3' (Danckwardt et al., 2011). Différentes mutations responsables de l'augmentation de l'efficacité de maturation en 3' de F2 ont été identifiées chez des

INTRODUCTION

patients atteints de thrombophilie. Une insertion de 11 thymines au niveau du DSE du gène F2 a été retrouvée chez un patient atteint du Syndrome de Budd-Chiari (Danckwardt et al., 2004; Danckwardt et al., 2006a) qui correspond à une maladie du foie résultant de l'obstruction d'une veine sus-hépatique généralement provoquée par une thrombose. Le DSE du pré-ARNm de F2 contient peu d'uridine et l'efficacité de maturation en 3' de ce pré-ARNm est physiologiquement faible. L'addition de 11 résidus de thymines au niveau de l'ADN provoque sur le pré-ARNm un enrichissement en uridines en aval du site de clivage, au niveau du site potentiel de CstF sur le pré-ARNm. Cet enrichissement entraîne une augmentation de l'efficacité de maturation en 3' responsable d'un accroissement pathologique de la protéine F2. L'augmentation de la quantité de F2 dans le sang est à l'origine de la thrombophilie de ce patient (Danckwardt et al., 2004; Danckwardt et al., 2006a).

D'autres mutations responsables de cette pathologie ont été identifiées au niveau du site de clivage du pré-ARNm de F2. Le site de clivage physiologique de F2 est composé d'une cytosine suivie d'une guanine alors que la plupart des sites de clivage sont composés d'une cytosine et d'une adénine. Il a été montré par des réactions de clivage *in vitro*, qu'un site de clivage cytosine guanine est moins efficace qu'un site cytosine adénine (Chen et al., 1995).

Une première mutation transformant le site de clivage de F2 en cytosine adénine a été identifiée (Gehring et al., 2001) causant une augmentation de la concentration en F2 dans le plasma. En effet, la mutation de la guanine en adénine augmente la faible efficacité de maturation en 3' de F2 ce qui accroît d'approximativement deux fois la quantité d'ARNm et de la protéine F2 (Danckwardt et al., 2004; Gehring et al., 2001). L'augmentation de l'efficacité de maturation en 3' de F2 due à cette mutation explique l'augmentation du risque de thrombophilie.

Une autre mutation transformant le site de clivage de F2 (CG→TG) a été identifiée (Balim et al., 2003; Danckwardt et al., 2004; Schrijver et al., 2003). Là encore, cette mutation augmente l'efficacité de maturation en 3' de F2 et donc le risque de thrombophilie (Danckwardt et al., 2004; Danckwardt et al., 2006b).

V-1-3 L'hyperéosinophilie idiopathique

L'hyperéosinophilie idiopathique se caractérise par une surproduction des éosinophiles polynucléaires dans la moelle osseuse conduisant à l'accumulation de ces cellules dans le sang. Cette accumulation occasionne une infiltration dans les tissus et endommagement des organes tel que le coeur, la rate ou le foie.

Il a été montré que l'une des causes de cette pathologie peut être un réarrangement provoqué par une délétion de 800 kb créant la fusion des gènes de la sous unité hFip1 de CPSF et de PDGFR α (« Platelet-Derived Growth Factor Receptor, alpha »). La protéine issue de cette fusion comprend 233 acides aminés de hFip1 dans sa partie N-terminale et 523 acides aminés de PDGFR α dans sa partie C-terminale (Cools et al., 2003). La protéine de fusion hFip1-PDGFR α se comporte comme un mutant constitutivement actif et ne nécessite plus la dimérisation dépendante des facteurs de croissance du récepteur à la tyrosine kinase PDGFR α . Ce mutant constitutif est à l'origine du syndrome d'hyperéosinophilie idiopathique.

La formation de la protéine de fusion entraîne la perte fonctionnelle de hFip1 ce qui stimule la maturation en 3' dans des conditions normales (Kaufmann et al., 2004). La stimulation de la maturation en 3' pourrait également expliquer en partie le phénotype (Danckwardt et al., 2008).

V-2 Les infections virales

V-2-1 Le virus de l'immunodéficience humaine de type 1

Le génome du VIH-1 est composé de neuf gènes dont trois principaux qui définissent la structure du virus et six autres gènes qui codent pour des protéines régulatrices. Parmi ces protéines régulatrices, la Tat (« Trans-Activator of Transcription ») est essentielle pour l'expression des gènes et la réplication virale du VIH-1. Il a été montré que la Tat augmente spécifiquement l'expression de la sous-unité de 73-kDa de CPSF (Calzado et al., 2004). L'augmentation de CPSF-73 par la Tat du VIH-1 permet à ce virus de stimuler la maturation en 3' provoquant alors une augmentation de l'expression des gènes cellulaires et viraux. En plus de la fonction de CPSF-73 sur la maturation en 3', il a été montré que dans ce cas, CPSF-73 réprime le promoteur du gène MDM2 (Murine Double Minute 2) (Calzado et al., 2004). La surexpression de CPSF-73 par la Tat est donc un mécanisme qui permet au virus d'augmenter l'expression de ses gènes en dérégulant la machinerie cellulaire.

V-2-2 L'Herpès simplex virus 1

Lors de l'infection par l'Herpès Simplex Virus 1 (HSV-1), la maturation en 3' de certains pré-ARNm viraux est stimulée (McLauchlan et al., 1989). Cette régulation spécifique de la maturation en 3' nécessite l'expression du gène viral précoce IE63 aussi appelé ICP27 (McLauchlan et al., 1992; Sandri-Goldin and Mendoza, 1992). IE63 peut également augmenter ou diminuer l'expression de différents gènes en fonction des signaux de maturation contenus sur le pré-ARNm. L'activation de l'expression des gènes par IE63 est corrélée avec la présence d'un hexamère AAUAAA sans DSE riche en uraciles ou en guanines et uraciles. L'inhibition de l'expression d'un gène est due à la présence d'introns dans son pré-ARNm (Sandri-Goldin and Mendoza, 1992). Le mécanisme de cette régulation post-transcriptionnelle est dû à l'augmentation du recrutement de CstF-64 au niveau des DSE de l'ensemble des pré-ARNm et à une redistribution éloignée des sites de transcription du facteur d'épissage snRNP U2 lors de l'expression de IE63 (McGregor et al., 1996). IE63

permet donc d'activer la maturation en 3' des pré-ARNm viraux ayant un signal de polyadénylation faible et inhibe l'épissage des pré-ARNm de la cellule hôte.

V-2-3 L'influenza A

Le virus grippale influenza A fournit un autre exemple de régulation de la maturation en 3' lors d'infections virales. La protéine NS1 (Non Structurale 1) du virus influenza A représente l'une des protéines les plus abondantes synthétisées dans les cellules infectées (Lazarowitz et al., 1971). NS1 interagit de manière fonctionnelle avec CPSF-30 empêchant l'interaction entre CPSF et le pré-ARNm substrat. Cela induit une inhibition de la maturation en 3' des pré-ARNm de la cellule hôte (Nemeroff et al., 1998). De plus, la protéine NS1 cible également la PABP, inhibant la synthèse de la queue poly(A) (Chen et al., 1999). La maturation en 3' des pré-ARNm représente une condition indispensable à l'export cytoplasmique (Stutz and Rosbash, 1998). Les pré-ARNm non clivés de la cellule hôte restent alors dans le noyau permettant à la protéine NS1 d'inhiber sélectivement la néosynthèse des ARNm cellulaires, et non viraux. (Chen et al., 1999; Nemeroff et al., 1998). En effet, la maturation en 3' des pré-ARNm viraux est réalisée par l'ARN transcriptase virale, qui retranscrit une séquence de 5 à 7 uraciles formant une queue poly(A) qui permet l'export vers le cytoplasme des ARNm.

En conséquence, en interférant avec l'activité essentielle des facteurs de maturation en 3', l'influenza A utilise un moyen efficace pour éteindre spécifiquement l'expression des gènes cellulaires et faciliter l'expression des gènes viraux qui ne dépendent pas de la machinerie de maturation en 3' cellulaire (Chen et al., 1999; Nemeroff et al., 1998).

V-3 Les cancers

Les dérégulations de la maturation en 3' dans les cancers peuvent s'expliquer de différentes manières. En effet, des mutations au niveau des séquences de polyadénylation des pré-ARNm peuvent être la cause de cette pathologie mais également une expression aberrante des facteurs protéiques de la maturation en 3'.

V-3-1 Polymorphisme de la N-acetyltransferase 1

L'enzyme N-acetyltransferase 1 (NAT1) peut N-acétyler ou O-acétyler les amines aromatiques (Hein et al., 1993; Probst et al., 1992). Les amines aromatiques (arylamines et hétérocycliques amines) sont en partie responsables d'un risque accru de cancers colorectaux liés aux agents cancérigènes présents dans l'alimentation, la fumée de cigarette, ou l'environnement (Giovannucci et al., 1994; Potter et al., 1993). La N-acétylation des arylamines est considérée de façon générale comme exerçant une détoxification relative puisque cette réaction les rend en effet moins actives, alors que la O-acétylation serait plutôt activatrice.

Au sein de la population humaine, il existe plusieurs polymorphismes du gène NAT1. Parmi eux, le variant NAT1*10 affecte l'hexanucléotide et la séquence en aval (AATAAATGTC → AAAAAAATGTA). Ce variant est considéré comme un important déterminant génétique de risque de cancer colorectal (Bell et al., 1995)

Cette étude montre que des mutations au niveau de l'hexamère de certains gènes peuvent augmenter le risque de développer des cancers.

V-3-2 Activation de la PAP

L'activité de la PAP est régulée de manière cycline-dépendante. En effet, cette protéine est phosphorylée par la kinase cdc2-cyclinB tout au long du cycle cellulaire. Mais durant la phase M, la PAP se retrouve hyperphosphorylée inhibant son activité au cours de cette phase dans des cellules normales (Colgan et al., 1996; Colgan et al., 1998; Zhao and Manley, 1998). De plus, la mutation des sites d'hyperphosphorylation de la PAP cause l'arrêt du cycle cellulaire dans les phases G0-G1 (Zhao and Manley, 1996).

INTRODUCTION

La PAP est surexprimée dans différents types de cancers. Ainsi, une augmentation de la quantité de l'ARNm de la PAP est retrouvée dans les cancers humains du sein, du côlon, des ovaires et du pancréas (Pendurthi et al., 1997). De plus, l'activité de la PAP est significativement augmentée dans les lymphocytes B des leucémies aiguës de forme agressive, par rapport aux lymphocytes B des leucémies chroniques moins agressives et les lymphocytes B normaux (Trangas et al., 1989). Une augmentation de l'activité de la PAP a également été retrouvée dans des cancers du sein susceptibles de former des métastases ganglionnaires comparativement à des formes moins agressives de cancer du sein (Scorilas et al., 2000). L'activité de la PAP reflète donc probablement l'activité proliférative des cellules et sert d'indicateur de la différenciation. Par conséquent, elle peut servir de biomarqueur pronostique dans différentes formes de cancers (Jacob et al., 1989; Pangalis et al., 1985; Sasaki et al., 1990; Scorilas, 2002). En particulier, des anticorps dirigés contre la PAP retrouvés dans le sérum de patients atteints de cancer semblent avoir un potentiel prédictif dans l'apparition et la progression du cancer (Stetler et al., 1981). Toutefois, ces anticorps ne sont pas spécifiques du cancer et ont également été observés chez des patients atteints de troubles rhumatismaux (Stetler et al., 1987).

A l'heure actuelle, on ne sait pas si la stimulation de l'activité de la PAP contribue à la tumorigenèse ou si cela correspond à une conséquence de l'adaptation cellulaire à une prolifération rapide.

V-3-3 Raccourcissement de la taille des 3'UTR

Les séquences 3'UTR permettent de réguler la localisation, la stabilité ainsi que la traduction des ARNm. Au cours du développement, la taille des 3'UTR de l'ensemble des gènes augmente. Cette régulation globale est contrôlée par la polyadénylation alternative qui correspond à la présence de plusieurs signaux de polyadénylation disposés en tandem au niveau de l'exon 3' terminal. L'augmentation de la taille des 3'UTR due à une diminution de l'efficacité de la maturation en 3' corrèle avec la différenciation ainsi que la prolifération cellulaire (Ji et al., 2009). Cette diminution d'efficacité de maturation en 3' s'accompagne de l'utilisation du site de polyadénylation distal. De manière plus générale, on observe un raccourcissement global de la taille des 3'UTR dans des cellules en cours de

INTRODUCTION

prolifération (Sandberg et al., 2008). Plus récemment, il a été mis en évidence un raccourcissement similaire dans des cellules cancéreuses (Mayr and Bartel, 2009; Singh et al., 2009). Ce raccourcissement de la taille du 3' UTR permet à un ARNm d'être à la fois plus stable et traduit plus efficacement. En effet, cette régulation permet d'échapper à la régulation due aux micro-RNA ou aux séquences riches en adénines et uraciles qui inhibent la traduction et stimule la dégradation des ARNm (Eulalio et al., 2008; Grosset et al., 2004; Shaw and Kamen, 1986). Les ARNm contenant des 3'UTR plus courts produisent jusqu'à 10 fois plus de protéines que leurs homologues contenant des 3'UTR plus longs (Mayr and Bartel, 2009). En particulier, l'expression de l'isoforme courte du proto-oncogène IGF2BP1 (« Insulin-like growth factor 2 mRNA binding protein 1 ») entraîne sa transformation en oncogène.

Le mécanisme de régulation de la maturation en 3' alternatif ne semble pas être médié par des mutations de séquence. En effet, ce mécanisme est global et affecte plusieurs gènes. La mutation d'un signal de polyadénylation ne suffit pas à expliquer cette globalité. De plus, aucune mutation qui aurait changé la force d'un signal de polyadénylation n'a été identifiée sur différents gènes subissant ce mécanisme (Mayr and Bartel, 2009). Il a donc été proposé que le mécanisme de cette régulation passe par une régulation des facteurs agissant en *trans*. L'analyse de la quantité d'ARNm des différents facteurs de maturation en 3' dans des lignées saines comparées à des lignées cancéreuses montre des différences d'expression. Notamment, la quantité d'ARNm de CPSF 160 et de CstF 64 augmente dans les cellules cancéreuses (Mayr and Bartel, 2009). Cette augmentation des facteurs de maturation en 3' pourrait entraîner une augmentation de l'efficacité de maturation en 3' conduisant à un raccourcissement des 3' UTR.

RESULTATS

I Rôle essentiel de l'interaction entre hnRNP H/F et un G-quadruplexe pour maintenir la maturation en 3' du pré-ARNm de p53 lors de dommages à l'ADN

I-1 Présentation et objectifs de l'étude

Le facteur de transcription P53 est une protéine nucléaire constituée de cinq domaines fonctionnels principaux : un domaine de transactivation, situé à l'extrémité N-terminale, une région riche en proline, composée de cinq répétitions PXXP, un domaine central requis pour la fixation spécifique de P53 à l'ADN, un domaine d'oligomérisation permettant la formation d'un tétramère essentiel pour l'activité transcriptionnelle et un domaine C-terminal qui contient un signal de localisation nucléaire (Hulboy and Lozano, 1994; Slee et al., 2004).

La P53 décrite comme le « gardien du génome », est un suppresseur de tumeur. En effet, cette protéine est présente dans toutes les cellules de l'organisme et est activée lorsqu'une cellule subit un stress comme l'exposition aux radiations ionisantes, aux UV, le traitement par des drogues chimiothérapeutiques, l'hypoxie, l'activation d'oncogènes, l'absence de facteurs de survie. Une fois activée, la fonction de P53 est de réguler la transcription de nombreux gènes impliqués dans les mécanismes d'arrêt du cycle cellulaire, de réparation des dommages, ou d'apoptose. Le choix de la réponse cellulaire se fait en fonction de l'intensité des dommages à l'ADN (Offer et al., 2002), de la quantité de P53 accumulée (Chen et al., 1996), et du type cellulaire (Burns and El-Deiry, 2003; Fei et al., 2002).

En condition normale, la demi-vie courte (20 min) de P53 la rend difficilement détectable dans les cellules et les tissus normaux (Vogelstein et al., 2000). Cette demi-vie courte est due à la dégradation de P53 sous contrôle de la protéine MDM2 (Murine Double Minute) qui en se fixant à P53 favorise son ubiquitination entraînant sa dégradation protéolytique par le protéasome (Deb, 2002; Yang et al., 2004). Lors de dommages à l'ADN, P53 est phosphorylée par ATM (Khosravi *et al.* 1999), ATR

(« Ataxia-Telangiectasia Related ») (Tibbetts et al., 1999), DNA-PK (« DNA-dependent protein kinase »), les caséines-kinases I ou II, ou ChK1/ChK2 (Chehab et al., 2000). La phosphorylation de P53 bloque sa liaison avec MDM2 et donc sa dégradation, entraînant une accumulation de P53 dans la cellule (Kruse and Gu, 2009).

Une augmentation de la transcription de p53 permet également de participer à son accumulation lors de dommages à l'ADN. Notamment, l'activation du facteur de transcription NF- κ B (« Nuclear Factor-kappa B ») (Janssens and Tschopp, 2006) induit la transcription de p53 et participe à son accumulation suite à des dommages à l'ADN (Hellin et al., 1998; Pei et al., 1999; Sun et al., 1995). La protéine RREB-1 (« Ras Responsive Element Binding protein-1 ») reconnaît également le promoteur de p53 et permet de contrôler son expression ainsi que l'apoptose médiée par P53 lors de stress génotoxiques (Liu et al., 2009). De même, lors de dommages à l'ADN, la protéine de liaison à la pol II, Che-1, phosphorylée par ATM, ATR et ChK2, s'accumule. Cette accumulation induit un recrutement spécifique de Che-1 sur le promoteur de p53 entraînant une augmentation de sa transcription (Bruno et al., 2006; Bruno et al., 2010).

L'accumulation de P53 lors de stress génotoxiques résulte aussi d'une stimulation de la traduction de son ARNm. L'ARNm de p53 possède une IRES dont l'activité de traduction augmente dans des cellules traitées avec des agents génotoxiques (Yang et al., 2006). D'autres études ont montré que la protéine ribosomale L26 (RPL26) qui interagit avec le 5'UTR de l'ARNm de p53 contrôle la traduction et l'accumulation de P53 lors de dommages à l'ADN (Takagi et al., 2005). La traduction régulée de p53 par RPL26 requiert l'appariement de séquences du 5'UTR avec le 3'UTR de p53 (Chen and Kastan, 2010). Le 3'UTR de p53 est également la cible du microRNA-125b inhibant sa traduction de p53. La quantité de ce microRNA diminue lors de dommages à l'ADN permettant une traduction de p53 efficace (Le et al., 2009). Enfin, lors de dommages à l'ADN la protéine de liaison aux ARN HuR, interagit avec des séquences riches en uraciles du 3'UTR de l'ARNm de p53 et augmente sa traduction (Mazan-Mamczarz et al., 2003).

Nous avons vu précédemment que lors de dommages à l'ADN, la maturation en 3' est inhibée de manière globale. Néanmoins, une augmentation de la quantité de transcrit de p53 est indispensable à l'accumulation de P53 lors de ce stress. Ceci

suggère que le pré-ARNm de p53 n'est pas affecté par l'inhibition de la maturation en 3' lors de dommages à l'ADN.

Ainsi, afin d'étendre la compréhension de la régulation post-transcriptionnelle du gène p53, nous nous sommes intéressés à la maturation en 3' de son pré-ARNm. Notre premier objectif a été de déterminer l'effet des dommages à l'ADN sur l'efficacité de la maturation en 3' du pré-ARNm de p53. Dans le but de caractériser le mécanisme qui permet au pré-ARNm de p53 de résister à l'inhibition de la maturation en 3' lors de dommages à l'ADN, nous avons analysé sa séquence. Les séquences en aval du signal de polyadénylation du pré-ARNm de p53 sont très conservées entre différentes espèces et ont la particularité d'être riches en guanines. Des séquences riches en guanines en aval du signal de polyadénylation de plusieurs gènes ont été décrites pour réguler la maturation en 3', en recrutant la famille de protéines de liaison aux ARN, les hnRNP H/F (« heterogeneous nuclear RiboNucleoProteins H/F ») (Alkan et al., 2006; Bagga et al., 1998; Caputi and Zahler, 2001; Dalziel et al., 2007). De plus, l'analyse bioinformatique de cette séquence révèle qu'elle est potentiellement capable de se structurer en G-quadruplexe. Dans un second temps, notre objectif a donc été de déterminer l'implication de l'interaction entre la structure de type G-quadruplexe et les hnRNP H/F dans la régulation de maturation en 3' du pré-ARNm de p53 lors de dommages à l'ADN.

I-2 Contexte de l'étude

I-2-1 La famille des hnRNP H/F

Les hnRNP H/F sont des hnRNP, ces dernières représentent un ensemble d'au moins 20 protéines qui lient les ARNm ou les pré-ARNm. Elles ont été regroupées sur la base de leurs associations aux transcrits primaires, leurs similitudes de séquence et leurs domaines structuraux caractéristiques (Dreyfuss et al., 1993; Krecic and Swanson, 1999). Elles contiennent des domaines de liaison à l'ARN appelé RRM (« RNA Recognition Motif ») ainsi que des domaines auxiliaires tels que les domaines riches en arginines et glycines, RGG (Cartegni et al., 1996). Les hnRNP sont des protéines ubiquitaires, néanmoins, leurs expressions varient en fonction des tissus et des types cellulaires.

La famille des hnRNP H/F comprend 4 membres codés par différents gènes : H1 ou H ; H2 ou H' ; H3 ou 2H9 et F. Ces protéines possèdent 2 (hnRNP H3) ou 3 domaines (hnRNP H1, H2 et F) RRM leur permettant d'interagir avec les séquences d'ARN riche en guanines (Alkan et al., 2006; Caputi and Zahler, 2001; Dominguez and Allain, 2006) et 2 domaines riches en glycines.

Les hnRNP H/F sont essentiellement impliquées dans l'épissage des pré-ARNm (Coles et al., 2009; Van Dusen et al., 2010) et en particulier des pré-ARNm de gènes régulant l'apoptose. En effet, les hnRNP H1 et H2 contrôlent l'épissage du pré-ARNm de a-raf régulant l'apoptose induite par la kinase MST2 (« Mammalian STE20-like kinase 2 ») (Rauch et al., 2011; Rauch et al., 2010). Les hnRNP H1 et F contrôlent également l'épissage alternatif du pré-ARNm du facteur anti-apoptotique BCL2L1 (« B-Cell Leukemia 2 Lymphoma 1 ») (Garneau et al., 2005). Les hnRNP H/F régulent également l'épissage de proto-oncogène comme SRC (sarcoma) (Min et al., 1995) ou HRAS (« Harvey Rat sarcoma Viral oncogene homolog ») (Camats et al., 2008). De plus, les protéines hnRNP H/F peuvent être dérégulées dans certains cancers. Une analyse comparative en immunohistochimie des hnRNP H/F dans les tissus sains *versus* les tissus cancéreux révèle que leur répartition nucléo-cytoplasmique est modifiée et qu'elles sont surexprimées dans les tissus cancéreux (Honore et al., 2004). Une implication plus directe des hnRNP H/F dans le cancer du sein est proposée par

deux études : une analyse protéomique d'une lignée cellulaire de cancer du sein (MCF-7) (Malorni et al., 2006) et un profil du transcriptome par SAGE de tumeurs invasives du sein (Abba et al., 2005).

Les hnRNP H/F sont aussi impliquées dans la transcription (Yoshida et al., 1999), l'export des ARNm (Kim et al., 2005) et la traduction IRES (« Internal Ribosomal Entry Site ») dépendante de l'ARNm de SHMT1 (« Serine Hydroxymethyltransferase 1 »), notamment lors de dommages à l'ADN (Fox et al., 2009; Fox and Stover, 2009). Lors de différents traitements génotoxiques, il a été mis en évidence une augmentation de l'expression des protéines hnRNP H/F (Decker et al., 2003; Fox et al., 2009; Rauch et al., 2010). De plus, hnRNP H est également phosphorylée lors de ce stress (Navakauskiene et al., 2004).

Enfin, les rôles des hnRNP H/F sur le clivage et la polyadénylation ont été mis en évidence au niveau de plusieurs pré-ARNm (Alkan et al., 2006; Arhin et al., 2002; Garneau et al., 2005; Millevoi et al., 2009). Les hnRNP H/F interagissent avec le pré-ARNm à proximité du signal de polyadénylation et stimulent la maturation en 3' en recrutant CstF (Bagga et al., 1998; Chen and Wilusz, 1998) et la PAP (Millevoi et al., 2009). Cependant, il a également été montré que hnRNP F peut inhiber la réaction de clivage en diminuant le recrutement de CstF sur le DSE au niveau du site de polyadénylation proximal du pré-ARNm des chaînes lourdes de l'IgM. (Veraldi et al., 2001). L'exemple le plus connu de stimulation de la maturation en 3' par les hnRNP H/F concerne le pré-ARNm de SV40L (« Simian virus 40 Large ») (Alkan et al., 2006; Arhin et al., 2002). Le pré-ARNm de SV40L possède un DSE aux riches en guanines reconnu par les hnRNP H/F. Cette séquence riche en guanines est potentiellement capable de se structurer en G-quadruplexe. Il est possible que les hnRNP H/F reconnaissent spécifiquement les G-quadruplexes ou qu'elles stimulent leur structuration (Zarudnaya et al., 2003).

I-2-2 Les G-quadruplexes

Les G-quadruplexes constitués d'empilements de G-quartets correspondant à l'association de 4 guanines par des liaisons « *hoogsteen* » peuvent être formés par le repliement d'une seule molécule d'acide nucléique (figure 14A), par l'association de 2 molécules (figure 14B) ou par 4 brins d'acides nucléiques (figure 14C) (Crnugelj et al., 2002).

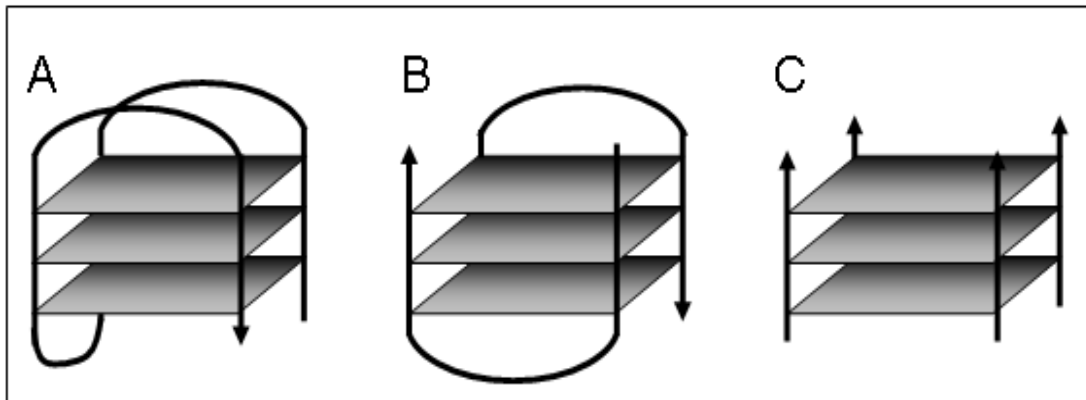


Figure 14 : Exemples de G-quadruplexe intra et intermoléculaire.

A : G-quadruplexe intramoléculaire. B : G-quadruplexe intermoléculaire à deux brins. C : G-quadruplexe intermoléculaire à quatre brins. Les flèches indiquent l'orientation du brin d'acide nucléique du 5' vers le 3'. Cette figure s'inspire de (Riou et al., 2003).

Les G-quadruplexes sont principalement décrits dans l'ADN au niveau des télomères (Parkinson et al., 2002; Wang and Patel, 1993). Ces G-quadruplexes régulent la taille des télomères en empêchant l'accès des nucléases et de la télomérase. La formation de G-quadruplexes télomériques intermoléculaires permettraient également l'alignement interchromosomique lors de la méiose (Oganesian and Bryan, 2007). Les G-quadruplexes sont aussi décrits au niveau des régions de commutation des gènes des chaînes lourdes des immunoglobulines (Sen and Gilbert, 1990) et des régions répétées de l'ADNr (Hanakahi et al., 1999). Enfin, les promoteurs de nombreux gènes (essentiellement des oncogènes) tels que *c-Myc*, *PDGF-A*, *c-Kit*, *bcl-2*, *Rb*, *RET*, *hTERT*, *K-Ras*, contiennent également des séquences capables de se structurer en G-quadruplexe et susceptibles d'inhiber la transcription. (Brooks et al., 2010; Eddy and Maizels, 2006). De part leur localisation (télomères et promoteur

d'oncogènes), les G-quadruplexes ont récemment été étudiés comme cible de traitement anti-cancéreux. De nombreuses classes de molécules capables de structurer les G-quadruplexes ont ainsi été décrites parmi lesquelles des porphyrines cationiques (TMPyP4), des dérivés de pérylène (PIPER), des acridines trisubstituées (BRACO19), des bisacridines, des acridines pentacycliques (RHPS4), des produits naturels (télomestatine), des dérivés d'éthidium, des dibenzophenanthrolines, des dérivés triazine (12459), des dérivés carboxamide (307A) des fluoroquinophenoxazines (QQ58) des porphyrines anioniques (NMM) ou encore la Quarfloxin® (CX-3543, Pharmaceuticals Cylene à San Diego, CA, USA) (Neidle, 2009; Yang and Okamoto, 2010). Ces drogues ciblent préférentiellement des séquences d'ADN capables de se structurer en G-quadruplexes et ont différentes affinités pour les G-quadruplexes en fonction de leurs conformations et de leurs stabilités.

Les G-quadruplexes ne se structurent pas uniquement sur des molécules d'ADN, mais peuvent également se former sur des ARN. En effet, les ARNm des plusieurs gènes codant la Sémaphorine 3F ou l'histone H4 possèdent des séquences capables de se structurer en G-quadruplexe. Ces G-quadruplexes sont la cible de la protéine FMRP (« Fragile X Mental Retardation Proteine ») absente dans le syndrome mental du chromosome X. La protéine FMRP régule la localisation et la traduction de ces ARNm cibles possédant un G-quadruplexe (Bechara et al., 2009; Subramanian et al., 2011).

Des G-quadruplexes dans les 5'UTR de nombreux gènes sont généralement décrits pour inhiber la traduction dépendante de la coiffe (Beaudoin and Perreault, 2010; Halder et al., 2009; Kumari et al., 2007). Des séquences capables de se structurer en G-quadruplexe ont aussi été retrouvées au niveau d'IRES des ARNm de FGF2 (Bonnal et al., 2003) et du VEGF (Morris et al., 2010).

Les G-quadruplexes permettent également de réguler la maturation des pré-ARNm. En effet, le gène TERT (« telomerase reverse transcriptase ») possède dans son intron 6, une séquence capable de former différents G-quadruplexes. L'ajout d'une drogue qui stabilise ces G-quadruplexes stimule l'épissage de cette région (Gomez et al., 2004). Une autre étude a montré que le pré-ARNm de p53 possède dans son intron 3, une séquence structurée en G-quadruplexe. Différentes délétions de piliers de ce G-quadruplexe augmentent la rétention de l'exon 2 situé en amont (Marcel et al., 2010). Un polymorphisme de ce G-quadruplexe appelé PIN3 consiste en une duplication de

RESULTATS

16 bases (ACCTGGAGGGCTGGGG) qui augmente potentiellement la structuration du G-quadruplexe. Ce polymorphisme est associé à une diminution d'expression du transcrit p53 et une augmentation du risque de cancers du cerveau, du poumon, des ovaires et des cancers colorectaux (Gemignani et al., 2004; Hung et al., 2006; Wang-Gohrke et al., 2002). En plus de l'épissage, les G-quadruplexes sont également supposés stimuler la maturation en 3' (Zarudnaya et al., 2003)

I-3 Stratégie expérimentale

I-3-1 Etude de l'effet des dommages à l'ADN sur la maturation en 3' du pré-ARNm de p53

Les modèles cellulaires choisis pour cette étude correspondent à différentes lignées cellulaires tumorales humaines (A549 issue d'adénocarcinome pulmonaire ; MCF7 issue d'une tumeur mammaires et HCT116 issue de cancer du colon) possédant le gène de la p53 non muté et déjà utilisé pour étudier l'effet de l'apoptose induite par P53 lors de dommages à l'ADN (Jung et al., 2010; Lu et al., 2001). Nous avons exposé ces cellules à différents agents génotoxiques afin de créer des dommages à l'ADN et observé par Western blot l'accumulation de P53.

Afin d'évaluer l'efficacité de la maturation en 3', nous avons mis au point une technique consistant à quantifier le pré-ARNm non clivé de différents gènes. Une reverse transcription suivie d'une PCR en temps réel (RT-PCRq) à l'aide d'oligonucléotides localisés de part et d'autre du site de clivage nous a permis d'évaluer la quantité de pré-ARNm d'un gène. Nous avons normalisé cette quantité par rapport à la quantité d'ARN total de ce gène (ARNm et pré-ARNm) grâce à des oligonucléotides localisés en amont du site de clivage. Cette technique nous a permis d'évaluer l'efficacité de la maturation en 3' de p53 et de TBP (« TATA box binding protein ») utilisé comme contrôle. Ces expériences nous ont permis de mettre en évidence la résistance à l'inhibition de la maturation en 3' lors de dommages à l'ADN du pré-ARNm de p53.

I-3-2 Caractérisation des séquences régulant la maturation en 3' du pré-ARNm de p53

Dans le but de déterminer l'implication des différentes régions de la séquence de p53 responsable de la résistance à l'inhibition de la maturation en 3', nous avons cloné la séquence correspondant au 3' UTR d'une longueur de 1270 nucléotides ainsi que les 300 nucléotides en aval. Nous avons cloné cette séquence en aval du gène de la luciférase *renilla*. Les variations des activités luciférases sont la conséquence de

mécanismes mettant en jeu les séquences ajoutées en 3' du gène rapporteur. Pour déterminer les séquences *cis*-régulatrices, nous avons réalisé différentes délétions des régions proches du signal de polyadénylation dans la séquence clonée en aval de la luciférase. Ces différentes constructions ont été transfectées de manière transitoire dans des cellules A549. Un plasmide contrôle contenant la luciférase *firefly* est également co-transfecté pour normaliser l'efficacité de transfection. La technique de quantification de l'efficacité de maturation en 3' n'étant plus possible pour ces constructions (délétion des régions d'hybridation avec les oligonucléotides de PCR quantitative) nous avons utilisé deux manières différentes pour évaluer l'efficacité de maturation en 3' :

-En quantifiant les activités luciférase, obtenues après la transfection de chacune des constructions. Cependant, cette technique ne permet pas de discriminer des régulations par les séquences localisées en amont du site de clivage sur d'autres étapes de l'expression des gènes comme la stabilité de l'ARNm ou la traduction.

-En utilisant une technique de protection aux ribonucléases permettant de visualiser les quantités de pré-ARNm non clivé et d'ARNm clivé.

Ces expériences nous ont permis d'identifier une région de 46 nucléotides localisée en aval du signal de polyadénylation jouant un rôle dans la maturation en 3' du pré-ARNm de p53. Par ailleurs, cette région est capable, *in silico*, de se structurer en G-quadruplexe.

I-3-3 Etude de la structuration en G-quadruplexe des séquences localisées en aval du site de clivage du pré-ARNm de p53

Afin de déterminer si la séquence localisée en aval du signal de polyadénylation de p53 et qui régule sa maturation en 3' est capable de se structurer en G-quadruplexe, nous avons utilisé une technique de « pause de reverse transcription dépendante de cations ». Cette technique consiste à reverse transcrire un ARN en présence de différents cations (Na^+ ou K^+) à l'aide d'une reverse transcriptase insensible aux cations. Sachant que les G-quadruplexes sont plus stables en présence de K^+ que de Na^+ , un arrêt de reverse transcription, uniquement en présence de K^+ est révélateur d'une structuration en G-quadruplexe. Pour visualiser cet arrêt, nous utilisons comme amorce de reverse transcription un oligonucléotide radiomarqué. Après la réaction de reverse transcription les différentes tailles des ADNc radiomarqués sont analysées par

migration sur gel dénaturant. Cette expérience nous a permis de montrer que les séquences d'ARN situées en aval du signal de polyadénylation de p53 sont capables de se structurer in vitro en G-quadruplexe.

I-3-4 Rôle du G-quadruplexe de p53 sur la résistance à l'inhibition de la maturation en 3' du pré-ARNm lors de dommages à l'ADN

Pour déterminer le rôle du G-quadruplexe localisé en aval du site de clivage sur le pré-ARNm de p53, nous avons muté sa séquence. Nous nous sommes assurés que la mutation déstructurait bien le G-quadruplexe par la technique de pause de reverse transcription dépendante de cations. Nous avons ensuite cloné la séquence mutée dans les plasmides décrits précédemment. Nous avons comparé les efficacités de maturation en 3' des plasmides contenant la région 3' de p53 sauvage ou mutée, mesurées par RT-PCRq dans des cellules exposées ou non aux UV. Cela nous a permis de déterminer que le G-quadruplexe de p53 est essentiel pour résister à l'inhibition de la maturation en 3' lors de dommages à l'ADN.

I-3-5 Etude de l'association des protéines hnRNP H/F avec le G-quadruplexe de p53

Pour étudier l'interaction entre les protéines hnRNP H/F et le G-quadruplexe en aval du site de clivage de p53, nous avons utilisé différentes techniques. Une chromatographie d'affinité à l'aide de petits ARN biotinylés contenant les séquences du G-quadruplexe de p53 sauvage et mutée, suivie d'un Western Blot nous a permis de montrer que la mutation qui déstructure le G-quadruplexe diminue l'interaction avec hnRNP H/F. Des couplages covalents aux UV suivis d'immunoprécipitation nous ont permis de démontrer que l'interaction entre les hnRNP H/F et le G-quadruplexe est directe. Enfin nous avons immunoprécipité les protéines hnRNP H/F à partir d'extraits issus de cellules traitées au formaldéhyde (entraînant des couplages covalents entre ARN et protéines) et exposées ou non aux UV. Ces expériences nous ont montré que la structure en G-quadruplexe des séquences en aval du signal de polyadénylation de p53 semble essentielle pour interagir avec les hnRNP H/F et que cette interaction est augmentée par les dommages à l'ADN.

I-3-6 Rôle des protéines hnRNP H/F dans l'accumulation et la fonction de p53 lors de dommages à l'ADN

Pour déterminer l'implication des hnRNP H/F sur l'accumulation et la fonction de P53 lors de dommages à l'ADN, nous avons réalisé des expériences d'extinction d'expression par des ARN interférant(ARNi) dirigés contre l'ARNm de hnRNP H1, H2 ou F dans des cellules exposées ou non aux UV. L'expression de P53 a ensuite été observée par Western Blot et l'efficacité de maturation en 3' du pré-ARNm de p53 analysée par RT-PCRq. Pour évaluer l'implication des hnRNP H/F dans l'apoptose induite par P53, nous avons diminué l'expression des hnRNP H/F par siARN sur une lignée cellulaire de cancer colorectal HCT116 possédant ou non le gène p53. L'apoptose a été évaluée par mesure du clivage de la PARP (Poly ADP Ribose Polymérase) par Western Blot ou par l'estimation du pourcentage de cellules en sub-G1 par cytométrie en flux.

I-4 Article I

« Essential role for the interaction between hnRNP H/F and a G-quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage »

Adrien Decorsière, Anne Cayrel, Stéphan Vagner and Stefania Millevoi



Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage

Adrien Decorsière, Anne Cayrel, Stéphan Vagner, et al.

Genes Dev. 2011 25: 220-225

Access the most recent version at doi:[10.1101/gad.607011](https://doi.org/10.1101/gad.607011)

Supplemental Material	http://genesdev.cshlp.org/content/suppl/2011/02/02/25.3.220.DC1.html
References	This article cites 34 articles, 17 of which can be accessed free at: http://genesdev.cshlp.org/content/25/3/220.full.html#ref-list-1
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here

To subscribe to *Genes & Development* go to:
<http://genesdev.cshlp.org/subscriptions>

Copyright © 2011 by Cold Spring Harbor Laboratory Press

RESEARCH COMMUNICATION

Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage

Adrien Decorsière,^{1,2,3} Anne Cayrel,^{1,2,3} Stéphane Vagner,^{1,2,3,4} and Stefania Millevoi^{1,2,3,5,6}

¹Institut National de la Santé et de la Recherche Médicale (INSERM) U563, Toulouse F-31000, France; ²Institut Claudius Regaud, Toulouse F-31052, France; ³Université Toulouse III Paul Sabatier, Toulouse F-31000, France

Following DNA damage, mRNA 3'-end formation is inhibited, contributing to repression of mRNA synthesis. Here we investigated how DNA-damaged cells accomplish p53 mRNA 3'-end formation when normal mechanisms of pre-mRNA 3'-end processing regulation are inhibited. The underlying mechanism involves the interaction between a G-quadruplex structure located downstream from the p53 cleavage site and hnRNP H/F. Importantly, this interaction is critical for p53 expression and contributes to p53-mediated apoptosis. Our results uncover the existence of a specific rescue mechanism of 3'-end processing regulation allowing stress-induced p53 accumulation and function in apoptosis.

Supplemental material is available for this article.

Received August 31, 2010; revised version accepted December 15, 2010.

The cellular response to DNA damage is a protective mechanism against disease. This involves a functional interaction of DNA repair and transcription machineries, resulting in a global decrease of mRNA levels and repair of the induced DNA lesions. Recently, it has been shown that pre-mRNA 3'-end processing plays an important role in the repression of mRNA synthesis.

mRNA 3'-end formation, a nuclear process consisting of cleavage of pre-mRNA 3' ends and poly(A) tail addition, is essential for mRNA functions, including stability, translocation to the cytoplasm, and translation. 3'-End processing proceeds through the recognition of *cis*-acting elements of the pre-mRNA (defined as the polyadenylation [pA] signal) by a complex machinery that is associated with a large number of protein factors regulating the efficiency and specificity of this process and mediating its

interaction with other nuclear events, including transcription and splicing (for recent reviews, see Danckwardt et al. 2008; Millevoi and Vagner 2010). Cleavage stimulation factor (CstF), an essential component of the basal 3'-end processing machinery (Takagaki et al. 1989), plays a coordinating role in the DNA damage response. Following DNA damage, CstF inhibits 3'-end processing by forming regulatory complexes with different factors, including RNA polymerase II (RNA Pol II), the BARD1/BRCA1 tumor suppressor, and PARN (Kleiman and Manley 1999, 2001; Kleiman et al. 2005; Kim et al. 2006; Mirkin et al. 2008; Cevher et al. 2010).

A paradigm of the DNA damage response is that transcripts of genes inducing cell cycle arrest, DNA repair, and apoptosis are expressed before the DNA repair process begins. This suggests that gene-specific compensatory mechanisms should be used on DNA damage-activated genes to ensure their transcription and proper formation of mRNA 3' ends, providing the cell with a way to react to genotoxic stresses and maintain genomic integrity. Among these, the p53 tumor suppressor protein is pivotal in regulating the biological response to DNA damage. Following genotoxic insult, p53 functions as a sequence-specific transcription factor that regulates the expression of downstream target genes required for cell cycle arrest, DNA repair, or apoptosis (Levine and Oren 2009). Although it is well accepted that the activity of p53 is controlled through post-translational modifications (Kruse and Gu 2009), recent studies have revealed that p53 expression is also regulated at the transcriptional (Bruno et al. 2006) and translational (Mazan-Mamczarz et al. 2003; Takagi et al. 2005; Ofir-Rosenfeld et al. 2008) levels in response to genotoxic stress. Of importance, these studies not only show that p53 activation occurs through specific mechanisms that allow the p53 gene to bypass general repression of gene expression, but also suggest that the 3'-end processing step must be tightly regulated to ensure proper formation of p53 mRNA 3' ends.

Here we investigated how DNA-damaged cells accomplish p53 mRNA 3'-end formation when normal mechanisms of pre-mRNA 3'-end processing regulation are inhibited. We show that p53 pre-mRNA resists general inhibition of 3'-end processing caused by DNA damage stresses. The underlying mechanism involves the interaction between hnRNP H/F and a G-quadruplex RNA structure at the p53 pA signal. We also show that hnRNP H/F depletion compromises p53 pre-mRNA 3'-end processing, protein expression, and p53-mediated apoptosis.

Results and Discussion

p53 pre-mRNA 3'-end processing resists DNA damage-dependent inhibition of mRNA 3'-end formation

To assess whether p53 mRNA 3'-end formation is inhibited during DNA damage, we exposed different cell lines to DNA-damaging agents and quantified 3'-end processing efficiency of the p53 primary transcript or the control TBP pre-mRNA. Exposure of A549, HCT116, and MCF7 cells to DNA-damaging agents doxorubicin and UV irradiation induced significant p53 protein accumulation (Supplemental Fig. 1). Real-time PCR analysis of p53 3'-end processing efficiency measured by the ratio

[**Keywords:** DNA damage, pre-mRNA 3'-end processing, polyadenylation, p53 tumor suppressor, hnRNP H, hnRNP F]
Present addresses: ⁴INSERM U981, Institut Gustave Roussy, 114 Rue Edouard Vaillant, 94800 Villejuif, France; ⁵INSERM U858, Bât. L3, B.P. 84225, 31432 Toulouse Cedex 4, France.
⁶Corresponding author.
 E-MAIL stefania.millevoi@inserm.fr; FAX 33-561-32-21-41.
 Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.607011>.

hnRNP H/F rescues p53 activation during DNA damage

of uncleaved (not polyadenylated) (Supplemental Fig. 2) to total RNA (or the ratio of cleaved and polyadenylated to total RNA) (Supplemental Fig. 3) revealed that p53 mRNA 3'-end formation is maintained in all cell lines after doxorubicin exposure or UV irradiation (Fig. 1). In contrast, and as expected from previous studies, TBP pre-mRNA 3'-end processing significantly decreased after exposure to both DNA-damaging agents in the different cell lines, as revealed by the increased ratio of uncleaved to total TBP RNA (Fig. 1). Similarly, p53 target gene *p21*, but not *hprt*, has been shown previously to be properly transcribed and processed at the 3' end following DNA damage by a mechanism that bypasses the requirement for CTD phosphorylation and the presence of several known RNA Pol II elongation factors (Gomes et al. 2006). These findings support the hypothesis that DNA-damaged cells use specific alternative regulatory pathways to counteract general repression of mRNA synthesis.

G-rich sequences located downstream from the p53 pre-mRNA cleavage site contribute to enhance 3'-end processing efficiency

To address how p53 pre-mRNA is efficiently processed during DNA damage, we first focused on identifying regulatory sequence elements within the regions flanking the pA signal of the p53 pre-mRNA. This revealed the presence of two *cis*-elements known to modulate 3'-end processing efficiency (Millevoi et al. 2009 and references therein): a U-rich element located upstream of the pA signal, and a G-rich region positioned downstream from the cleavage site (Fig. 2A). The function of these elements was tested by using a luciferase reporter vector containing the p53 3'-end region inserted downstream from the luciferase gene (Rluc/p53), and two deletion constructs in which either the U-rich (ΔU) or the 3'-end-containing G-rich region (D1) was deleted (Fig. 2A). The luciferase reporter assay showed a 2.5-fold decrease in luciferase expression when deleting the G-rich region, while a minor decrease was obtained with the ΔU construct (Fig. 2B). To

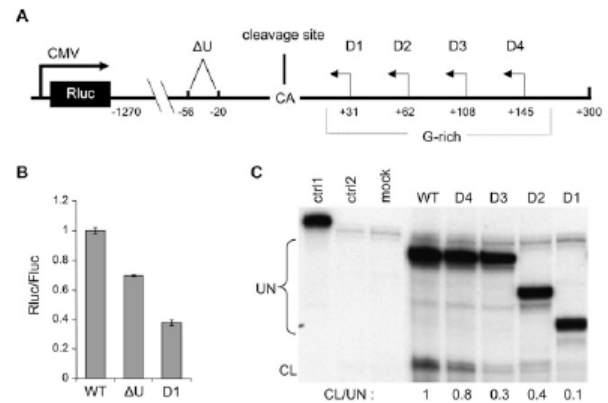


Figure 2. p53 pre-mRNA 3'-end processing is regulated by a downstream G-rich element. (A) Illustration of the luciferase reporter vector (Rluc/p53) containing the p53 pA signal and 3'-end-flanking regions inserted downstream from the Renilla luciferase gene (Rluc). Individual clones carrying deletion of the U-rich element (ΔU) or the G-rich region (D1–D4) are depicted, and their positions are indicated as positive and negative values relative to the cleavage site (CA). (CMV) Cytomegalovirus promoter. (B) The ratio of Renilla/firefly luciferase activities (Rluc/Fluc) was determined using HeLa cells cotransfected with the wild-type, ΔU , and D1 Rluc/p53 reporter plasmids and an internal control plasmid encoding the firefly luciferase (Fluc). (C) RPA of nuclear RNA isolated from HeLa cells transiently transfected with the indicated Rluc/p53 reporter plasmids. Protected bands representing the uncleaved (UN) and cleaved (CL) transcripts are shown. (Ctrl1) Without RNase; (Ctrl2) yeast RNA. Note that the length of the bands corresponding to uncleaved transcripts differs in each lane because the introduced deletions overlap with the protection probe, which contains wild-type 3'-flanking sequences. The cleavage efficiency—measured by the ratio of cleaved to uncleaved RNA and normalized to the wild-type construct—is shown below.

better characterize the functional role of this region in pre-mRNA 3'-end processing, we created a series of Rluc/p53 constructs in which the G-rich region was progressively deleted in a 3'-to-5' direction and analyzed mRNA 3'-end formation by RNase protection analysis (RPA). RPA of the transfected reporter plasmids showed that, whereas the D4 construct did not lead to a major effect on 3'-end processing efficiency, a fivefold decrease in 3'-end processing efficiency was observed with the D3 construct (Fig. 2C). This suggests that the sequence encompassing nucleotides +108 to +145 contains sequence or structural elements that are critical for p53 pre-mRNA 3'-end processing.

A G quadruplex in the p53 pre-mRNA 3'-flanking region is critical for maintaining p53 3'-end processing efficiency following UV irradiation

G-rich sequences are potential candidates to form G-quadruplex structures (G4), consisting of planar arrays of four guanines (G quartets), each guanine pairing with two neighbors by Hoogsteen bonding. These are specific DNA/RNA architectures adopted in telomeres, promoter regions, and 5' and 3' ends of the primary transcripts, and play a role in telomere lengthening, transcription, and translation (for a recent review, see Lipps and Rhodes 2009). Biocomputing analysis of the previously identified p53 pre-mRNA regulatory region using the QGRS Mapper program (<http://bioinformatics.ramapo.edu/QGRS/index.php>) revealed that the sequence encompassing

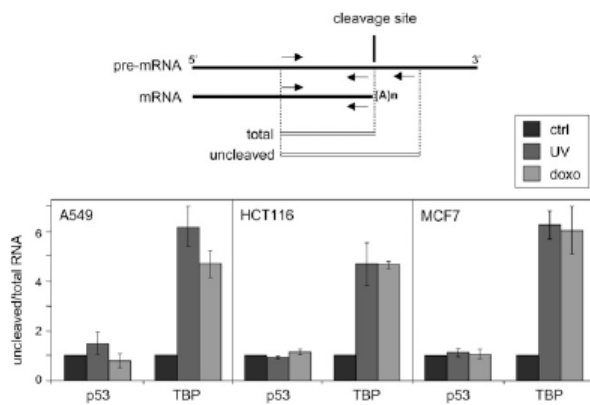


Figure 1. p53 pre-mRNA 3'-end processing is not affected by DNA-damaging agents. Reverse transcription of nuclear RNA from A549, HCT116, and MCF7 cells left untreated (ctrl), UV-irradiated (UV), or treated with doxorubicin (doxo) followed by real-time PCR with primers specific to either p53 or TBP located on either side of the cleavage site to amplify the uncleaved RNA, or upstream of the cleavage site to amplify both the uncleaved and cleaved RNA, as depicted.

Decorsière et al.

nucleotides +108 to +142 folds in a G4 structure. Since planar layers of G quartets are stabilized preferentially by K⁺ in comparison with Na⁺ (Sundquist and Klug 1989), we investigated the presence of a G4 by analyzing the influence of the nature of the cation on the pauses of reverse transcription catalyzed by the avian myeloblastosis virus (AMV) reverse transcriptase that has no monovalent cation requirement. As shown in Figure 3A, when NaCl was substituted for KCl in the reverse transcription buffer, the strong pauses observed with K⁺ disappeared. This result strongly suggests the occurrence of a G4 in the G-rich (+108 to +142) region. A GGGG-to-CACA mutation in the last G stretch of the predicted G4 (G4M) abolished the K⁺-dependent reverse transcriptional pausing (Fig. 3A), further demonstrating the presence of a G4 in this G-rich sequence element.

To determine whether this structure is involved in the mechanism allowing p53 pre-mRNA to be properly processed at the 3' end during DNA damage, we introduced the G4M mutation in the *luc/p53* vector. The G4M mutant induced a sixfold and fourfold inhibition of 3'-end processing efficiency compared with the wild-type sequence in untreated (Fig. 3B, graph i) and UV-irradiated (Fig. 3B, graph ii), cells respectively, suggesting that the ability to form a G4 at this position is critical for p53 pre-

mRNA 3'-end processing during DNA damage. The relationship between the G4 structure and its function in p53 3'-end processing before and after UV irradiation was further confirmed by using two additional G4 mutants (Supplemental Fig. 4). To further show the importance of G4 formation in p53 3'-end processing, we used TMPyP4, a compound known to bind and stabilize G4 structures (Lipps and Rhodes 2009). Cation-dependent reverse transcription showed that TMPyP4 increases the ability of the G4 formed at the +108 to +142 region to pause the AMV reverse transcriptase, suggesting an improved stability of the G4 in the presence of this compound (Fig. 3C). Treatment of A549 cells with TMPyP4 not only induced p53 protein expression, but also increased 3'-end processing of the p53 transcripts compared with TBP (Fig. 3D), suggesting that stabilization of the p53 G4 plays a role in stimulating p53 gene expression by stimulating the 3'-end processing step. Extending the role of G4 in gene expression, this set of results shows that a functional G-rich element at the p53 pA signal is structured in a G4 conformation that is not only critical to stimulate p53 3'-end processing, but also necessary to counteract 3'-end formation inhibition induced by DNA damage.

The G4 at the p53 pA signal interacts with hnRNP H/F in a UV-dependent manner

In order to determine the protein factors that bind to the p53 G4 and may be involved in the mechanism by which the G4 affects 3'-end processing during DNA damage, we performed RNA affinity chromatography experiments. Since hnRNP H/F was shown by us (Millevoi et al. 2009) and others (Bagga et al. 1998; Alkan et al. 2006; Dalziel et al. 2007) to bind G-rich sequences and plays a role in 3'-end processing, we analyzed whether this factor was able to bind the G-rich regulatory element at the p53 pA signal. The G-rich sequence located downstream from the SV40 pA signal that binds hnRNP H/F (Bagga et al. 1998; Alkan et al. 2006) and forms a G4 (Supplemental Fig. 5) served as positive control. As shown in Figure 4A, hnRNP H/F was pulled down by both the p53 and SV40 G-rich RNAs. This interaction was almost completely abolished when using the G4M mutation of the p53 G-rich element or a SV40 RNA containing a G-to-C mutation in a G stretch that hampers G4 formation. Conversely, hnRNP A1, which has been shown previously to bind RNA G4 structures (Abdul-Manan et al. 1996), was not retained by either of the two RNAs. Using UV cross-linking assays followed by immunoprecipitation (IP) with the hnRNP H/F antibody, we showed that the hnRNP H/F-G4 interaction is the result of a direct contact between the protein factor and the RNA folded in a G4 (Fig. 4B). This effect is not specific of the p53 G4 since it occurred also for the SV40 RNA (Fig. 4B). Addition of G4-stabilizing compound TMPyP4 to the nuclear extract increased the interaction between hnRNP H/F and the p53 G4, further confirming that the functional G-rich element at the p53 pA signal, when structured in a G4 conformation, interacts with hnRNP H/F in a specific manner (Fig. 4C). Finally, we tested whether DNA damage inducers affect the hnRNP H/F-G4 interaction at the p53 pA signal. Using UV cross-linking/IP assays with nuclear extract from A549 exposed or not to UV radiation, we showed that DNA damage increased the binding of hnRNP H/F to the p53 G4 (Fig. 4C). This effect is specific to the p53 G4 only, since it did not occur for the TBP RNA (Fig. 4C). To determine whether DNA damage stimulates

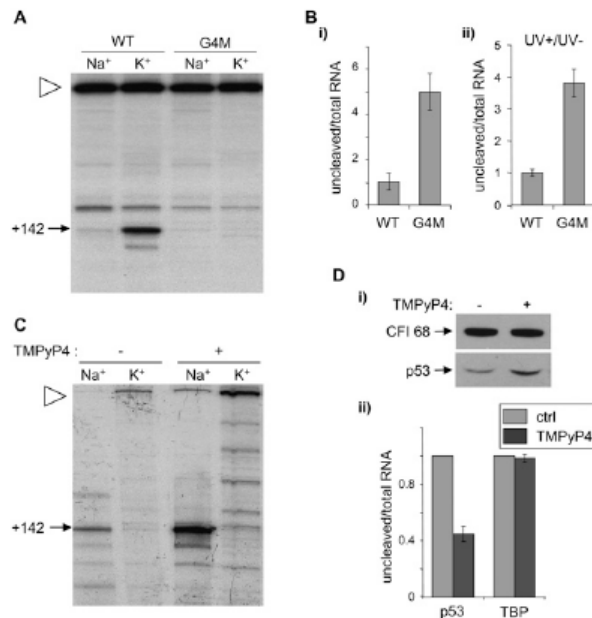


Figure 3. A G4 structure at the p53 pA signal is required to counteract UV-induced 3'-end processing inhibition. (A) Cation-dependent termination of reverse transcription at p53 RNA wild-type or mutated G4 (G4M). Strong pauses of reverse transcriptase are indicated by arrowheads, with their positions within the p53 RNA 3'-end sequence. Full-length extension products are indicated by open triangles. (B) Transfection of the wild-type or G4M *luc/p53* reporter plasmids in A549 cells left untreated (graph i) or UV-irradiated (graph ii), followed by analysis of 3'-end processing efficiency by RT-qPCR. (C) Cation-dependent pauses of reverse transcriptase at the p53 G4 (indicated by arrowheads), depending on TMPyP4 addition. (D) Western blot analysis of p53 protein accumulation (graph i) and RT-qPCR analysis of endogenous p53 3'-end processing efficiency (graph ii) in either TMPyP4-treated or untreated A549 cells.

hnRNP H/F rescues p53 activation during DNA damage

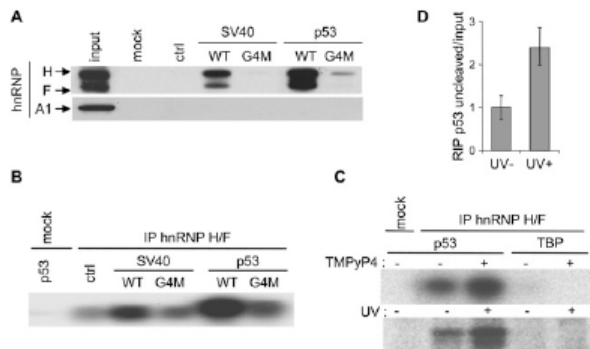


Figure 4. The G4 at the p53 pA signal interacts with hnRNP H/F in a UV-dependent manner. (A) RNA affinity chromatography using the G-rich element at the p53 or SV40 late pA signal, either wild-type or mutated (G4M), in the last G stretch and HeLa nuclear extracts, followed by Western blot analysis with the hnRNP H/F or hnRNP A1 antibodies. (B) UV cross-linking assays using the various RNAs as in A, followed by IP with the hnRNP H/F antibody. (C, top panel) UV cross-linking of the p53 or TBP RNAs in HeLa nuclear extracts, either TMPyP4-treated or untreated, followed by IP with the hnRNP H/F antibody. (Bottom panel) UV cross-linking of p53 or TBP RNA in nuclear extracts from A549 cells, either UV irradiated or left untreated, followed by IP with the hnRNP H/F antibody. (D) IP of the in vivo cross-linked RNA–protein complexes (RIP) in A549 cells, UV-irradiated or mock-treated, using the hnRNP H/F antibody followed by RT-qPCR analysis of the p53 pre-mRNA in the input and the RIP samples.

hnRNP H/F binding to the p53 transcript in vivo, we performed an RNP IP (RIP) assay using cell extracts from A549 cells UV-irradiated or left untreated and monitored the endogenous p53 primary transcript contained in the IPs. For this purpose, IPs were carried out with A549 cell lysates after formaldehyde cross-linking to specifically assay for in vivo RNA–protein interactions, with antibodies directed against hnRNP H/F. As shown in Figure 4D, endogenous p53 pre-mRNA was enriched in IP samples derived from UV-irradiated cells compared with mock-treated samples. Taken together, these results reveal that, following DNA damage, hnRNP H/F more closely associates with the G4 at the 3' end of the p53 primary transcript. Our results also show that TMPyP4 increases p53 3'-end processing efficiency through the stabilization of the p53 G4 and, as a consequence, the hnRNP H/F–p53 G4 interaction, contributing to p53 protein accumulation (Figs. 3, 4). These findings, together with previous observations showing increased p53 expression by G4 ligands (Lixia et al. 2008), prompt us to propose that, in addition to inhibiting telomerase activity, G4 ligands stimulate p53 accumulation via the 3'-end processing step, and p53 in turn represses telomerase activity directly (by interacting with the telomerase) (Li et al. 1999) or indirectly (by inhibiting telomerase mRNA synthesis) (Kanaya et al. 2000; Xu et al. 2000). Thus, the process generating mRNA 3' ends could be involved in the anti-tumor action of G4-interacting ligands via the p53 pathway.

hnRNP H/F regulates p53 expression and function in apoptosis during DNA damage

We next investigated the functional importance of the DNA damage-dependent interaction between hnRNP H/F and the G4 at the p53 pre-mRNA 3' end. To this end, we

performed siRNA-mediated depletion of hnRNP H, hnRNP F, or both proteins in A549 cells exposed or not to UV irradiation, and measured p53 expression at the protein and 3'-end processing levels. As shown in Figure 5A, siRNA-mediated silencing of hnRNP H and/or hnRNP F reduced p53 protein expression in UV-irradiated but not mock-treated cells. Similar results were obtained by exposing different cell lines to doxorubicin (Supplemental Fig. S6), suggesting that hnRNP H/F depletion counteracts DNA damage-dependent p53 accumulation regardless of cell type and the DNA-damaging agent. It is important to note that, in agreement with previous studies (Decker et al. 2003; Fox and Stover 2009), exposure of A549 cells to UV irradiation resulted in an increased hnRNP H/F expression (Fig. 5A). Importantly, loss of p53 protein expression in DNA-damaged A549 cells depleted of hnRNP H

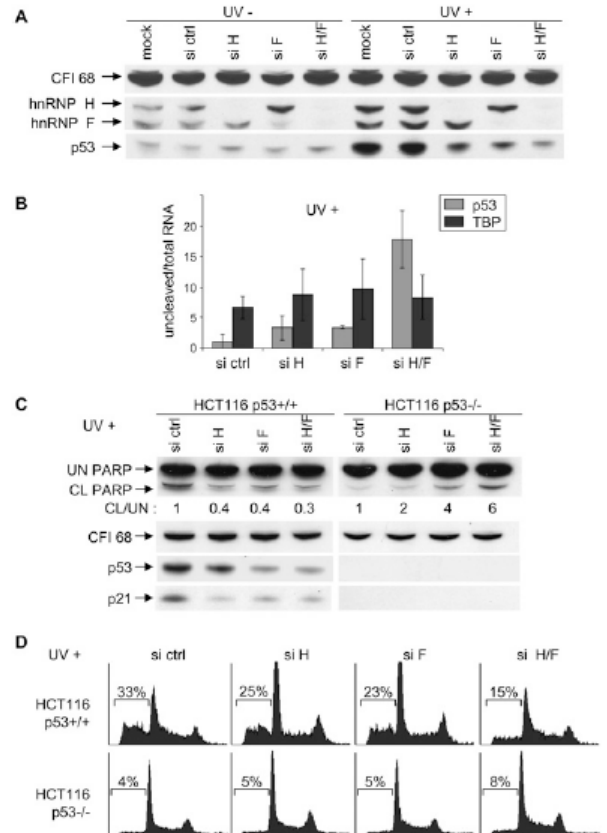


Figure 5. hnRNP H/F regulates p53 expression and function in apoptosis. (A) Western blot analysis of p53, hnRNP H/F, and CFI 68 (loading control) in A549 cells treated or not with siRNA control or siRNA against hnRNP H and/or hnRNP F, followed or not by UV irradiation. (B) RT-qPCR analysis of p53 and TBP pre-mRNA 3'-end processing efficiency in A549 cells treated or not with siRNA against hnRNP H and/or hnRNP F, followed by UV irradiation. (C) Western blot analysis of PARP uncleaved (UN) and cleaved (CL) p53, p21, and CFI 68 (loading control) in UV-irradiated HCT116 cells expressing p53 (HCT116 p53^{+/+}) or not (HCT116 p53^{-/-}) and treated with siRNA control or siRNA against hnRNP H and/or hnRNP F. (D) FACS analysis of HCT116 p53^{+/+} and p53^{-/-} cells treated with siRNA control or siRNA against hnRNP H and/or hnRNP F, followed by UV irradiation. The percentage of cells with a sub-G1 DNA content is indicated.

Decorsière et al.

and/or hnRNP F appeared to be due, at least in part, to a significant reduction in p53 pre-mRNA 3'-end processing efficiency, which reached a 15-fold maximum when both hnRNPs were silenced (Fig. 5B; Supplemental Fig. 7). This effect is specific to the p53 pre-mRNA, since processing of TBP 3' ends was not affected by siRNA-mediated depletion of hnRNP H/F (Fig. 5B). It is important to note that the effect of hnRNP H/F depletion on p53 3'-end processing is not due to altered p53 transcription (Supplemental Fig. 8) and does not depend on the p53 mutational status (Supplemental Fig. 9). These findings support a model whereby DNA damage-induced hnRNP H/F expression promotes its association with the p53 G4, resulting in maintained 3'-end processing efficiency. This can occur by recruiting factors that are essential for the cleavage and pA reaction; namely, CstF (Bagga et al. 1998; Chen and Wilusz 1998) and poly(A) polymerase (PAP) (Millevoi et al. 2009). Since DNA damage induces CstF sequestration in complexes with BRCA1/BARD1 (Kleiman and Manley 1999, 2001), RNA Pol II (Kleiman et al. 2005), and PARN (Cevher et al. 2010), which inhibit 3'-end processing, the function of hnRNP H/F in DNA-damaged cells could be to prevent CstF from being hijacked in alternative protein complexes. This hypothesis is strongly supported by our findings showing that hnRNP H increases CstF 64 binding to the p53 pA signal (Supplemental Fig. 10). Since hnRNP H/F affected p53 protein expression, and p53 has well-documented proapoptotic activities, we asked whether this effect had consequences on p53 function in apoptosis. To test this hypothesis, we monitored PARP cleavage, a hallmark of apoptotic cell death, in the human HCT116 colon carcinoma cell line expressing (HCT116 p53^{+/+}) or not (HCT116 p53^{-/-}) p53. The isogenic cell lines were first treated with the siRNAs against hnRNP H and/or hnRNP F—or with a negative control siRNA—and then exposed to UV irradiation. As shown in Figure 5C, depletion of hnRNP H and/or hnRNP F in UV-irradiated HCT116 cells (+/+) attenuated PARP cleavage and decreased expression of p53 and p21, a main transcriptional target of p53. Importantly, the effect of hnRNP H/F inhibition on PARP processing is p53-dependent, as it was not observed in p53-null HCT116 cells. The proapoptotic effect of hnRNP H/F silencing in HCT116 (p53^{-/-}) cells (Fig. 5C) is in agreement with previous data showing that hnRNP H is able to block MST2-dependent apoptosis by controlling the splicing of the A-Raf transcription factor (Rauch et al. 2010). The role of hnRNP H/F in UV-induced apoptosis was also evident by the increased percentage of cells with sub-G1 DNA content in HCT116 (p53^{+/+}) cells treated with siRNAs against hnRNP H and/or hnRNP F, indicative of apoptosis (Fig. 5D). Again, this effect is absent in HCT116 (p53^{-/-}) cells, further supporting the importance of hnRNP H/F in p53-dependent apoptosis. Overall, these results reveal that hnRNP H/F depletion reduced p53 expression mostly by affecting the 3'-end processing step, and this contributes to halting p53-dependent apoptosis induced by UV irradiation. Notably, hnRNP H/F has been shown to influence the expression of apoptotic genes by acting mainly on the splicing process (Garneau et al. 2005; Camats et al. 2008; Fox and Stover 2009; Rauch et al. 2010), and its expression is modified in response to genotoxic agents (Decker et al. 2003; Fox and Stover 2009; Rauch et al. 2010). Although the mechanistic details of hnRNP H/F activation under DNA damage awaits further investigations, our study extends the notion of a link between hnRNP H/F and apoptosis to the process

generating 3' ends, and unraveled a new level of complexity in the p53 regulation of expression in response to genotoxic stress.

In conclusion, in this study, we provide evidence for the existence of a specific compensatory mechanism that (1) enables tumor suppressor p53 pre-mRNA to be properly processed at the 3' end following DNA damage stress, and (2) contributes to p53-dependent apoptosis. The proposed rescue mechanism of p53 pre-mRNA 3'-end processing and function by the hnRNP H/F-G4 interaction provides fertile ground for further investigation and, together with previous findings (Burns and Richter 2008), supports the notion that p53 3'-end processing can control its expression and biological activity.

Materials and methods

Cell culture, treatment, and transfection

Subconfluent A549, MCF-7, or HCT116 cells cultured in RPMI, DMEM, and McCoy's media, respectively, with 10% fetal bovine serum were exposed to UV light (254 nm) at 50 J/m² or 1 mg/mL doxorubicin for 16 h. TMPyP4 was added at a final concentration of 20 μM for 24 h. Plasmid DNA (1 μg) was transfected for 24 h in 60-mm plates with the JetPEI transfection reagent (Polyplus Transfection). siRNAs (2.5 nM) against hnRNP H, hnRNP F (Alkan et al. 2006), and a control siRNA (5'-GGUCCGGUCCUCCCAAAUGtt-3') (Eurogentec) were transfected with Lipofectamine RNAiMAX (Invitrogen) for 72 h.

pre-mRNA 3'-end processing efficiency analysis

RNA (5 μg) from nuclei pellet isolated with the TRIzol RNA reagent (Invitrogen) was reverse transcribed using SuperScript III Reverse Transcriptase (Invitrogen) and the random hexamer protocol. Real-time PCR was performed using a LightCycler system (Roche) with DNA MasterPLUS SYBR Green I kit (Roche). Data were then analyzed as 2exp(-Ct total RNA - Ct uncleaved). RPA was performed as in Dalziel et al. (2007) using 5 μg of nuclear RNA and ~60,000 cpm (counts per minute) [³²P]UTP radiolabeled probe.

Cell cycle analysis

Attached cells were trypsinized, combined with floating cells, stained with propidium iodide (Sigma), and analyzed in a FACScan flow cytometry system using the CellFit software (Becton Dickinson). Data were obtained from 10⁴ viable cells.

Acknowledgments

We thank Sandra Bernat for technical help, and Olivier Sordet, Anne Pradines, and Anne-Françoise Tilkin-Mariamé for materials and discussions. We thank INSERM, Institut Claudius Regaud, FRM [Equipe FRM, Soutenu par la Fondation Recherche Médicale], and Fondation RITC [Recherche et Innovation Thérapeutique en Cancérologie] for financial support. A.D. was a recipient of a studentship from ARC [Association pour la Recherche sur le Cancer] and FRM.

References

- Abdul-Manan N, O'Malley SM, Williams KR. 1996. Origins of binding specificity of the A1 heterogeneous nuclear ribonucleoprotein. *Biochemistry* 35: 3545-3554.
- Alkan SA, Martincic K, Milcarek C. 2006. The hnRNPs F and H2 bind to similar sequences to influence gene expression. *Biochem J* 393: 361-371.
- Bagga PS, Arhin GK, Wilusz J. 1998. DSEF-1 is a member of the hnRNP H family of RNA-binding proteins and stimulates pre-mRNA cleavage and polyadenylation in vitro. *Nucleic Acids Res* 26: 5343-5350.
- Bruno T, De Nicola F, Iezzi S, Lecis D, D'Angelo C, Di Padova M, Corbi N, Dimizzi L, Zannini L, Jekimovs C, et al. 2006. Che-1 phosphorylation by ATM/ATR and Chk2 kinases activates p53 transcription and the G2/M checkpoint. *Cancer Cell* 10: 473-486.

hnRNP H/F rescues p53 activation during DNA damage

- Burns DM, Richter JD. 2008. CPEB regulation of human cellular senescence, energy metabolism, and p53 mRNA translation. *Genes Dev* 22: 3449–3460.
- Camats M, Guil S, Kokolo M, Bach-Elias M. 2008. P68 RNA helicase (DDX5) alters activity of *cis*- and *trans*-acting factors of the alternative splicing of H-Ras. *PLoS ONE* 3: e2926. doi: 10.1371/journal.pone.0002926.
- Cevher MA, Zhang X, Fernandez S, Kim S, Baquero J, Nilsson P, Lee S, Virtanen A, Kleiman FE. 2010. Nuclear deadenylation/polyadenylation factors regulate 3' processing in response to DNA damage. *EMBO J* 29: 1674–1687.
- Chen F, Wilusz J. 1998. Auxiliary downstream elements are required for efficient polyadenylation of mammalian pre-mRNAs. *Nucleic Acids Res* 26: 2891–2898.
- Dalziel M, Nunes NM, Furger A. 2007. Two G-rich regulatory elements located adjacent to and 440 nucleotides downstream of the core poly(A) site of the intronless melanocortin receptor 1 gene are critical for efficient 3' end processing. *Mol Cell Biol* 27: 1568–1580.
- Dankwardt S, Hentze MW, Kulozik AE. 2008. 3' end mRNA processing: Molecular mechanisms and implications for health and disease. *EMBO J* 27: 482–498.
- Decker ED, Zhang Y, Cocklin RR, Witzmann FA, Wang M. 2003. Proteomic analysis of differential protein expression induced by ultraviolet light radiation in HeLa cells. *Proteomics* 3: 2019–2027.
- Fox JT, Stover PJ. 2009. Mechanism of the internal ribosome entry site-mediated translation of serine hydroxymethyltransferase 1. *J Biol Chem* 284: 31085–31096.
- Gameau D, Revil T, Fiset JF, Chabot B. 2005. Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J Biol Chem* 280: 22641–22650.
- Gomes NP, Bjerke G, Llorente B, Szostek SA, Emerson BM, Espinosa JM. 2006. Gene-specific requirement for P-TEFb activity and RNA polymerase II phosphorylation within the p53 transcriptional program. *Genes Dev* 20: 601–612.
- Kanaya T, Kyo S, Hamada K, Takakura M, Kitagawa Y, Harada H, Inoue M. 2000. Adenoviral expression of p53 represses telomerase activity through down-regulation of human telomerase reverse transcriptase transcription. *Clin Cancer Res* 6: 1239–1247.
- Kim HS, Li H, Cevher M, Parmelee A, Fonseca D, Kleiman FE, Lee SB. 2006. DNA damage-induced BARD1 phosphorylation is critical for the inhibition of messenger RNA processing by BRCA1/BARD1 complex. *Cancer Res* 66: 4561–4565.
- Kleiman FE, Manley JL. 1999. Functional interaction of BRCA1-associated BARD1 with polyadenylation factor CstF-50. *Science* 285: 1576–1579.
- Kleiman FE, Manley JL. 2001. The BARD1–CstF-50 interaction links mRNA 3' end formation to DNA damage and tumor suppression. *Cell* 104: 743–753.
- Kleiman FE, Wu-Baer F, Fonseca D, Kaneko S, Baer R, Manley JL. 2005. BRCA1/BARD1 inhibition of mRNA 3' processing involves targeted degradation of RNA polymerase II. *Genes Dev* 19: 1227–1237.
- Kruse JP, Gu W. 2009. Modes of p53 regulation. *Cell* 137: 609–622.
- Levine AJ, Oren M. 2009. The first 30 years of p53: Growing ever more complex. *Nat Rev Cancer* 9: 749–758.
- Li H, Cao Y, Berndt MC, Funder JW, Liu JP. 1999. Molecular interactions between telomerase and the tumor suppressor protein p53 in vitro. *Oncogene* 18: 6785–6794.
- Lipps HJ, Rhodes D. 2009. G-quadruplex structures: In vivo evidence and function. *Trends Cell Biol* 19: 414–422.
- Lixia G, Fei Y, Jiajia J, Jianhui L. 2008. Triethylene tetramine, a novel ligand of G-quadruplex, induces senescence of MCF-7 cells. *Biotechnol Lett* 30: 47–53.
- Mazan-Mamczarz K, Galban S, Lopez de Silanes I, Martindale JL, Atasoy U, Keene JD, Gorospe M. 2003. RNA-binding protein HuR enhances p53 translation in response to ultraviolet light irradiation. *Proc Natl Acad Sci* 100: 8354–8359.
- Millevoi S, Vagner S. 2010. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res* 38: 2757–2774.
- Millevoi S, Decorsiere A, Loulergue C, Iacovoni J, Bernat S, Antoniou M, Vagner S. 2009. A physical and functional link between splicing factors promotes pre-mRNA 3' end processing. *Nucleic Acids Res* 37: 4672–4683.
- Mirkin N, Fonseca D, Mohammed S, Cevher MA, Manley JL, Kleiman FE. 2008. The 3' processing factor CstF functions in the DNA repair response. *Nucleic Acids Res* 36: 1792–1804.
- Ofir-Rosenfeld Y, Boggs K, Michael D, Kastan MB, Oren M. 2008. Mdm2 regulates p53 mRNA translation through inhibitory interactions with ribosomal protein L26. *Mol Cell* 32: 180–189.
- Rauch J, O'Neill E, Mack B, Matthias C, Munz M, Kolch W, Gires O. 2010. Heterogeneous nuclear ribonucleoprotein H blocks MST2-mediated apoptosis in cancer cells by regulating A-Raf transcription. *Cancer Res* 70: 1679–1688.
- Sundquist WI, Klug A. 1989. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* 342: 825–829.
- Takagaki Y, Ryner LC, Manley JL. 1989. Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes Dev* 3: 1711–1724.
- Takagi M, Absalon MJ, McLure KG, Kastan MB. 2005. Regulation of p53 translation and induction after DNA damage by ribosomal protein L26 and nucleolin. *Cell* 123: 49–63.
- Xu D, Wang Q, Gruber A, Bjorkholm M, Chen Z, Zaid A, Selivanova G, Peterson C, Wiman KG, Pisa P. 2000. Downregulation of telomerase reverse transcriptase mRNA expression by wild type p53 in human tumor cells. *Oncogene* 19: 5123–5133.

SUPPLEMENTAL MATERIAL

Essential role for the interaction between hnRNP H/F and a G-quadruplex in maintaining p53 pre-mRNA 3' end processing and function during DNA damage. Adrien Decorsière, Anne Cayrel, Stéphan Vagner and Stefania Millevoi

Supplemental Figure Legends

Supplemental Figure 1 (related to Fig. 1): p53 activation following DNA damage.

(A) Western blot showing p53 accumulation in A549, HCT116 and MCF7 cells left untreated (ctrl), UV irradiated (UV) or treated with doxorubicin (doxo). CFI 68 was used as loading control.

Supplemental Figure 2 (related to Fig. 1): The uncleaved p53 and TBP transcripts do not correspond to alternatively polyadenylated mRNAs.

Reverse transcription analysis primed with oligo(dT) or random oligonucleotides followed by PCR with primers to amplify the uncleaved (UN) and total (tot) p53 or TBP cDNA products, as depicted.

Supplemental Figure 3 (related to Fig. 1): Analysis of the cleaved to total RNA ratio confirms that p53 3' end processing is maintained during DNA damage.

Reverse transcription of nuclear RNA from A549 cells left untreated (ctrl), UV irradiated (UV) or treated with doxorubicin (doxo), followed by real-time PCR with primers specific to either p53 or TBP located upstream of the cleavage site to amplify the cleaved and the total RNA using as template cDNA from oligo (dT) and random primed reverse transcription, respectively, as depicted.

Supplemental Figure 4 (related to Fig. 3): Structure-function relationship of additional G-rich mutants in p53 3' end processing during DNA damage.

(A) DNA sequence of the G-rich element containing the G4 G-stretches (in bold) wild-type (WT) or mutated (small underlined caps) to prevent the G4 formation (G4M and G4D) or to leave this structure unaffected (G4S). (B) Cation-dependent termination of

reverse transcription at the +142 position in the p53 RNA WT or mutated G4 (G4M,G4S and G4D) showing the presence of strong pauses with K⁺ for the WT and G4S but not for the G4M and G4D RNAs. (C) Transfection of the WT, G4M, G4S or G4D RLuc/p53 reporter plasmids in A549 cells, followed by measurement of the RLuc/Fluc activity (i) or RT-qPCR analysis of 3' end processing efficiency (ii) (D) RLuc/Fluc activity (i) and RT-qPCR analysis of 3' end processing efficiency (ii) as described in (C) using protein or RNA extracted from A549 cells exposed (UV+) or not (UV-) to UV irradiation.

Supplemental Figure 5 (related to Fig. 4): The G-rich element at the SV40 polyadenylation signal element folds in a G-quadruplex structure. (A) DNA sequence of the SV40 G-rich element (GRS) containing the G4 forming G-stretches (in bold) wild-type (WT) or mutated (small underlined caps) to prevent the G4 formation (G4M), (B) Cation-dependent termination of reverse transcription at SV40 RNA WT or mutated G4 (G4M). Strong pauses of reverse transcriptase are indicated by an arrowhead with its position relative to the cleavage site. Full-length extension products are indicated by an open triangle.

Supplemental Figure 6 (related to Fig. 5): hnRNP H/F depletion counteracts p53 accumulation after doxorubicin treatment. Western blot analysis of p53 and CFI 68 (loading control) in A549, HCT116 and MCF-7 cells treated with siRNA control or siRNAs against hnRNP H and/or hnRNP F, followed by doxorubicin (Doxo+) treatment.

Supplemental Figure 7 (related to Fig. 5): Analysis of the p53 to TBP mRNA ratio confirms p53 3' end processing inhibition by hnRNP H/F depletion. Oligo (dT) primed reverse transcription of nuclear RNA from A549 cells treated with siRNA control (si ctrl) or siRNAs against hnRNP H and/or hnRNP F and UV irradiated, followed by real-time PCR with primers specific to either p53 or TBP located upstream of the cleavage site to amplify the cleaved and polyadenylated RNAs.

Supplemental Figure 8 (related to Fig. 5): hnRNP H/F depletion does not affect p53 transcription. (A) ChIP analysis of RNA polymerase II occupancy of the p53 or p21 promoter using nuclear extracts from A549 cells treated with siRNA control (si ctrl) or siRNAs against hnRNP H and hnRNP F (si H/F) , followed by UV irradiation. Values are expressed as signal to noise ratio. (B) Reverse transcription of nuclear RNA from A549 cells treated with siRNA control (si ctrl) or siRNAs against hnRNP H and hnRNP F (si H/F) and UV irradiated, followed by real-time PCR analysis of the p53 pre-mRNA to TBP total RNA ratio using primers as depicted.

Supplemental Figure 9 (related to Fig. 5): The regulation of p53 expression by hnRNP H/F does not depend on the p53 mutational status. Random primed reverse transcription of nuclear RNA from HaCaT cells (harbouring two mutant p53 alleles) treated with siRNA control (si ctrl) or siRNAs against hnRNP H and hnRNP F (si H/F) and UV irradiated, followed by real-time PCR with primers specific to either p53 or TBP located on either side of the cleavage site (CA) (to amplify the uncleaved RNA) or upstream of the cleavage site (to amplify both the uncleaved and cleaved RNA), as depicted in Fig. 1. (B) Western blot analysis of p53, hnRNP H/F and CFI 68 (loading control) in HaCaT cells treated with siRNA control (si ctrl) and siRNAs against hnRNP H and hnRNP F (si H/F), followed by UV irradiation.

Supplemental Figure 10 (related to Fig. 5): hnRNP H increases CstF 64 binding at the p53 polyadenylation signal. UV crosslinking of the uniformly *in vitro* transcribed ³²P-labeled p53 RNAs using HeLa nuclear extracts in the absence or in the presence of recombinant purified hnRNP H, followed by IP with the CstF 64 antibody or with the hnRNP H antibody, and SDS-PAGE analysis.

Supplemental Materials and Methods

RNA structural-probing

In vitro transcribed RNA (4 pmol) was annealed with 10^5 cpm of ATP-labeled primer (5'-AGGGAGAGAACCTTGTGCCA-3'), followed by reverse transcription with 5 U of AMV reverse transcriptase (Promega) as in (Brunel and Romby 2000). Where indicated, TMPyP4 was added to a final concentration of 20 μ M in the annealing buffer.

Plasmid construction

The p53 3' end region was generated by PCR on genomic DNA of MCF7 cell using the primers F-WTp53 (5'-ttactagtcacactgaagtcacaaaagg-3') with R-WTp53 (5'-ttagatctagggagagaaggttgcca-3'), followed by enzymatic digestion with SpeI and BglII. The amplified fragments were cloned in pRL-CMV vector using XbaI and BamHI sites to generate the Rluc/p53 plasmid containing the p53 3' end region in place of SV40 polyadenylation signal. The mutants G4M were generated by a splicing PCR from the Rluc/p53 using the primers F-WTp53 with R-G4M (5'-tgtgtcccgaacccgggact-3') and F-G4M (5'-cccgggttcgggacacaaacaaaggctggagactggg-3') with R-WTp53.

Luciferase Activities

24 h after transfection cells rinsed twice with 1X PBS were harvested and lysed in 1X Passive Lysis Buffer (Promega). Reporter measurement assays were done with the Dual-Luciferase Reporter Assay System (Promega) according to manufacturer's instructions.

Western Blot Analysis

Protein analyses by Western blot on whole cell lysates were performed with mAbs against hnRNP A1 (4B10; Abcam), hnRNP H/F (1G11; Abcam), p53 (DO-7, DakoCytomation), P21 (SX118, BD Pharmingen), CstF 64 (3A1; kindly provided by C. McDonald), a polyclonal antibody against PARP (Cell Signaling), a polyclonal antibody against CFI 68 (Bethyl).

Subcellular Fractionation

This experiment was performed as described previously (Lewis et al. 2007). ~30 millions cells were washed in PBS, resuspended in 400 µl of buffer A (10 mM HEPES-KOH, pH 7.5, 10 mM KCl, 1 mM DTT, 1 mM PMSF) containing RNase inhibitors, were incubated on ice for 15 min. Nonidet P-40 was added to a final concentration of 0.3% and cells were incubated for an additional 10 min on ice. Nuclei were pelleted by centrifugation at $1500 \times g$ for 5 min, and the cytoplasmic fraction was collected to a new tube and clarified by centrifugation at $13,000 \times g$ for 15 min. Nuclear or cytoplasmic RNAs were isolated with the TRIzol RNA reagent (Invitrogen) according to the manufacturer's protocol.

In vitro RNA-protein interactions

Nuclear extract (10 µg) were combined with in vitro transcribed ^{32}P -labeled RNAs (150,000 cpm), UV irradiated and immunoprecipitated with the hnRNP H/F antibody (1G11; Abcam) as in (Millevoi et al. 2006). Where indicated, TMPyP4 (0.5µM) or recombinant hnRNP H (purified as described in Millevoi et al. 2009) was added to the nuclear extract. RNA affinity chromatography was performed by combining *in vitro* transcribed Biotin-11-UTP-labelled RNA with 100 µg of nuclear extract from HeLa cells and 20 µl of streptavidin acrylamide beads (Pierce Chemical, Rockford, IL) in a buffer containing 100 mM KCl, 20 mM Hepes pH 7, 0.2 mM EDTA, 3 mM DTT, 0.05% Nonidet P-40, for 1 h at 4°C. The beads were collected by centrifugation, washed four times with 1 ml of the binding buffer, resuspended in 10 µl of loading buffer, and boiled for 5 min. After centrifugation, the supernatant was loaded on to an SDS-polyacrylamide gel electrophoresis (PAGE) gel and analyzed by Western Blot.

RNA Immunoprecipitation (RIP)

Cells ($\sim 10^7$) were harvested and ribonucleoprotein complexes were cross-linked by adding formaldehyde to the cell culture medium to a final concentration of 1% and incubating the mixture for 10 minutes at room temperature. The cross-linking reaction was stopped by adding 125 mM glycine for 5 min at room temperature. Cells were washed twice with ice-cold PBS and resuspended in 1 ml of RIPA buffer (50 mM Tris HCl pH 7.4, 1% Nonidet P-40, 0.5% NaDeoxycolate, 0.05% SDS, 1 mM EDTA,

RESULTATS

150mM NaCl). Cellular extracts were precleared with 100 µg of E.coli tRNA for 1 h. For immunoprecipitation, 40µl protein G sepharose beads were preincubated 1 h at 4°C with protease inhibitor cocktail (Roche) and 5 µl of hnRNP H/F mAb (α1G11, Abcam) in 1ml of RIPA buffer. 200 µl of precleared extract was incubated with washed protein G sepharose beads, for 4 h at 4°C. After washing, cross-linked immunoprecipitates were reverse cross-linked by heating at 70°C 45 min in REV buffer (25mM Tris pH 7, 2.5 mM EDTA, 5mM DTT, 0.5% SDS). The immunoprecipitated RNA was isolated with the TRIzol RNA reagent (Invitrogen) according to the manufacturer's protocol and real-time RT-PCR quantification. Data were then analyzed as $2^{\text{exp}(\text{Ct input} - \text{Ct IP})}$.

Chromatin immunoprecipitation (ChIP)

ChIP assays were performed as in (Tyteca et al. 2006) with the following modifications. In all, 150 µg of chromatin was immunoprecipitated using 2 µg of rabbit polyclonal antibody against RNA polymerase II (N-20; Santa Cruz sc-899) or without antibody (mock). After washing, the beads-chromatin complexes were resuspended in TE buffer (10 mM Tris, 0.5 mM EDTA pH 8.0) and formaldehyde crosslinking was reversed in the presence of 0.5% SDS, overnight at 70°C. Immunoprecipitated and input DNA was phenol/chlorophorm extracted, precipitated and analyzed in duplicate by real-time PCR analysis using primers indicated below.

Real-time PCR oligonucleotides:

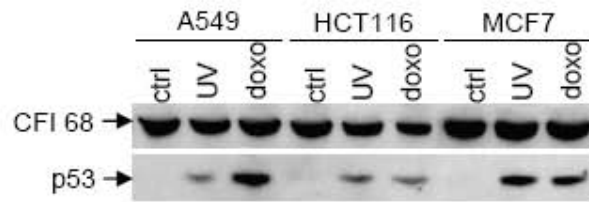
qPCR			
oligonucleotides	PCR product	forward (5'→3')	reverse (5'→3')
3' end processing	p53 total	AGGCGATCCACCTGTCTCA	CAGATGTGCTTGCAGAATGT
	p53 uncleaved	AGGCGATCCACCTGTCTCA	TAGCCTGCACTGGCGTTC
	TBP total	GGAAGGGGCATTATTTGTG	GCCCAGATAGCAGCACGGTA
	TBP uncleaved	CAAGTTGGTTTGAGGGAGAAA	CAGTATGATCACATGACTCTTACAAGG
ChIP RNA polII	promoter p53	GGCGGATTA CTGCCCTTAC	TGGACGGTGGCTCTAGACTT
	promoter p21	TATATCAGGGCCGCGCTG	GGCTCCACAAGGAACTGACTTC
transcript analysis	p53 pre-mRNA	GGCCTCTGATTCTCACTGATTG	TCTGTCATCCAAATACTCCACACG

Supplemental References

- Brunel, C. and Romby, P. 2000. Probing RNA structure and RNA-ligand complexes with chemical probes. *Methods Enzymol* **318**: 3-21.
- Lewis, S.M., Veyrier, A., Hosszu Ungureanu, N., Bonnal, S., Vagner, S., and Holcik, M. 2007. Subcellular relocalization of a trans-acting factor regulates XIAP IRES-dependent translation. *Mol Biol Cell* **18**(4): 1302-1311.
- Millevoi, S., Bernat, S., Telly, D., Fouque, F., Gladieff, L., Favre, G., Vagner, S., and Toulas, C. 2009. The c.5242C>A BRCA1 missense variant induces exon skipping by increasing splicing repressors binding. *Breast Cancer Res Treat* **120**(2): 391-399.
- Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S.Z., Keller, W., Antoniou, M., and Vagner, S. 2006. An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. *Embo J* **25**(20): 4854-4864.
- Tyteca, S., Vandromme, M., Legube, G., Chevillard-Briet, M., and Trouche, D. 2006. Tip60 and p400 are both required for UV-induced apoptosis but play antagonistic roles in cell cycle progression. *EMBO J* **25**(8): 1680-1689.

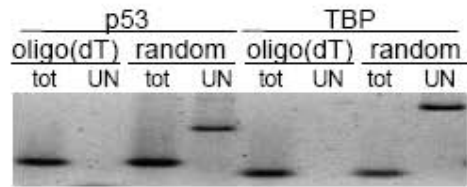
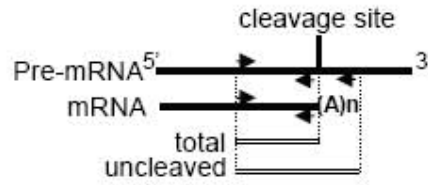
RESULTATS

Decorsière et al., Supplemental Fig. 1



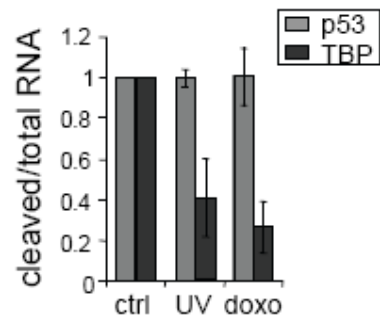
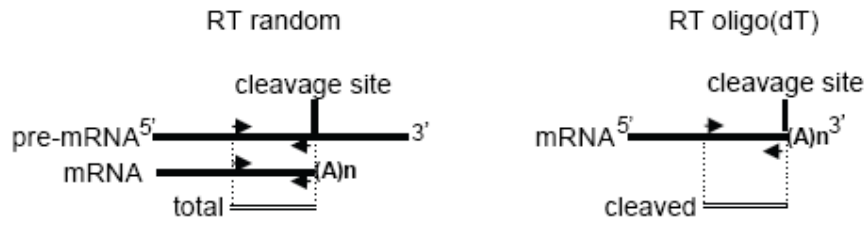
RESULTATS

Decorsière et al., Supplemental Fig. 2



RESULTATS

Decorsière et al., Supplemental Fig. 3

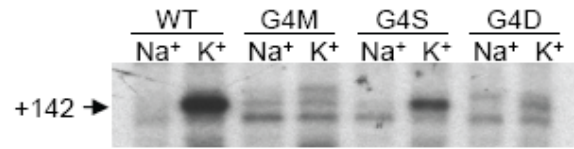


A

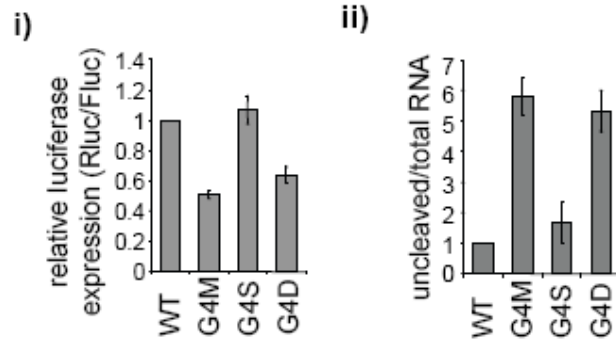
```

WT      ATGAGTGTGGGCTGGGGGGGCCAGTGCCCGGGTTCCGGGAGGGGAACAA
G4M     ATGAGTGTGGGCTGGGGGGGCCAGTGCCCGGGTTCGGGAAcacaAACAA
G4S     ATGAGTGTGGGCTGGGGGGGCCCcacaCCCGGGTTCCGGGAGGGGAACAA
G4D     ATGAGTGTGGGCTcacacaCCAGTGCCCGGGTTCCGGGAGGGGAACAA
    
```

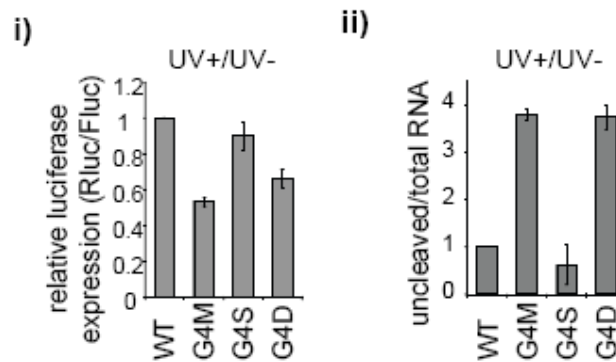
B



C



D



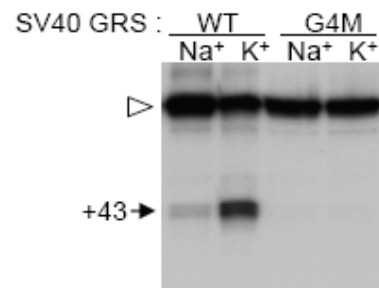
RESULTATS

Decorsière et al., Supplemental Fig. 5

A

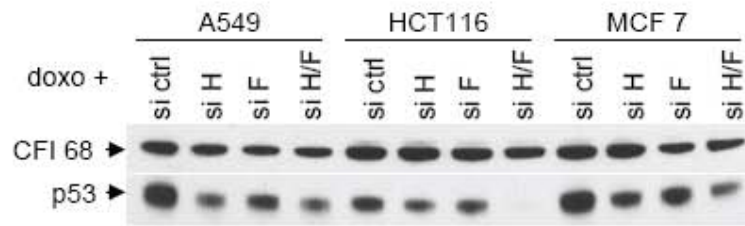
WT TTCAGGGGGAGGTGTGGGAGGTTTTT
G4M TTCccccAGGTGTGGGAGGTTTTT

B



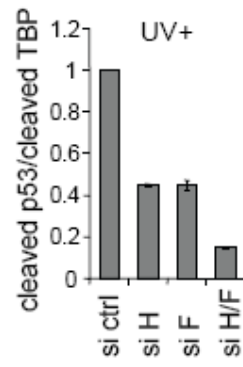
RESULTATS

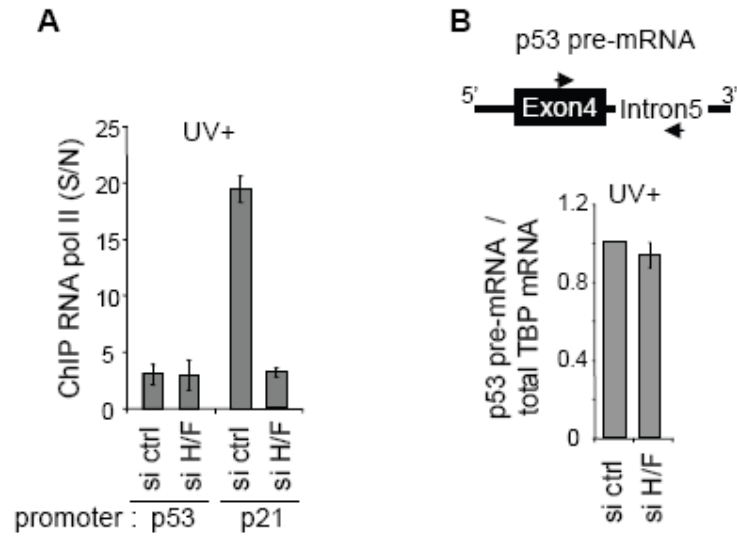
Decorsière et al., Supplemental Fig. 6



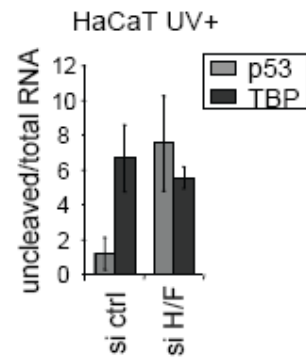
RESULTATS

Decorsière et al., Supplemental Fig. 7

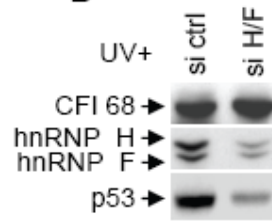




A

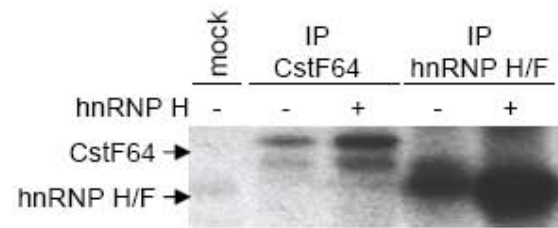


B



RESULTATS

Decorsière et al., Supplemental Fig. 10



I-5 Discussions et perspectives

Dans cette étude, nous avons mis en évidence que le gène codant le suppresseur de tumeur p53 n'est pas affecté par l'inhibition globale de la maturation en 3' lors de dommages à l'ADN. Le mécanisme de maintien de la maturation en 3' implique le recrutement des protéines hnRNP H et F sur une structure de type G-quadruplexe située en aval du signal de polyadénylation de p53. Cette interaction est nécessaire à l'accumulation de P53 et permet l'apoptose induite par p53. Ces résultats ainsi que des travaux antérieurs (Burns and Richter, 2008) montrent que la maturation en 3' de p53 peut réguler son expression protéique et son activité biologique.

I-5-1 Implication des hnRNP H/F dans l'apoptose

Notre étude a montré que lors de dommages à l'ADN, la déplétion des hnRNP H/F entraîne une diminution de l'apoptose induite par P53. A l'inverse, dans des cellules n'ayant pas de P53, la déplétion de hnRNP H/F a un effet pro-apoptotique.

Ces données s'intègrent à un ensemble d'études contradictoires concernant le rôle des hnRNP H/F dans l'induction de l'apoptose (Garneau et al., 2005; Rauch et al., 2011; Rauch et al., 2010). En effet, la déplétion de hnRNP H1 et H2 stimule l'apoptose induite par la kinase MST2 en absence de dommage à l'ADN (Rauch et al., 2010). La surexpression de hnRNP H1, dans des cellules n'exprimant pas P53, diminue l'apoptose lors de dommages à l'ADN (Rauch et al., 2011; Rauch et al., 2010). Dans ces études, où l'induction de l'apoptose par P53 n'est pas prise en compte, les hnRNP H/F ont un effet anti-apoptotique que nous retrouvons dans des cellules n'exprimant pas P53.

Il a également été montré que la déplétion de hnRNP H1 et F stimule la production de la forme anti-apoptotique BclXs de la protéines BCL2L1 et diminue la forme pro-apoptotique BclXI (Garneau et al., 2005). Dans cette étude, l'effet de la déplétion de hnRNP H1 et F sur l'apoptose n'est pas quantifié. En conclusion, il semble donc que les hnRNP H/F stimule l'épissage et la maturation en 3' de gènes pro ou anti-apoptotique. Il en résulte, en absence d'apoptose induite par P53, que les hnRNP H/F auraient un effet anti-apoptotique. Néanmoins, nous avons montré que la déplétion des hnRNP H/F entraîne une diminution de l'apoptose induite par P53.

I-5-2 Implication des hnRNP H/F dans la résistance à l'inhibition de la maturation en 3' lors de dommages à l'ADN

Nous avons montré que l'interaction entre les hnRNP H/F et la séquence capable de se structurer en G-quadruplexe en aval du pré-ARNm de p53 est indispensable pour la résistance à l'inhibition de la maturation en 3' lors de dommages à l'ADN.

Afin d'approfondir le rôle des hnRNP H/F sur la stimulation de la maturation en 3' de p53 ainsi que sur l'apoptose induite par P53, un ensemble d'expériences pourrait être réalisé afin de déterminer si le recrutement des hnRNP H/F en aval du signal de polyadénylation est suffisant pour résister à l'inhibition de la maturation en 3' lors de dommages à l'ADN. Nous pourrions utiliser une stratégie de « recrutement hétérologue » comme précédemment réalisé dans l'équipe (Millevoi et al., 2006). Sur le gène rapporteur contenant les séquences de polyadénylation de p53, le G-quadruplexe pourra être remplacé par une séquence r-17 qui interagit avec une haute affinité et spécificité avec la protéine phagique R-17. Ce rapporteur sera alors transfecté dans des cellules surexprimant les protéines de fusion R17-hnRNP H/F, et traitées ou non aux UV. Cette expérience pourra également être réalisée sur des gènes dont la maturation en 3' est inhibée lors de dommages à l'ADN tel que TBP. Ces expériences nous permettront de définir si le recrutement de hnRNP H/F est suffisant pour maintenir la maturation en 3' lors de dommages à l'ADN. De plus, ces expériences nous permettront de localiser les domaines des hnRNP H/F impliqués dans la stimulation de la maturation en 3' pour ensuite identifier plus précisément les partenaires protéiques éventuels tels que CstF et PAP (Bagga et al., 1998; Chen and Wilusz, 1998; Millevoi et al., 2009). Nous pourrions également étudier l'effet de mutations dans les sites de phosphorylation de hnRNP H1 et H2 (Navakauskiene et al., 2004) et d'acétylation de hnRNP F (Choudhary et al., 2009).

I-5-3 Les G-quadruplexes : cible thérapeutique

Notre étude a montré qu'une drogue, le TMPyP4 (porphyrin 5,10,15,20-tetra(N-méthyl-4-pyridyl)porphin) connue pour interagir avec les G-quadruplexes

(Arora and Maiti, 2009; Ofer et al., 2009), augmente la structuration du pré-ARNm de p53 en G-quadruplexe. Le traitement de lignées cellulaires avec cette drogue accroît le recrutement des hnRNP H/F sur la région en aval du signal de polyadénylation du pré-ARNm de p53. De plus, l'ajout de cette drogue sur des cellules entraîne une plus forte expression de la protéine P53 en partie due à une augmentation de l'efficacité de maturation en 3' du pré-ARNm.

La localisation et les conséquences de la structuration de ces G-quadruplexes en font une cible moléculaire pour de nombreuses drogues anticancéreuses (Neidle, 2009; Yang and Okamoto, 2010). En effet, la structuration de ces G-quadruplexes peut empêcher l'élongation de la taille des télomères en inhibant l'activité de la télomérase (Zahler et al., 1991). La structuration de ces G-quadruplexes au niveau des promoteurs de nombreux oncogènes (Brooks et al., 2010; Eddy and Maizels, 2006) diminue la transcription des gènes qu'ils contrôlent (Ou et al., 2007; Sun et al., 2008).

La majorité de ces drogues cible des séquences d'ADN, avec différentes affinités pour les G-quadruplexes en fonction de leurs conformations et de leurs stabilités. Par exemple, la Telomestatin, un puissant inhibiteur de la télomérase, stabilise la structuration des G-quadruplexe de « type panier » retrouvés au niveau des télomères (Kim et al., 2002; Shin-ya et al., 2001). Certaines de ces drogues ont une cytotoxicité relativement faible et pourraient être utilisées dans des traitement anticancéreux. Notamment, la molécule Quarfloxacin® (CX-3543, Pharmaceuticals Cylene à San Diego, CA, USA) est actuellement en essai clinique de phase II (Drygin et al., 2009). Ces drogues anticancéreuses peuvent cibler préférentiellement des séquences d'ARN. Il a été montré que la drogue antitumorale MPP (Macrocyclic Pyridyl Polyoxazoles) interagit de manière préférentielle avec le G-quadruplexe contenue sur l'ARNm de Aurora plutôt qu'avec les G-quadruplexes des séquences d'ADN télomériques (Rzuczek et al., 2010).

Connaissant le rôle suppresseur de tumeur de P53, notre étude a donc permis de définir une nouvelle cible pour les drogues anticancéreuses visant les G-quadruplexes mais également à mieux caractériser les effets de ces drogues sur l'inhibition de la prolifération cellulaire.

I-5-4 Généralisation du mécanisme de résistance à l'inhibition de la maturation en 3' lors de dommages à l'ADN

Lors de dommages à l'ADN, l'activation de P53 permet la transcription et l'accumulation de gènes de réponse aux stress tel que p21. Cela suppose que ces gènes résistent à l'inhibition de la maturation en 3'. En effet, il a été montré que la maturation en 3' des pré-ARNm de p21, n'est pas affectée par les dommages à l'ADN, contrairement aux autres gènes tel que HPRT utilisé comme contrôle (Gomes et al., 2006). L'accumulation du facteur de transcription P53 permet la transcription de gènes indépendamment de la phosphorylation de la serine 2 de la CTD de la pol II ce qui permettrait de résister à l'inhibition de la maturation en 3' (Gomes et al., 2006).

Cependant, certains gènes de réponse aux dommages à l'ADN ne sont pas activés par une transcription dépendante de P53. Afin d'identifier d'autres pré-ARNm polyadénylés par le mécanisme que nous avons mis en évidence pour p53, il serait nécessaire de mieux appréhender l'implication des G-quadruplexes dans le maintien de l'efficacité de la maturation en 3' lors de dommage à l'ADN.

Le G-quadruplexe du pré-ARNm de p53 pourrait être placé à différentes positions en aval mais aussi en amont du site de clivage sur le pré-ARNm de TBP ou de p53 délété de son G-quadruplexe. Nous pourrions également réaliser ces expériences avec différentes séquences capables de se structurer en G-quadruplexe telles que les séquences en aval du signal de polyadénylation de SV40L ou encore une séquence synthétique. L'analyse de l'efficacité de la maturation en 3' lors de dommages à l'ADN de ces différentes constructions nous permettra de mieux comprendre l'implication de la structure en G-quadruplexe dans le mécanisme de résistance à l'inhibition de la maturation en 3'. En effet, nous déterminerons si la position et si la séquence des G-quadruplexes sont importantes au maintien de la maturation en 3', mais également si une structure en G-quadruplexe est suffisante pour permettre ce mécanisme.

L'ensemble de ces expériences nous permettra de mieux sélectionner d'autres gènes capables de résister à l'inhibition de la maturation en 3' lors de dommages à l'ADN par le même mécanisme que le pré-ARNm p53. Nous pourrions alors sélectionner différents gènes impliqués dans la réponse aux dommages à l'ADN possédant un G-quadruplexe en aval ou en amont du signal de polyadénylation. Une

RESULTATS

fois les gènes sélectionnés, nous pourrons analyser leur efficacité de maturation en 3' lors de dommages à l'ADN par RT-PCRq.

II Une duplication de 20 nucléotides retrouvée dans des familles atteintes du syndrome de Lynch diminue l'efficacité de maturation en 3' du pré-ARNm de MSH6

II-1 Présentation et objectifs de l'étude

La protéine du système de réparation des mésappariements (MMR) MSH6 (« MutS Homolog 6 ») est une protéine nucléaire constituée de cinq domaines fonctionnels principaux : deux motifs PIP (« Proliferating cell nuclear antigen Interaction Motif ») situés à l'extrémité N-terminale et C-terminale permettant d'interagir avec la polymérase δ , un domaine PWWP et un domaine central comprenant 3 NLS et un domaine d'interaction avec l'ADN (Clark et al., 2007; Gassman et al., 2011)

La protéine MSH6 est indispensable au maintien de la stabilité génétique. En effet, elle permet d'initier le système de réparation des mésappariements qui corrige les bases non appariées et les petites insertions/délétions après la réplication de l'ADN. Bien qu'extrêmement fidèles, les polymérases introduisent toutefois de rares erreurs au sein des séquences qu'elles ont synthétisées. Ces erreurs sont reconnues par l'hétérodimère MSH2 MSH6 (Hsieh, 2001) déclenchant le recrutement des MLH1 (« MutL Homolog 1 ») et PSM2 (« Postmeiotic Segregation increased 2 ») pour former le complexe de réparation des mésappariements appelé réparosome (Jiricny, 1998). Le mécanisme permettant l'identification du brin neosynthétisé n'est pas clairement établi. Il existe probablement un couplage avec la réplication permettant l'identification du brin neosynthétisé au travers d'une forme de « mémoire » de la réplication (Jiricny, 1998). Le réparosome recrute l'exonucléase I qui, assistée par des hélicases, dégrade le brin neosynthétisé (Szankasi and Smith, 1995; Tran et al., 2001b). La polymérase δ peut alors repolymériser le brin d'ADN complémentaire supprimant ainsi le mésappariement. (Harfe and Jinks-Robertson, 2000)

L'inactivation du MMR due à des mutations dans l'un des quatre gènes du MMR (MLH1, MSH2, MSH6 et PSM2) provoque le syndrome de Lynch. Le syndrome de Lynch également appelé cancer colorectal héréditaire sans polypose est un syndrome héréditaire responsable d'un risque accru de cancers colorectaux, de l'endomètre, des ovaires, de l'estomac, de l'intestin grêle, du foie et de l'appareil urinaire supérieur. L'inactivation des gènes du MMR provoque des instabilités géniques notamment aux niveaux des séquences microsatellites. Ces séquences présentes sur l'ensemble du génome sont formées par la répétition de dix à cent fois, d'un motif composé de deux à dix nucléotides.

Des mutations du gène *msh6* sont à l'origine de 10 à 20% des syndromes de Lynch et 0,4% de l'ensemble des cancers colorectaux (Baglietto et al., 2010; Barnetson et al., 2006). Ces mutations sont principalement décrites dans la phase codante mais également au niveau des régions correspondant au 3'UTR du gène de MSH6 (InSiGHT database). L'effet de ces mutations au niveau du 3'UTR n'a pour l'instant pas été défini.

Le 3'UTR de l'ARNm du gène MSH6 est constitué de 85 bases, les mutations dans ces régions pourraient affecter la stabilité de l'ARNm, l'efficacité de traduction mais également la maturation en 3'.

Le service d'Oncogénétique de l'Institut Claudius Regaud a identifié une nouvelle mutation dans le 3'UTR de MSH6 chez des patients atteints du syndrome de Lynch. Cette mutation correspond à une duplication des 20 nucléotides précédant le signal de polyadénylation. Nous avons donc souhaité déterminer si cette mutation modifie l'expression de la protéine MSH6 par l'intermédiaire d'une régulation de la maturation en 3'.

II-2 Stratégie expérimentale

II-2-1 Identification de mutations dans la séquence correspondant au 3' UTR de MSH6

Afin d'identifier de nouvelles mutations dans les séquences correspondantes au 3' UTR du gène MSH6, le service d'Oncogénétique de l'Institut Claudius Regaud dirigé par le Dr Toulas et le Pr Favre a réalisé, par génotypage, le séquençage du derniers exon de MSH6 incluant la région correspondant au 3'UTR. Ce séquençage a été fait grâce à une banque d'ADN issue de 189 patients ayant le syndrome de Lynch, exempts de mutation dans la phase codante ou dans les régions introniques à proximité des exons pour les gènes MSH2, MLH1 et MSH6. Ceci a permis de mettre en évidence une duplication de 20 nucléotides précédant le signal de polyadénylation chez deux patients issus de familles distinctes.

II-2-2 Etude de l'implication de la duplication dans le 3'UTR sur l'expression de MSH6

Dans le but de déterminer l'implication de cette duplication sur l'expression protéique, nous avons cloné la séquence correspondant au 3' UTR de MSH6 d'une longueur de 85 nucléotides ainsi que les 300 nucléotides en aval. Nous avons cloné cette séquence en aval du gène de luciférase *firefly*. Nous avons également cloné la même séquence avec les 20 nucléotides précédant le signal de polyadénylation, répétée deux fois et quatre fois. Ces différentes constructions ont été transfectées de manière transitoire dans différentes lignées cellulaires tumorales. Un plasmide contrôle contenant la luciférase *renilla*, est également co-transfecté pour normaliser l'efficacité de transfection. La quantification des différentes activités luciférase nous a permis de montrer que les répétitions des 20 nucléotides précédant le signal de polyadénylation de MSH6 entraînent une diminution de l'expression des gènes rapporteurs.

II-2-3 Etude du mécanisme impliqué dans la régulation de l'expression de MSH6

Afin de déterminer l'étape dérégulée par les duplications dans la région correspondant au 3' UTR de MSH6, nous avons commencé par évaluer l'efficacité de maturation en 3' des différentes constructions, par RT-PCRq, en quantifiant le pré-ARNm non clivé normalisé par l'ARN total (ARNm et pré-ARNm) respective de chaque rapporteur. Ces expériences nous ont permis de mettre en évidence que ces répétitions des 20 nucléotides précédant le signal de polyadénylation entraînent une diminution de l'efficacité de maturation en 3'. Nous avons ensuite évalué l'efficacité de traduction des ARNm issus des différentes constructions en normalisant les activités luciférase par la quantité d'ARNm respective de chaque rapporteur. Cette expérience a permis d'exclure une régulation traductionnelle engendrée par les répétitions des 20 nucléotides.

II-3 Article II :

« Decreased efficiency of MSH6 mRNA polyadenylation linked to a 20-nucleotides duplication found in French Lynch syndrome families »

Adrien Decorsière, Françoise Fouque, Anne-Françoise Tilkin-Mariamé, Janick Selves, Rosine Guimbaud, Edith Chipoulet, Caroline Delmas, Gilles Favre, Stephan Vagner, Stefania Millevoi, Christine Toulas.



Decreased efficiency of MSH6 mRNA polyadenylation linked to a 20-nucleotides duplication found in French Lynch syndrome families

Journal:	<i>Human Mutation</i>
Manuscript ID:	Draft
Wiley - Manuscript type:	Brief Report
Date Submitted by the Author:	n/a
Complete List of Authors:	Decorsieres, Adrien; INSERM U1037 Fouque, Françoise; Institut Claudius Regaud, Oncogenetic Tilkin-Mariamé, Anne Françoise; INSERM U1037 Selves, Janick; CHU Guimbaud, Rosine; INSERMU1037, Institut Claudius Regaud Chipoulet, Edith; Institut Claudius Regaud, Oncogenetic Delmas, Caroline; INSERMU1037, Institut Claudius Regaud Favre, Gilles; INSERMU1037, Institut Claudius Regaud Vagner, Stephan; INSERM U981 Millevoi, Stefania; INSERM U1037 Toulas, Christine; INSERMU1037, Institut Claudius Regaud
Key Words:	MSH6, Lynch syndrome, 3'UTR, polyadenylation

SCHOLARONE™
Manuscripts

1
2
3
4 **Decreased efficiency of *MSH6* mRNA polyadenylation linked to a 20-**
5 **nucleotides duplication found in French Lynch syndrome families.**
6
7
8
9

10
11 Adrien Decorsière^{1,2,3}, Françoise Fouque², Anne-Françoise Tilkin-Mariamé^{1,2,3}, Janick
12 Selves^{1,4}, Rosine Guimbaud^{1,2,3,5}, Edith Chipoulet^{2,5}, Caroline Delmas^{1,2}, Gilles Favre^{1,2,3,5},
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Stephan Vagner^{1,2,3,#}, Stefania Millevoi^{1,2,3,*}, Christine Toulas^{1,2,3&}.

¹Institut National de la Santé et de la Recherche Médicale (INSERM) U1037, Toulouse, F-31000, France

²Institut Claudius Regaud, Toulouse, F-31052, France

³Université Toulouse III Paul Sabatier, Toulouse, F-31000, France

⁴Service d'anatomie et cytologie pathologiques, hôpital Purpan, place du Docteur-Baylac, Toulouse, France

⁵Digestive Oncology and Oncogenetic. Pôle digestif. CHU Toulouse, France

***Co-corresponding authors:**

[#]Present address: INSERM U981, Institut Gustave Roussy, 114 Rue Edouard Vaillant, 94800 Villejuif, France Tel: (33) 1-42-11-65-10, Fax: (33) 1-42-11-60-94, e-mail:

Stephan.vagner@inserm.fr

^{*} Present address: INSERM U1037, Bât. L3, B.P. 84225, 31432 Toulouse Cedex 4, France
Tel: (33) 5-61-32-21-46, Fax: (33) 5-61-32-21-41 e-mail: stefania.millevoi@inserm.fr

[&]Christine Toulas, Institut Claudius Regaud, INSERMU1037, Laboratoire d'Oncogénétique, 20-24 rue du Pont St-Pierre, 31052 Toulouse, France. Tel : (33) 5-61-42-42-75, Fax : (33) 5-61-42-46-31 e-mail: toulas.christine@claudiusregaud.fr

ABSTRACT

The aim of this work was to determine the prevalence of *MSH6* 3'UTR variants and their potential contribution to Lynch syndrome. We performed sequencing of the 3' UTR of *MSH6* in a population of 189 non mutated-*MMR* patients fulfilling Amsterdam criteria. We identified for two different families the duplication of the same 20 bp sequence located on c.*49-69 on the wild-type *MSH6* 3'UTR on two adjacent positions closed to the *MSH6* polyadenylation signal. We found that this duplication affects *MSH6* mRNA level expression *in vivo*. By using reporter plasmids carrying the *MSH6* 3' end region, we demonstrated that the duplication of this sequence decreases *MSH6* expression by a mechanism reducing polyadenylation efficiency. Our data strongly suggests that this duplication may be a causal mutation of Lynch syndrome and that analyzing the *MSH6* gene mutations would benefit from investigating the 3' UTR region of this gene.

INTRODUCTION

Lynch syndrome is a rare autosomal dominant inherited syndrome that confers increased risks of colorectal, endometrial and to a lower extent of stomach, small intestine, hepatobiliary, kidney and ovarian cancers. Germline mutations in one of the four mismatch repair (*MMR*) genes *MLH1*, *MSH2*, *MSH6* and *PMS2* have been associated with this syndrome. Inactivation of these *MMR* genes induces somatic instability in particular in microsatellite region (MSI). Mutations in *MSH6* account for 10%–20% of Lynch syndrome colorectal cancers and the risk to develop cancer for mutation carriers is far to be negligible. Indeed, among *MSH6* mutations carriers, four of ten men and two of ten women will be diagnosed for colorectal cancer before 80-years old, four women on ten will be diagnosed for endometrial cancer before this age (Baglietto et al., 2010).

Deleterious mutations are mainly located either within the coding sequence or at the exon-intron junctions of *MMR* genes, but large deletions and duplications encompassing exons occur. More recently *MLH1* and *MSH2* constitutive promotor hypermethylations have also been reported (Niessen et al., 2009). However, in many patients suspected to belong to high risk Lynch syndrome families presenting MSI, no mutation nor hypermethylation of *MMR* genes can be detected in the exon/intron sequences and promoter region, respectively, suggesting the presence of mutations in other regions controlling *MSH6* expression. This prompted us to look for novel mutations at the 3' side of the *MSH6* open reading frame region. Indeed, this region, often neglected during the genetic screening of disease-associated candidate genes, may play an important role in various diseases (Conne et al., 2000). 3'UTR regions are involved in the post-transcriptional control of gene expression either at the pre-mRNA processing level (splicing or polyadenylation) or at the mRNA level (stability,

1
2
3 translation). Any sequence variations in the 3'UTR may have strong consequences on gene
4
5 expression and on the etiology of a given disease.
6

7
8 The 3'UTR of *MMR* genes have been poorly explored so far. Mao et al investigated the
9
10 presence of variants in the 3'UTR of the *MLH1* mRNA and identified a three nucleotides-
11
12 deletion which reduces *MLH1* expression in white blood cells from leukaemia patient,
13
14 probably conferring leukaemia relapse (Mao et al., 2008) but the pathogenicity of this variant
15
16 is controversial in the Lynch syndrome (Tinat et al., 2008). Even if some variants have been
17
18 reported in the 3'UTR of the *MSH6* mRNA (see the InSiGHT database: [http://www.insight-
21
22 group.org/home/](http://www.insight-
19
20 group.org/home/)), no investigation has been conducted to describe the impact of these variants
23
24 on MSH6 expression and, as a consequence, their role in Lynch syndrome is still unknown.
25

26
27 The purpose of this study was to determine the prevalence of mRNA variants in the
28
29 *MSH6* 3'UTR in patients fulfilling Amsterdam criteria and their potential contribution to
30
31 MSH6 expression and Lynch syndrome.
32

33 34 35 36 MATERIAL AND METHODS

37
38 Detailed information on Material and Methods is available in Supplementary Data
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

RESULTS

We investigated the presence of variants in the 3'UTR of the *MSH6* gene for 189 patients fulfilling Amsterdam criteria or selected because of a suggestive familial history associated to a MSI phenotype in the tumour. Among these patients, the MSI status was negative for 63, low for 4 patients, positive for 39 patients and unknown for 83 patients. Our analysis did not reveal any single nucleotide substitution (SNP) in the *MSH6* 3' UTR of this population. However, we identified, for two different patients, the duplication of the same 20 bp sequence located on c.*49-69 on the wild-type *MSH6* 3' UTR on two adjacent positions, i.e c.*47-67InsTTCAGACAACATTATGATCT or c.*48-68 Dup in the vicinity of the 3' end formation signal (Supplementary Figure 1A). These variants have never been reported in France nor in the various international databases listing the identified *MMR* mutations. This duplication has not been identified in the control population composed of 51 patients from high risk breast/ovarian cancers families. *In silico* analysis reveals that this 20bp duplication 5'TTCAGACAACATTATGATCT-3' (Dup20bp) which is adjacent to the *MSH6* polyadenylation signal in human, is highly conserved among higher vertebrates (Supplementary Figure 1B). The two probands carrying these two variants have a strong history of cancers associated with Lynch syndrome (Fig 1A). The first patient (patient 1) has developed a sigmoid colon cancer at 57, her mother a rectal cancer at 38 and her father a colon cancer at 69. The second proband (patient 2) has also developed a sigmoid colon cancer at 38, her mother an ovarian cancer at 55 and a colon cancer at 62. Patient 1's tumor was MSI-H and the expression of *MSH6* was detected. The MSI status was not performable for patient 2's tumor due to tumor fixation conditions.

To investigate whether this duplicated sequence might modify *MSH6* expression, we generated an EBV-immortalized lymphoblastoid cell line from patient 1's lymphocytes (P1 cell line) and performed real time RT-qPCR to compare *MSH6* mRNA expression in P1

1
2
3 versus 3 different EBV-immortalized lymphoblastoid cell lines from patients not belonging to
4
5 Lynch syndrome families (Fig1B). The level of *MSH6* mRNA is quite variable in the different
6
7 control cell lines but the amount of *MSH6* mRNA is lower in P1 cell line compared to the
8
9 control patients as previously shown by others (Auclair et al., 2007).
10

11
12 To better decipher the role of this 20bp duplication on *MSH6* expression, we constructed a
13
14 reporter vector in which the *MSH6* 3' end region (including the 3'UTR and 3' flanking
15
16 region) was inserted downstream of the *Firefly* luciferase (FLuc) and two variants plasmids in
17
18 which the Dup20pb was duplicated (2ins) or quadruplicated (4ins) (Fig. 2A). Relative
19
20 luciferase activity showed that *MSH6* reporter expression decreased when the 20 bp region
21
22 was duplicated and further diminished when the 20 bp region was quadruplicated, compared
23
24 to the wild type construct, suggesting that *MSH6* expression decreases as the number of
25
26 duplication increases. This effect is not cell-specific since it was observed in the different
27
28 cancer cell lines including colon and ovarian cancers belonging to Lynch syndrome (Fig. 2B).
29
30 To determine whether the Dup20bp affects translational efficiency, we normalized luciferase
31
32 activity to luciferase RNA levels (Vasudevan and Steitz, 2007). We observed that the
33
34 normalized luciferase activities of the reporter constructs containing the two different
35
36 duplications (2ins or 4ins) were equivalent to the one measured following the transfection of
37
38 the wild type plasmids (Supplementary Figure2), showing that the duplications in the *MSH6*
39
40 3' UTR do not modify the ability of the *MSH6* reporter mRNA to be translated. The
41
42 proximity of the Dup20bp to the *MSH6* polyadenylation signal prompted us to investigate
43
44 whether the 20 bp duplication interferes with the process leading to *MSH6* mRNA
45
46 polyadenylation. Real-time RT-qPCR analysis of *MSH6* mRNA polyadenylation efficiency,
47
48 measured by the ratio of pre-mRNA to total RNA, revealed that the duplication of the 20bp
49
50 sequence reduces *MSH6* mRNA polyadenylation. This effect was potentiated with the
51
52 construct containing four repetitions (Fig. 2C). Overall, these results suggest that the 20bp
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 duplicated sequence reduces MSH6 expression by lowering the efficiency of *MSH6* mRNA
4
5
6 polyadenylation .
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

DISCUSSION

Our work investigated for the first time whether *MSH6* 3'UTR sequence modifications might exist in Lynch syndrome families and if so, whether these potential sequence variations might affect *MSH6* expression and in consequence, be causal of Lynch syndrome. Even if SNPs have been reported in *MSH6* 3'UTR as c.*85T>A (rs2020906) (Perez-Cabornero et al., 2009), we did not detect any SNP in *MSH6* 3'UTR of the 189 patients accrued in our study, revealing the low mutability of the 3' UTR of *MSH6* in these families. The only *MSH6* 3'UTR modification detected in our population concerns the duplication of the c.*49-69 sequence of wild-type *MSH6* 3'UTR located on two adjacent positions in two different patients. This duplication was not observed in the *MSH6* 3'UTR of 51 patients from high risk breast/ovarian families without any history of familial colorectal cancer. An insertion of 10bp in *MSH6* 3'UTR has previously been reported in a family presenting endometrial and ovarian cancer respectively MSI-low and MSI-high, without any loss of *MSH6* expression in the tumors (Berends et al., 2002). In our study, tumor of the Dup20bp bearing patient was MSI-high but no extinction of *MSH6* expression could have been detected. Unfortunately, we could not access to Patient 2's tumor to perform MSI test and examine *MSH6* expression within the tumor. Due to the small size of the two families presenting this Dup20pb, we could not perform any segregation analysis.

In vitro analysis showed that this Dup20bp affects the polyadenylation step of *MSH6* gene expression. A possible explanation of this result is that this variant alters the binding of polyadenylation regulatory proteins. This hypothesis is supported by a number of evidences showing that diseases-causing mutations in 3' UTR affects polyadenylation by interfering with the accessibility of regulatory proteins to their binding elements in the vicinity of the polyadenylation signal (Danckwardt et al., 2008) . The mechanism underlying the effect of the mutation on polyadenylation efficiency may involve a default in the direct interaction of

1
2
3 these regulatory proteins with factors of the polyadenylation machinery (as in (Millevoi et al.,
4 2002; Millevoi et al., 2006)). Alternatively, the mutation can interfere with the ability of the
5 regulatory protein to recruit additional regulatory factors that, in turn, affects the assembly of
6 the polyadenylation machinery (as in (Millevoi et al., 2009)). Disease-causing mutation may
7 also facilitate the binding of repressor proteins, as we recently showed for the alteration
8 within BRCA1 exon 18 (Millevoi and Vagner, 2010). In this case, the presence of multiple
9 binding sites would increase the repressor protein occupancy nearby the polyadenylation
10 signal, resulting in polyadenylation inhibition.
11
12
13
14
15
16
17
18
19
20

21 In conclusion, our work exploring for the first time the involvement of *MSH6* 3'UTR in
22 Lynch syndrome reveals the existence of variants containing a duplicated sequence which
23 regulates *MSH6* expression by lowering the efficiency of *MSH6* pre-mRNA polyadenylation.
24 This 20bp duplication-dependent *MSH6* expression regulation strongly suggests that this
25 duplication may be a causal mutation of Lynch syndrome. Segregation analyses in large
26 families bearing this type of duplication would reinforce this data. Noteworthy, our work
27 strongly suggests that analyzing the involvement of the *MSH6* gene mutations in Lynch
28 syndrome should not be restricted to the detection of the mutations located in exons or at
29 exon-intron junctions (completed by rearrangement analysis), but may benefit from
30 investigating the 3'UTR region of *MSH6* gene.
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

48 ACKNOWLEDGMENTS

49 This work was supported by INSERM and FRM (Equipe FRM, soutenue par la Fondation
50 Recherche Médicale; AD, SM, SV) and the Groupe de Recherche de l' Institut Claudius
51 Regaud (FF, CT, SM, SV, RG). A.D. was a recipient of a studentship from ARC (Association
52 pour la recherche sur le cancer) and FRM. The authors sincerely thank Chantal Darnau,
53 Pauline Dejean and Isabelle Gitlaw for their efficient technical assistance.
54
55
56
57
58
59
60

FIGURE LEGENDS**Figure 1 : Identification of the Dup20bp at *MSH6* 3'UTR in two Lynch syndrome families**

A: Pedigrees of patient 1's and patient 2's families bearing the Dup20bp sequence. Square symbols indicate males, round symbols females. Filled symbols indicate breast (BC), ovarian (OC), rectal (RC), colon (CC) and lung (LC) cancer affected individuals. Ages of diagnosis are indicated below the individuals.

B: Real-time RT-qPCR analysis of *MSH6* mRNA in lymphoblastoids immortalized cells from patient 1 (P1) or 3 control patients (control) without any familial history of colorectal cancer. Bars represented the mean \pm SD of the $\Delta\Delta C_t$ (the reference was control 1) of three different experiments.

Figure 2 : The Dup20bp repetition at the *MSH6* 3' UTR inhibits mRNA polyadenylation

A: Illustration of luciferase reporter vector containing *MSH6* 3' UTR and 3' end flanking regions inserted downstream from the *Firefly luciferase* gene (*Fluc*). Individual clones carrying the region of interest wild type (WT), duplicated (2ins) or quadruplicated (4ins) are depicted.

B: The ratio of Firefly/Renilla luciferase activities (Fluc/Rluc) was determined using different cells lines (A549, HCT116, OVCAR, SKOV or Hela) cotransfected with WT, 2ins or 4ins plasmids and an internal control plasmid encoding Renilla luciferase. Bars represented the mean \pm SD of the Fluc/Rluc ratio of three different experiments.

C: Reverse transcription of RNA from SKOV or OVCAR cells transfected with WT, 2ins or 4ins plasmid followed by real-time qPCR with primer specific to located on either side of the cleavage site to amplify the uncleaved RNA or upstream of the cleavage site to amplify both

1
2
3 the uncleaved and cleaved RNA, as depicted. Bars represented the mean \pm SD of the 3'
4
5 processing efficiency of three different experiments.
6
7
8
9

10 REFERENCES

- 11
12
13
14 1. Auclair J, Leroux D, Desseigne F, Lasset C, Saurin JC, Joly MO, Pinson S, Xu XL,
15 Montmain G, Ruano E, Navarro C, Puisieux A, Wang Q. 2007. Novel biallelic mutations in
16 MSH6 and PMS2 genes: gene conversion as a likely cause of PMS2 gene inactivation. *Hum*
17 *Mutat* 28:1084-1090.
18
19
20 2. Baglietto L, Lindor NM, Dowty JG, White DM, Wagner A, Gomez Garcia EB,
21 Vriends AH, Cartwright NR, Barnetson RA, Farrington SM, Tenesa A, Hampel H, Buchanan
22 D, Arnold S, Young J, Walsh MD, Jass J, Macrae F, Antill Y, Winship IM, Giles GG,
23 Goldblatt J, Parry S, Suthers G, Leggett B, Butz M, Aronson M, Poynter JN, Baron JA, Le
24 ML, Haile R, Gallinger S, Hopper JL, Potter J, de la Chapelle A, Vasen HF, Dunlop MG,
25 Thibodeau SN, Jenkins MA. 2010. Risks of Lynch syndrome cancers for MSH6 mutation
26 carriers. *J Natl Cancer Inst* 102:193-201.
27
28
29 3. Berends MJ, Wu Y, Sijmons RH, Mensink RG, van der Sluis T, Hordijk-Hos JM, de
30 Vries EG, Hollema H, Karrenbeld A, Buys CH, van der Zee AG, Hofstra RM, Kleibeuker JH.
31 2002. Molecular and clinical characteristics of MSH6 variants: an analysis of 25 index
32 carriers of a germline variant. *Am J Hum Genet* 70:26-37.
33
34
35 4. Conne B, Stutz A, Vassalli JD. 2000. The 3' untranslated region of messenger RNA: A
36 molecular 'hotspot' for pathology? *Nat Med* 6:637-641.
37
38
39 5. Danckwardt S, Hentze MW, Kulozik AE. 2008. 3' end mRNA processing: molecular
40 mechanisms and implications for health and disease. *EMBO J* 27:482-498.
41
42
43 6. Mao G, Pan X, Gu L. 2008. Evidence that a mutation in the MLH1 3'-untranslated
44 region confers a mutator phenotype and mismatch repair deficiency in patients with relapsed
45 leukemia. *J Biol Chem* 283:3211-3216.
46
47
48 7. Millevoi S, Decorsiere A, Loulergue C, Iacovoni J, Bernat S, Antoniou M, Vagner S.
49 2009. A physical and functional link between splicing factors promotes pre-mRNA 3' end
50 processing. *Nucleic Acids Res* 37:4672-4683.
51
52
53 8. Millevoi S, Geraghty F, Idowu B, Tam JL, Antoniou M, Vagner S. 2002. A novel
54 function for the U2AF 65 splicing factor in promoting pre-mRNA 3'-end processing. *EMBO*
55 *Rep* 3:869-874.
56
57
58 9. Millevoi S, Loulergue C, Dettwiler S, Karaa SZ, Keller W, Antoniou M, Vagner S.
59 2006. An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing
60 machineries. *EMBO J* 25:4854-4864.
10. Millevoi S, Vagner S. 2010. Molecular mechanisms of eukaryotic pre-mRNA 3' end
processing regulation. *Nucleic Acids Res* 38:2757-2774.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

11. Niessen RC, Hofstra RM, Westers H, Ligtenberg MJ, Kooi K, Jager PO, de Groot ML, Dijkhuizen T, Olderode-Berends MJ, Hollema H, Kleibeuker JH, Sijmons RH. 2009. Germline hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome. *Genes Chromosomes Cancer* 48:737-744.

12. Perez-Cabornero L, Velasco E, Infante M, Sanz D, Lastra E, Hernandez L, Miner C, Duran M. 2009. A new strategy to screen MMR genes in Lynch Syndrome: HA-CAE, MLPA and RT-PCR. *Eur J Cancer* 45:1485-1493.

13. Tinat J, Baert-Desurmont S, Latouche JB, Vasseur S, Martin C, Bouvignies E, Frebourg T. 2008. The three nucleotide deletion within the 3'untranslated region of MLH1 resulting in gene expression reduction is not a causal alteration in Lynch syndrome. *Fam Cancer* 7:339-340.

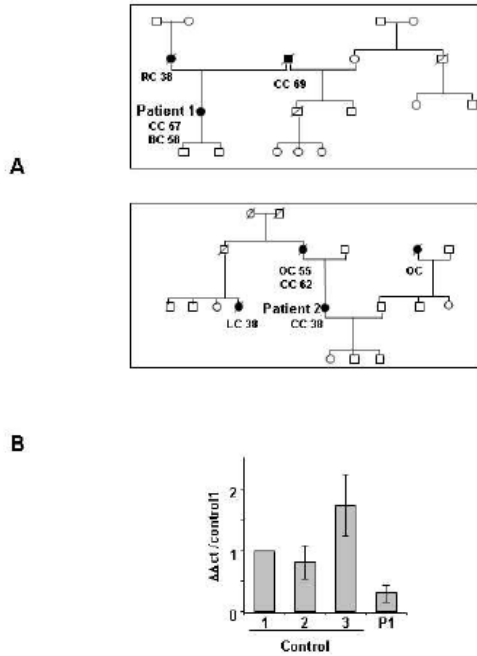
14. Vasudevan S, Steitz JA. 2007. AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. *Cell* 128:1105-1118.

For Peer Review

Human Mutation

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1

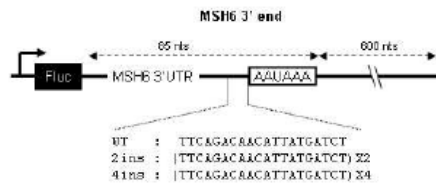


190x275mm (96 x 96 DPI)

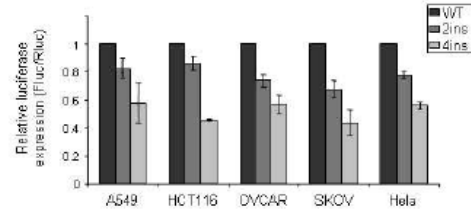
Human Mutation

Figure 2

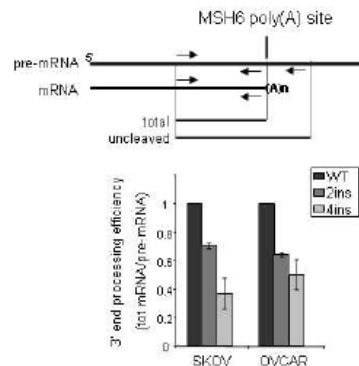
A



B



C



190x275mm (96 x 96 DPI)

SUPPLEMENTARY DATA**MATERIAL AND METHODS****Patient population**

Patients enrolled in this study were ascertained at the Institute Claudius Regaud and the Hospital of Toulouse, France between 2005 and 2010. All these 189 patients fulfilled Amsterdam criteria or selected because of a suggestive familial history associated to a MSI phenotype in the tumour. Blood samples were collected following informed consent. The control group consisted of 51 patients belonging to high risk breast/ovarian cancer families without any familial history of colorectal cancer. All the patients accrued in this study did not carry any mutation on *MSH2*, *MLH1* or *MSH6* in the exons or the intron-exon junctions nor *MLH1* or *MSH2* rearrangement.

MSI detection and IHC staining

Tumors were analyzed for MSI and for *MLH1*, *MSH2* and *MSH6* expression by immunohistochemistry as previously described (Danjoux et al., 2006). Tumors were scored as high level of MSI (MSI-H), low level of MSI (MSI-L) or microsatellite stable (MSS) as usually recommended (Umar et al., 2004; Vasen et al., 1991). MSI-H was considered as MSI-positive; MSI-L and MSS as MSI-negative.

Genotyping

The 3' UTR (or 3' flank) of *MSH6* was amplified by PCR using M13-flagged primers : forward primer 5'-GGAAGGGATGATGCACTATG-3'; reverse primer 5'-AAGAAAATGGAAAATGGTCA-3') including exon 10. Bi-directional sequencing was performed on a automated Sequencer ABI Prism 3130XL (Applied Biosystems).

1
2
3 The DNA numbering is based on the Human Genome Variation Society's
4
5 recommendations on the cDNA sequence U54777.
6
7
8
9

10 **Immortalization of lymphoblastic cell lines**

11
12 Lymphoblastoid cell lines were immortalized by Epstein Barr virus as previously described
13 (Tilkin et al., 1986) from Patient 1's lymphocytes and from patients without any familial
14 history of colorectal cancer.
15
16
17
18
19

20 **Luciferase reporter and transfection**

21
22 The *MSH6* 3' flank region was generated by PCR on genomic DNA of MCF7 cell using the
23 primers F-WT-MSH6 (5'-tttactagtagtactgactacattggaagctt-3') and R-WT-MSH6 (5'-
24 aaaggatccagctctgtattgaaactctg-3'), followed by enzymatic digestion with SpeI and Bam HI.
25
26 The amplified fragments were cloned in pGL3-TK vector using XbaI and BamHI sites to
27 generate the Fluc/MSH6 plasmid. The mutant 2ins was generated by a splicing PCR from the
28 Fluc/MSH6 using the primers F-WT-MSH6 with R-2ins (5'-
29 ttttaaaaaataaagtttattagatcataatggtgtctgaaagatcataatggtgtctgaattaccaccttgcagaag-3') and F-2ins
30 (5'-gatctaataaactttatttttaaaaatgacc-3') with R-WT-MSH6, and the mutant 4ins were generated
31 with the same technique from the Fluc/2ins using the primers F-WT-MSH6 with R-4ins (5'-
32 aaagtttattagatcataatggtgtctgaaagatcataatggtgtctgaaagatcataatggtgtctgaaagatcataat-3') and F-4ins
33 (5'-gacaacattatgatctaataaactttatttttaaaaatga-3') with R-WT-MSH6. Plasmid DNA (1 µg) was
34 transfected for 24 h in 12 well plates with the JetPEI transfection reagent (Polyplus
35 Transfection). The reporters were transfected in different cancer cell lines and FLuc activity
36 was normalized by the RLuc activity obtained after cotransfection of a *Renilla* luciferase
37 reporter construct (referred as relative activity).
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

***MSH6* mRNA quantification by real-time RT-qPCR**

1
2
3 RNA was extracted from the Patient 1's immortalized lymphoblastoid cells using the RNeasy
4 purification kit (Qiagen) and reverse transcription carried out using the iScript cDNA
5 synthesis kit (Bio-Rad), following the manufacturers' instructions. Quantitative reverse
6 transcription-polymerase chain reaction (qRT-PCR) was performed using Glyceraldehyde 3-
7 phosphate dehydrogenase (GAPDH) as house-keeping gene on the thermocycler iCycler (Bio-
8 Rad) with the SYBR Green master mix (Bio-Rad). *MSH6* mRNA expression was evaluated
9 according to the $\Delta\Delta C_t$ Livak method (Livak and Schmittgen, 2001)
10
11
12
13
14
15
16
17
18
19

20 21 22 **Pre-mRNA 3'-end processing efficiency analysis**

23
24 3'-end processing efficiency was analyzed as previously described (Decorsiere et al., 2011).
25
26
27

28 29 **REFERENCES :**

- 30
31
32
33 1. Danjoux M, Guimbaud R, Al ST, Meggetto F, Carrere N, Portier G, Delsol G, Selves J.
34 2006. Contribution of microdissection for the detection of microsatellite instability in
35 colorectal cancer. *Hum Pathol* 37:361-368.
36
37 2. Decorsiere A, Cayrel A, Vagner S, Millevoi S. 2011. Essential role for the interaction
38 between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing
39 and function during DNA damage. *Genes Dev* 25:220-225.
40
41 3. Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-
42 time quantitative PCR and the $2^{-(\Delta\Delta C(T))}$ Method. *Methods* 25:402-408.
43
44 4. Tilkin AF, Bagot M, Picard F, Dreyfus F, Vernant JP, Levy JP. 1986. B cells from
45 leukemic patients are relatively resistant to in-vitro EBV transformation. *Leuk Res* 10:1059-
46 1060.
47
48 5. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, Fishel R,
49 Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, Jass J, Lindblom A, Lynch HT,
50 Peltomaki P, Ramsey SD, Rodriguez-Bigas MA, Vasen HF, Hawk ET, Barrett JC, Freedman
51 AN, Srivastava S. 2004. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal
52 cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* 96:261-268.
53
54 6. Vasen HF, Mecklin JP, Khan PM, Lynch HT. 1991. The International Collaborative
55 Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis Colon Rectum*
56 34:424-425.
57
58
59
60

Supplementary figures legends

Figure S1 : Detection of the two duplications in the 3'UTR sequence of MSH6

A : Reverse sequencing analyses of the *MSH6* PCR product for a patient carrying a wild-type sequence (Control) and the two patients carrying the same duplication of the 20bp sequence (5'TTCAGACAACATTATGATCT-3') . Red arrows indicate the respective positions of these insertions (c.*47 for patient 1 (*47) and c.*48 for patient 2 (*48)).

B: Sequence alignment of vertebrates 3'UTR sequences of *MSH6* mRNA. The duplicated sequence detected in our two patients (5'TTCAGACAACATTATGATCT-3') is surrounded, mentioned in bold when homology is maintained between human and other vertebrates.

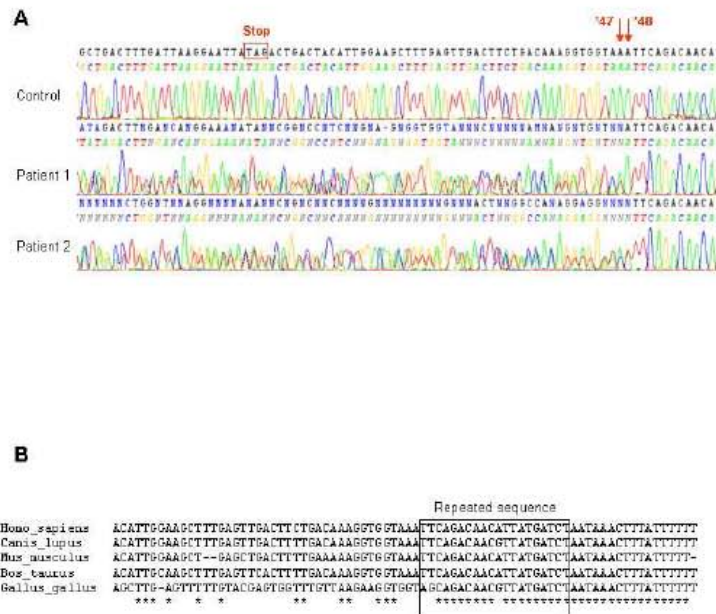
Figure S2 : The Dup 20bp sequence does not affect translation efficiency

Translational efficiency of the WT, 2ins and 4ins *MSH6* reporter vectors transfected in the ovarian cancer cell lines, SKOV and OVCAR, was measured by normalization of the relative luciferase activity (Fluc/Rluc) to luciferase RNA levels quantified by real time RT-qPCR analysis.

Human Mutation

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

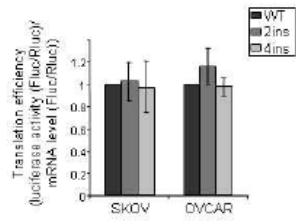
Figure S1



190x275mm (96 x 96 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure S2



190x275mm (96 x 96 DPI)

II-4 Discussions perspectives

Cette étude nous a permis d'identifier une nouvelle mutation localisée dans la séquence correspondant au 3'UTR du gène MSH6. Cette mutation a été identifiée chez des patients atteints du syndrome de Lynch. Nous avons montré que cette mutation affecte l'expression de gènes rapporteurs dans différentes lignées cellulaires. De plus, aucune autre mutation dans les gènes du MMR n'a été détectée chez ces patients. Ces résultats suggèrent que cette mutation pourrait être à l'origine du syndrome de Lynch.

Nous avons également révélé que la diminution de l'expression de gènes rapporteurs ayant cette mutation est due à une diminution de l'efficacité de maturation en 3' des pré-ARNm issus de ces constructions. Nous avons donc montré pour la première fois qu'une dérégulation de la maturation en 3' du pré-ARNm de MSH6 est associée au syndrome de Lynch.

Nous ne pouvons pas exclure avec cette étude qu'il y ait un effet additionnel de la duplication sur le recrutement de microRNA. En effet, la duplication dans le 3'UTR de l'ARNm de MSH6 pourrait créer un site d'appariement des microRNA ou dupliquer un site déjà existant. Cela aurait également pour effet une diminution de l'expression de la protéine MSH6. Pour vérifier que la duplication n'augmente pas le recrutement de microRNA, nous pourrions immunoprécipiter la protéine Ago 2 (Argonaute 2), le composant effecteur du complexe miRISC (« RNA Induced Silencing Complex ») (Glorian et al., 2011). En comparant les quantités d'ARNm des différents rapporteur immunoprécipitées à partir de cellules transfectées avec nos constructions contenant les 20 nucléotides précédant le signal de polyadénylation, répétée deux fois et quatre fois, nous pourrions déterminer si la duplication augmente le recrutement de microRNA.

De plus, le mécanisme de l'inhibition de la maturation en 3' de MSH6 n'est pas établi et ne semble pas, au vu de la séquence, modifier un élément de séquence de maturation en 3'. Néanmoins, il est probable que cette duplication affecte le recrutement de facteurs régulateurs de la maturation en 3'. La duplication pourrait entraîner une diminution du recrutement des facteurs de maturation en 3' canoniques ou de facteurs de stimulation de la maturation en 3', mais également augmenter le recrutement de facteurs inhibant la maturation en 3'. Afin de vérifier ces hypothèses, nous pourrions envisager de réaliser des chromatographies d'affinité à l'aide d'ARN biotinylés contenant les séquences du 3'UTR de MSH6 et la séquence en aval avec ou

RESULTATS

sans la duplication. Les protéines ainsi retenues pourraient alors être analysées par spectrométrie de masse et la comparaison entre les ARN contenant ou non la duplication renseignerait sur les facteurs impliqués dans la différence d'efficacité de maturation en 3'. Le décryptage de ce mécanisme pourrait permettre de mettre au point un système permettant de contrecarrer l'inhibition de la maturation en 3' et donc le syndrome de Lynch associé.

DISCUSSION ET CONCLUSIONS

La maturation en 3' des pré-ARNm : Importance dans le contrôle de l'expression des gènes et implications dans les pathologies

Mes travaux de thèse ont permis de mettre en évidence que la maturation en 3' de p53 peut réguler son expression protéique et son activité biologique. Le gène codant le suppresseur de tumeur p53 n'est pas affecté par l'inhibition globale de la maturation en 3' lors de dommages à l'ADN. Le mécanisme de maintien de la maturation en 3' implique le recrutement des protéines hnRNP H et F sur une séquence de type G-quadruplexe située en aval du signal de polyadénylation de p53. Des mutations de cette séquence, empêchant la structuration en G-quadruplexe sur le pré-ARNm de p53, entraînent une perte d'interaction avec les protéines hnRNP H et F, diminuant l'efficacité de maturation en 3' du pré-ARNm de p53 et donc l'accumulation de la protéine lors de dommages à l'ADN.

Nous avons également identifié une nouvelle mutation dans la séquence correspondant au 3'UTR du gène MSH6 impliqué dans le système de réparation des mésappariements. Cette mutation, retrouvée chez des patients atteints du syndrome de Lynch, diminue l'efficacité de maturation en 3' du pré-ARNm de MSH6 ce qui affecte l'expression de la protéine et pourrait de ce fait être à l'origine de cette pathologie.

Ces travaux ont donc contribué à montrer que les séquences de part et d'autre du site de polyadénylation sont essentielles pour la maturation en 3'. Leurs mutations peuvent conduire à une diminution de la maturation en 3' et donc de l'expression protéique. Plusieurs études ont montré un lien entre des polymorphismes dans ces séquences à proximité du site de maturation en 3' de nombreux gènes et certaines pathologies comme résumé dans le tableau suivant.

Tableau 4: Exemples de pathologies associées au polymorphisme dans les séquences entourant le site de polyadénylation susceptibles d'inhiber la maturation en 3'.

Gènes	Pathologie associée	Références
GIMAP5 (« GTPase, IMAP family member 5 »)	Diabètes de type I	(Shin et al., 2007)
FABP3 (« fatty acid-binding protein 3 »)	Résistances à l'insuline et des diabètes de type II	(Shin et al., 2003)
SLC7A1 (« solute carrier family 7 member A 1 »)	Hypertension	(Yang and Kaye, 2009)
HBA2 (« hemoglobin, alpha 2 »)	α -thalassémie	(Harteveld et al., 1994; Higgs et al., 1983)
HBB (« hemoglobin beta »)	β -thalassémie	(Jankovic et al., 1990; Millevoi et al., 2009)
DUX4 (« double homeobox 4 »)	Myopathie facio-scapulo-humérale	(Lemmers et al., 2010)
LOX1 (« lectin-like oxidized low-density lipoprotein receptor 1 »)	Cardiopathies ischémiques	(Chen et al., 2003)
SNCA (« synuclein, alpha »)	Maladie Alzheimer	(Lambert et al., 2003)
IL-1 α (« Interleukin-1 alpha »)	Maladie Alzheimer	(Sotiriou et al., 2009)
ARMS2 (« age-related maculopathy susceptibility 2 »)	Réduction du risque d'insuffisance rénale chronique terminale	(Bensen et al., 2003)
SERT (« Serotonin transporter »)	Dégénérescences maculaires liées à l'âge	(Fritsche et al., 2008)
IGF1 (« insulin-like growth factor 1 »)	Troubles paniques	(Gyawali et al., 2010)
SCN1B (« sodium channel, voltage-gated, type I, beta »)	Altération de la croissance	(Coutinho et al., 2007)
LDLR (« low-density lipoprotein receptor »)	Epilepsie généralisée avec convulsions fébriles	(Fendri-Kriaa et al., 2011)
CD19 (« cluster of differentiation 19 »)	Modification de la réponse suite à la contamination par le virus de l'hépatite C	(Hennig et al., 2002)
	Maladies auto-immunes comme la polyarthrite rhumatoïde, le lupus érythémateux disséminé et à la maladie de Crohn	(Kuroki et al., 2002)

DISCUSSION ET CONCLUSIONS

IL-12 (« interleukin-12 »)	Maladie de Crohn	(Seegers et al., 2002)
FG γ -H2 (« Fibrinogen gamma haplotype 2 »)	Thrombophilie	(Uitte de Willige et al., 2005)
F2 (« coagulation factor II »)	Thrombophilie	(Danckwardt et al., 2004)
FOXP3 (« forkhead box P3 »)	Syndrome IPEX (immunodysregulation polyendocrinopathy enteropathy X-linked syndrome)	(Bennett et al., 2001)
NAT1 (« N-acetyltransferase 1 »)	Maladie de Fabry	(Yasuda et al., 2003)
PTGS2/COX2 (« prostaglandin synthase 2/cyclooxygenase 2 »)	Cancer colorectal	(Bell et al., 1995)
TLR4 (« Toll-like receptor 4 »)	Cancer colorectal	(Cox et al., 2004)
TYMS (« thymidylate synthase »)	Cancer de la prostate	(Zheng et al., 2004)
	Carcinomes	(Zhang et al., 2004)

Gènes	Pathologie associée	Références
GIMAP5 (« GTPase, IMAP family member 5 »)	Diabètes de type I	(Shin et al., 2007)
FABP3 (« fatty acid-binding protein 3 »)	Résistances à l'insuline et des diabètes de type II	(Shin et al., 2003)
SLC7A1 (« solute carrier family 7 member A 1 »)	Hypertension	(Yang and Kaye, 2009)
HBA2 (« hemoglobin, alpha 2 »)	α -thalassémie	(Harteveld et al., 1994; Higgs et al., 1983)
HBB (« hemoglobin beta »)	β -thalassémie	(Jankovic et al., 1990; Millevoi et al., 2009)
DUX4 (« double homeobox 4 »)	Myopathie facio-scapulo-humérale	(Lemmers et al., 2010)
LOX1 (« lectin-like oxidized low-density lipoprotein receptor 1 »)	Cardiopathies ischémiques	(Chen et al., 2003)
SNCA (« synuclein, alpha »)	Maladie Alzheimer	(Lambert et al., 2003)
	Maladie Alzheimer	(Sotiriou et al., 2009)

DISCUSSION ET CONCLUSIONS

IL-1 α (« Interleukin-1 alpha »)	Réduction du risque d'insuffisance rénale chronique terminale	(Bensen et al., 2003)
ARMS2 (« age-related maculopathy susceptibility 2 »)	Dégénérescences maculaires liées à l'âge	(Fritsche et al., 2008)
SERT (« Serotonin transporter »)	Troubles paniques	(Gyawali et al., 2010)
IGF1 (« insulin-like growth factor 1 »)	Altération de la croissance	(Coutinho et al., 2007)
SCN1B (« sodium channel, voltage-gated, type I, beta »)	Epilepsie généralisée avec convulsions fébriles	(Fendri-Kriaa et al., 2011)
LDLR (« low-density lipoprotein receptor »)	Modification de la réponse suite à la contamination par le virus de l'hépatite C	(Hennig et al., 2002)
CD19 (« cluster of differentiation 19 »)	Maladies auto-immunes comme la polyarthrite rhumatoïde, le lupus érythémateux disséminé et à la maladie de Crohn	(Kuroki et al., 2002)
IL-12 (« interleukin-12 »)	Maladie de Crohn	(Seegers et al., 2002)
FG γ -H2 (« Fibrinogen gamma haplotype 2 »)	Thrombophilie	(Uitte de Willige et al., 2005)
F2 (« coagulation factor II »)	Thrombophilie	(Danckwardt et al., 2004)
FOXP3 (« forkhead box P3 »)	Syndrome IPEX (immunodysregulation polyendocrinopathy enteropathy X-linked syndrome)	(Bennett et al., 2001)
NAT1 (« N-acetyltransferase 1 »)	Maladie de Fabry Cancer colorectal	(Yasuda et al., 2003) (Bell et al., 1995)
CCND1 (Cycline D1)	Lymphomes non hodgkiniens	(Wiestner et al., 2007)
PTGS2/COX2 (« prostaglandin synthase 2/cyclooxygenase 2 »)	Cancer colorectal	(Cox et al., 2004)
TLR4 (« Toll-like receptor 4 »)	Cancer de la prostate	(Zheng et al., 2004)
TYMS (« thymidylate synthase »)	Carcinomes	(Zhang et al., 2004)

Ces différentes études ainsi que nos travaux prouvent l'importance des séquences à proximité du site de clivage dans le diagnostic de nombreuses pathologies héréditaires.

En effet, nous avons mis en évidence que la duplication de 20 nucléotides précédant le signal de polyadénylation de MSH6 est associée au syndrome de Lynch chez deux patients issus de familles distinctes.

Nous avons également montré que les séquences localisées en aval du signal de polyadénylation du pré-ARNm p53 peuvent être essentielles pour la maturation en 3' de p53 ainsi que pour son accumulation et ses fonctions de suppresseur de tumeur lors de dommages à l'ADN. Nous n'avons pas recherché des mutations dans cette séquence, néanmoins, en regardant dans la base de données « dbSNP », on peut constater qu'il existe plusieurs polymorphismes identifiés dans ces régions. Le gène codant la p53 est le plus fréquemment muté dans les cancers (Soussi et al., 2006). De plus, des mutations de ce gène sont responsables du syndrome de Li-Fraumeni prédisposant au cancer des tissus mous, des seins, des os, de la peau, du côlon, du pancréas, des surrénales, du cerveau et du sang (Palmero et al., 2010). Néanmoins, certains patients atteints du syndrome de Li-Fraumeni n'ont pas de mutation dans la séquence codante de p53. Il pourrait être intéressant de rechercher chez ces patients des polymorphismes au niveau des séquences localisées en aval du signal de polyadénylation et d'analyser l'effet de ces polymorphismes sur la maturation en 3' ainsi que sur l'expression de P53 lors de dommages à l'ADN. De même que le gène F2 dans lequel des mutations de son DSE sont à l'origine d'une thrombose (Danckwardt et al., 2004; Danckwardt et al., 2006a), des mutations ou des polymorphismes dans le DSE aux de p53 pourraient être à l'origine de cancers ou du syndrome de Li-Fraumeni.

En conclusion, mes travaux de thèse ainsi que les différentes études citées dans le tableau 4 ont permis de mettre en évidence que les séquences encadrant le site de polyadénylation sont essentielles à l'expression de nombreux gènes mis en cause dans différentes pathologies. Ces séquences pourraient faire l'objet d'analyse afin d'améliorer la prévention ou d'établir un diagnostic précoce de ces pathologies.

REFERENCE

REFERENCES

- Abba, M.C., Hu, Y., Sun, H., Drake, J.A., Gaddis, S., Baggerly, K., Sahin, A., and Aldaz, C.M. (2005). Gene expression signature of estrogen receptor alpha status in breast cancer. *BMC Genomics* 6, 37.
- Ahn, S.H., Kim, M., and Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol Cell* 13, 67-76.
- Alkan, S.A., Martincic, K., and Milcarek, C. (2006). The hnRNPs F and H2 bind to similar sequences to influence gene expression. *Biochem J* 393, 361-371.
- Arhin, G.K., Boots, M., Bagga, P.S., Milcarek, C., and Wilusz, J. (2002). Downstream sequence elements with different affinities for the hnRNP H/H' protein influence the processing efficiency of mammalian polyadenylation signals. *Nucleic Acids Res* 30, 1842-1850.
- Arora, A., and Maiti, S. (2009). Differential biophysical behavior of human telomeric RNA and DNA quadruplex. *J Phys Chem B* 113, 10515-10520.
- Ashe, M.P., Pearson, L.H., and Proudfoot, N.J. (1997). The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. *EMBO J* 16, 5752-5763.
- Awasthi, S., and Alwine, J.C. (2003). Association of polyadenylation cleavage factor I with U1 snRNP. *RNA* 9, 1400-1409.
- Azzouz, T.N., Gruber, A., and Schumperli, 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.
- Bagga, P.S., Arhin, G.K., and Wilusz, J. (1998). DSEF-1 is a member of the hnRNP H family of RNA-binding proteins and stimulates pre-mRNA cleavage and polyadenylation in vitro. *Nucleic Acids Res* 26, 5343-5350.
- Bagga, P.S., Ford, L.P., Chen, F., and Wilusz, J. (1995). The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3' end pre-mRNA processing through a trans-acting factor. *Nucleic Acids Res* 23, 1625-1631.
- Baglietto, L., Lindor, N.M., Dowty, J.G., White, D.M., Wagner, A., Gomez Garcia, E.B., Vriends, A.H., Cartwright, N.R., Barnetson, R.A., Farrington, S.M., *et al.* (2010). Risks of Lynch syndrome cancers for MSH6 mutation carriers. *J Natl Cancer Inst* 102, 193-201.
- Bai, Y., Auperin, T.C., Chou, C.Y., Chang, G.G., Manley, J.L., and Tong, L. (2007). Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors. *Mol Cell* 25, 863-875.
- Balbo, P.B., and Bohm, A. (2007). Mechanism of poly(A) polymerase: structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis. *Structure* 15, 1117-1131.
- Balim, Z., Kosova, B., Falzon, K., Bezzina Wettinger, S., and Colak, Y. (2003). Budd-Chiari syndrome in a patient heterozygous for the point mutation C20221T of the prothrombin gene. *J Thromb Haemost* 1, 852-853.
- Barabino, S.M., Hubner, W., Jenny, A., Minvielle-Sebastia, L., and Keller, W. (1997). The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes Dev* 11, 1703-1716.

REFERENCES

- Barnetson, R.A., Tenesa, A., Farrington, S.M., Nicholl, I.D., Cetnarskyj, R., Porteous, M.E., Campbell, H., and Dunlop, M.G. (2006). Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. *N Engl J Med* *354*, 2751-2763.
- Beaudoin, J.D., and Perreault, J.P. 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res* *38*, 7022-7036.
- Beaudoin, J.D., and Perreault, J.P. (2010). 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res* *38*, 7022-7036.
- Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M., and Gautheret, D. (2000). Patterns of variant polyadenylation signal usage in human genes. *Genome Res* *10*, 1001-1010.
- Bechara, E.G., Didiot, M.C., Melko, M., Davidovic, L., Bensaid, M., Martin, P., Castets, M., Pognonec, P., Khandjian, E.W., Moine, H., *et al.* (2009). A novel function for fragile X mental retardation protein in translational activation. *PLoS Biol* *7*, e16.
- Bell, D.A., Stephens, E.A., Castranio, T., Umbach, D.M., Watson, M., Deakin, M., Elder, J., Hendrickse, C., Duncan, H., and Strange, R.C. (1995). Polyadenylation polymorphism in the acetyltransferase 1 gene (NAT1) increases risk of colorectal cancer. *Cancer Res* *55*, 3537-3542.
- Bennett, C.L., Brunkow, M.E., Ramsdell, F., O'Briant, K.C., Zhu, Q., Fuleihan, R.L., Shigeoka, A.O., Ochs, H.D., and Chance, P.F. (2001). A rare polyadenylation signal mutation of the FOXP3 gene (AAUAAA-->AAUGAA) leads to the IPEX syndrome. *Immunogenetics* *53*, 435-439.
- Bensen, J.T., Langefeld, C.D., Li, L., McCall, C.E., Cousart, S.L., Dryman, B.N., Freedman, B.I., and Bowden, D.W. (2003). Association of an IL-1A 3'UTR polymorphism with end-stage renal disease and IL-1 alpha expression. *Kidney Int* *63*, 1211-1219.
- Bentley, D. (1999). Coupling RNA polymerase II transcription with pre-mRNA processing. *Curr Opin Cell Biol* *11*, 347-351.
- Bienroth, S., Keller, W., and Wahle, E. (1993). Assembly of a processive messenger RNA polyadenylation complex. *Embo J* *12*, 585-594.
- Birchmeier, C., Schumperli, D., Sconzo, G., and Birnstiel, M.L. (1984). 3' editing of mRNAs: sequence requirements and involvement of a 60-nucleotide RNA in maturation of histone mRNA precursors. *Proc Natl Acad Sci U S A* *81*, 1057-1061.
- Birnstiel, M.L., Busslinger, M., and Strub, K. (1985). Transcription termination and 3' processing: the end is in site! *Cell* *41*, 349-359.
- Blechingberg, J., Lykke-Andersen, S., Jensen, T.H., Jorgensen, A.L., and Nielsen, A.L. (2007). Regulatory mechanisms for 3'-end alternative splicing and polyadenylation of the Glial Fibrillary Acidic Protein, GFAP, transcript. *Nucleic Acids Res* *35*, 7636-7650.
- Blencowe, B.J., Issner, R., Nickerson, J.A., and Sharp, P.A. (1998). A coactivator of pre-mRNA splicing. *Genes Dev* *12*, 996-1009.
- Boelens, W.C., Jansen, E.J., van Venrooij, W.J., Stripecke, R., Mattaj, I.W., and Gunderson, S.I. (1993). The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell* *72*, 881-892.
- Bonnal, S., Schaeffer, C., Creancier, L., Clamens, S., Moine, H., Prats, A.C., and Vagner, S. (2003). A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *J Biol Chem* *278*, 39330-39336.

REFERENCES

- Brackenridge, S., and Proudfoot, N.J. (2000). Recruitment of a basal polyadenylation factor by the upstream sequence element of the human lamin B2 polyadenylation signal. *Mol Cell Biol* 20, 2660-2669.
- Bramham, C.R., and Wells, D.G. (2007). Dendritic mRNA: transport, translation and function. *Nat Rev Neurosci* 8, 776-789.
- Brodsky, A.S., and Silver, P.A. (2000). Pre-mRNA processing factors are required for nuclear export. *RNA* 6, 1737-1749.
- Brooks, T.A., Kendrick, S., and Hurley, L. (2010). Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J* 277, 3459-3469.
- Brown, K.M., and Gilmartin, G.M. (2003). A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. *Mol Cell* 12, 1467-1476.
- Bruno, T., De Nicola, F., Iezzi, S., Lecis, D., D'Angelo, C., Di Padova, M., Corbi, N., Dimiziani, L., Zannini, L., Jekimovs, C., *et al.* (2006). Che-1 phosphorylation by ATM/ATR and Chk2 kinases activates p53 transcription and the G2/M checkpoint. *Cancer Cell* 10, 473-486.
- Bruno, T., Desantis, A., Bossi, G., Di Agostino, S., Sorino, C., De Nicola, F., Iezzi, S., Franchitto, A., Benassi, B., Galanti, S., *et al.* (2010). Che-1 promotes tumor cell survival by sustaining mutant p53 transcription and inhibiting DNA damage response activation. *Cancer Cell* 18, 122-134.
- Buratowski, S. (2005). Connections between mRNA 3' end processing and transcription termination. *Curr Opin Cell Biol* 17, 257-261.
- Burns, D.M., and Richter, J.D. (2008). CPEB regulation of human cellular senescence, energy metabolism, and p53 mRNA translation. *Genes Dev* 22, 3449-3460.
- Burns, T.F., and El-Deiry, W.S. (2003). Microarray analysis of p53 target gene expression patterns in the spleen and thymus in response to ionizing radiation. *Cancer Biol Ther* 2, 431-443.
- Calzado, M.A., Sancho, R., and Munoz, E. (2004). Human immunodeficiency virus type 1 Tat increases the expression of cleavage and polyadenylation specificity factor 73-kilodalton subunit modulating cellular and viral expression. *J Virol* 78, 6846-6854.
- Camats, M., Guil, S., Kokolo, M., and Bach-Elias, M. (2008). P68 RNA helicase (DDX5) alters activity of cis- and trans-acting factors of the alternative splicing of H-Ras. *PLoS One* 3, e2926.
- Caputi, M., and Zahler, A.M. (2001). Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J Biol Chem* 276, 43850-43859.
- Carswell, S., and Alwine, J.C. (1989). Efficiency of utilization of the simian virus 40 late polyadenylation site: effects of upstream sequences. *Mol Cell Biol* 9, 4248-4258.
- Cartegni, L., Maconi, M., Morandi, E., Cobianchi, F., Riva, S., and Biamonti, G. (1996). hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J Mol Biol* 259, 337-348.
- Castelo-Branco, P., Furger, A., Wollerton, M., Smith, C., Moreira, A., and Proudfoot, N. (2004). Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol Cell Biol* 24, 4174-4183.
- Cevher, M.A., Zhang, X., Fernandez, S., Kim, S., Baquero, J., Nilsson, P., Lee, S., Virtanen, A., and Kleiman, F.E. (2010). Nuclear deadenylation/polyadenylation factors regulate 3' processing in response to DNA damage. *EMBO J* 29, 1674-1687.

REFERENCES

- Chehab, N.H., Malikzay, A., Appel, M., and Halazonetis, T.D. (2000). Chk2/hCds1 functions as a DNA damage checkpoint in G(1) by stabilizing p53. *Genes Dev* 14, 278-288.
- Chen, F., MacDonald, C.C., and Wilusz, J. (1995). Cleavage site determinants in the mammalian polyadenylation signal. *Nucleic Acids Res* 23, 2614-2620.
- Chen, F., and Wilusz, J. (1998). Auxiliary downstream elements are required for efficient polyadenylation of mammalian pre-mRNAs. *Nucleic Acids Res* 26, 2891-2898.
- Chen, J., and Kastan, M.B. (2010). 5'-3'-UTR interactions regulate p53 mRNA translation and provide a target for modulating p53 induction after DNA damage. *Genes Dev* 24, 2146-2156.
- Chen, Q., Reis, S.E., Kammerer, C., Craig, W.Y., LaPierre, S.E., Zimmer, E.L., McNamara, D.M., Pauly, D.F., Sharaf, B., Holubkov, R., *et al.* (2003). Genetic variation in lectin-like oxidized low-density lipoprotein receptor 1 (LOX1) gene and the risk of coronary artery disease. *Circulation* 107, 3146-3151.
- Chen, X., Ko, L.J., Jayaraman, L., and Prives, C. (1996). p53 levels, functional domains, and DNA damage determine the extent of the apoptotic response of tumor cells. *Genes Dev* 10, 2438-2451.
- Chen, Z., Li, Y., and Krug, R.M. (1999). Influenza A virus NS1 protein targets poly(A)-binding protein II of the cellular 3'-end processing machinery. *EMBO J* 18, 2273-2283.
- Cherrington, J., and Ganem, D. (1992). Regulation of polyadenylation in human immunodeficiency virus (HIV): contributions of promoter proximity and upstream sequences. *Embo J* 11, 1513-1524.
- Chodchoy, N., Pandey, N.B., and Marzluff, W.F. (1991). An intact histone 3'-processing site is required for transcription termination in a mouse histone H2a gene. *Mol Cell Biol* 11, 497-509.
- Chou, Z.F., Chen, F., and Wilusz, J. (1994). Sequence and position requirements for uridylylate-rich downstream elements of polyadenylation signals. *Nucleic Acids Res* 22, 2525-2531.
- Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325, 834-840.
- Clark, A.B., Deterding, L., Tomer, K.B., and Kunkel, T.A. (2007). Multiple functions for the N-terminal region of Msh6. *Nucleic Acids Res* 35, 4114-4123.
- Colegrove-Otero, L.J., Minshall, N., and Standart, N. (2005). RNA-binding proteins in early development. *Crit Rev Biochem Mol Biol* 40, 21-73.
- Coles, J.L., Hallegger, M., and Smith, C.W. (2009). A nonsense exon in the Tpm1 gene is silenced by hnRNP H and F. *RNA* 15, 33-43.
- Colgan, D.F., Murthy, K.G., Prives, C., and Manley, J.L. (1996). Cell-cycle related regulation of poly(A) polymerase by phosphorylation. *Nature* 384, 282-285.
- Colgan, D.F., Murthy, K.G., Zhao, W., Prives, C., and Manley, J.L. (1998). Inhibition of poly(A) polymerase requires p34cdc2/cyclin B phosphorylation of multiple consensus and non-consensus sites. *EMBO J* 17, 1053-1062.
- Collart, C., Ramis, J.M., Down, T.A., and Smith, J.C. (2009). Smc1 is required for phosphorylation of RNA polymerase II and affects 3'-end processing of RNA at the midblastula transition in *Xenopus*. *Development* 136, 3451-3461.
- Coller, J.M., Gray, N.K., and Wickens, M.P. (1998). mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. *Genes Dev* 12, 3226-3235.

REFERENCES

- Connelly, S., and Manley, J.L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev* 2, 440-452.
- Cooke, C., and Alwine, J.C. (1996). The cap and the 3' splice site similarly affect polyadenylation efficiency. *Mol Cell Biol* 16, 2579-2584.
- Cooke, C., Hans, H., and Alwine, J.C. (1999). Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. *Mol Cell Biol* 19, 4971-4979.
- Cools, J., DeAngelo, D.J., Gotlib, J., Stover, E.H., Legare, R.D., Cortes, J., Kutok, J., Clark, J., Galinsky, I., Griffin, J.D., *et al.* (2003). A tyrosine kinase created by fusion of the PDGFRA and FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. *N Engl J Med* 348, 1201-1214.
- Coutinho, D.C., Coletta, R.R., Costa, E.M., Pachi, P.R., Boguszewski, M.C., Damiani, D., Mendonca, B.B., Arnhold, I.J., and Jorge, A.A. (2007). Polymorphisms identified in the upstream core polyadenylation signal of IGF1 gene exon 6 do not cause pre- and postnatal growth impairment. *J Clin Endocrinol Metab* 92, 4889-4892.
- Cox, D.G., Pontes, C., Guino, E., Navarro, M., Osorio, A., Canzian, F., and Moreno, V. (2004). Polymorphisms in prostaglandin synthase 2/cyclooxygenase 2 (PTGS2/COX2) and risk of colorectal cancer. *Br J Cancer* 91, 339-343.
- Cramer, P., Srebrow, A., Kadener, S., Werbajh, S., de la Mata, M., Melen, G., Nogues, G., and Kornbliht, A.R. (2001). Coordination between transcription and pre-mRNA processing. *FEBS Lett* 498, 179-182.
- Crnugelj, M., Hud, N.V., and Plavec, J. (2002). The solution structure of d(G(4)T(4)G(3))(2): a bimolecular G-quadruplex with a novel fold. *J Mol Biol* 320, 911-924.
- Dalziel, M., Nunes, N.M., and Furger, A. (2007). Two G-rich regulatory elements located adjacent to and 440 nucleotides downstream of the core poly(A) site of the intronless melanocortin receptor 1 gene are critical for efficient 3' end processing. *Mol Cell Biol* 27, 1568-1580.
- Danckwardt, S., Gantzert, A.S., Macher-Goeppinger, S., Probst, H.C., Gentzel, M., Wilm, M., Grone, H.J., Schirmacher, P., Hentze, M.W., and Kulozik, A.E. (2011). p38 MAPK controls prothrombin expression by regulated RNA 3' end processing. *Mol Cell* 41, 298-310.
- Danckwardt, S., Gehring, N.H., Neu-Yilik, G., Hundsdoerfer, P., Pforsich, M., Frede, U., Hentze, M.W., and Kulozik, A.E. (2004). The prothrombin 3' end formation signal reveals a unique architecture that is sensitive to thrombophilic gain-of-function mutations. *Blood* 104, 428-435.
- Danckwardt, S., Hartmann, K., Gehring, N.H., Hentze, M.W., and Kulozik, A.E. (2006a). 3' end processing of the prothrombin mRNA in thrombophilia. *Acta Haematol* 115, 192-197.
- Danckwardt, S., Hartmann, K., Katz, B., Hentze, M.W., Levy, Y., Eichele, R., Deutsch, V., Kulozik, A.E., and Ben-Tal, O. (2006b). The prothrombin 20209 C->T mutation in Jewish-Moroccan Caucasians: molecular analysis of gain-of-function of 3' end processing. *J Thromb Haemost* 4, 1078-1085.
- Danckwardt, S., Hentze, M.W., and Kulozik, A.E. (2008). 3' end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J* 27, 482-498.
- Danckwardt, S., Kaufmann, I., Gentzel, M., Foerstner, K.U., Gantzert, A.S., Gehring, N.H., Neu-Yilik, G., Bork, P., Keller, W., Wilm, M., *et al.* (2007). Splicing

REFERENCES

- factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals. *EMBO J* 26, 2658-2669.
- Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P., and Reed, R. (2007). SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol Cell* 26, 867-881.
- Dass, B., Tardif, S., Park, J.Y., Tian, B., Weitlauf, H.M., Hess, R.A., Carnes, K., Griswold, M.D., Small, C.L., and Macdonald, C.C. (2007). Loss of polyadenylation protein tauCstF-64 causes spermatogenic defects and male infertility. *Proc Natl Acad Sci U S A* 104, 20374-20379.
- de Vries, H., Ruegsegger, U., Hubner, W., Friedlein, A., Langen, H., and Keller, W. (2000). Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *Embo J* 19, 5895-5904.
- Deb, S.P. (2002). Function and dysfunction of the human oncoprotein MDM2. *Front Biosci* 7, d235-243.
- Decker, E.D., Zhang, Y., Cocklin, R.R., Witzmann, F.A., and Wang, M. (2003). Proteomic analysis of differential protein expression induced by ultraviolet light radiation in HeLa cells. *Proteomics* 3, 2019-2027.
- Derry, M.C., Yanagiya, A., Martineau, Y., and Sonenberg, N. (2006). Regulation of poly(A)-binding protein through PABP-interacting proteins. *Cold Spring Harb Symp Quant Biol* 71, 537-543.
- DeZazzo, J.D., Kilpatrick, J.E., and Imperiale, M.J. (1991). Involvement of long terminal repeat U3 sequences overlapping the transcription control region in human immunodeficiency virus type 1 mRNA 3' end formation. *Mol Cell Biol* 11, 1624-1630.
- Dominguez, C., and Allain, F.H. (2006). NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition. *Nucleic Acids Res* 34, 3634-3645.
- 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.
- Dominski, Z., and Marzluff, W.F. (2007). Formation of the 3' end of histone mRNA: getting closer to the end. *Gene* 396, 373-390.
- Dominski, Z., Yang, X.C., and Marzluff, W.F. (2005a). The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. *Cell* 123, 37-48.
- Dominski, Z., Yang, X.C., Purdy, M., Wagner, E.J., and Marzluff, W.F. (2005b). A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. *Mol Cell Biol* 25, 1489-1500.
- Dower, K., and Rosbash, M. (2002). T7 RNA polymerase-directed transcripts are processed in yeast and link 3' end formation to mRNA nuclear export. *RNA* 8, 686-697.
- Dreyfuss, G., Matunis, M.J., Pinol-Roma, S., and Burd, C.G. (1993). hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* 62, 289-321.
- Drygin, D., Siddiqui-Jain, A., O'Brien, S., Schwaebe, M., Lin, A., Bliesath, J., Ho, C.B., Proffitt, C., Trent, K., Whitten, J.P., *et al.* (2009). Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer Res* 69, 7653-7661.
- Eddy, J., and Maizels, N. (2006). Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res* 34, 3887-3896.

REFERENCES

- Edwalds-Gilbert, G., and Milcarek, C. (1995). The binding of a subunit of the general polyadenylation factor cleavage-polyadenylation specificity factor (CPSF) to polyadenylation sites changes during B cell development. *Nucleic Acids Symp Ser*, 229-233.
- Edwalds-Gilbert, G., Veraldi, K.L., and Milcarek, C. (1997). Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res* 25, 2547-2561.
- Edwards, R.A., Lee, M.S., Tsutakawa, S.E., Williams, R.S., Nazeer, I., Kleiman, F.E., Tainer, J.A., and Glover, J.N. (2008). The BARD1 C-terminal domain structure and interactions with polyadenylation factor CstF-50. *Biochemistry* 47, 11446-11456.
- Enriquez-Harris, P., Levitt, N., Briggs, D., and Proudfoot, N.J. (1991). A pause site for RNA polymerase II is associated with termination of transcription. *EMBO J* 10, 1833-1842.
- Eulalio, A., Huntzinger, E., and Izaurralde, E. (2008). Getting to the root of miRNA-mediated gene silencing. *Cell* 132, 9-14.
- Fei, P., Bernhard, E.J., and El-Deiry, W.S. (2002). Tissue-specific induction of p53 targets in vivo. *Cancer Res* 62, 7316-7327.
- Fendri-Kriaa, N., Kammoun, F., Salem, I.H., Kifagi, C., Mkaouar-Rebai, E., Hsairi, I., Rebai, A., Triki, C., and Fakhfakh, F. (2011). New mutation c.374C>T and a putative disease-associated haplotype within SCN1B gene in Tunisian families with febrile seizures. *Eur J Neurol* 18, 695-702.
- Flaherty, S.M., Fortes, P., Izaurralde, E., Mattaj, I.W., and Gilmartin, G.M. (1997). Participation of the nuclear cap binding complex in pre-mRNA 3' processing. *Proc Natl Acad Sci U S A* 94, 11893-11898.
- Ford, L.P., Bagga, P.S., and Wilusz, J. (1997). The poly(A) tail inhibits the assembly of a 3'-to-5' exonuclease in an in vitro RNA stability system. *Mol Cell Biol* 17, 398-406.
- Fox, J.T., Shin, W.K., Caudill, M.A., and Stover, P.J. (2009). A UV-responsive internal ribosome entry site enhances serine hydroxymethyltransferase 1 expression for DNA damage repair. *J Biol Chem* 284, 31097-31108.
- Fox, J.T., and Stover, P.J. (2009). Mechanism of the internal ribosome entry site-mediated translation of serine hydroxymethyltransferase 1. *J Biol Chem* 284, 31085-31096.
- Fritsche, L.G., Loenhardt, T., Janssen, A., Fisher, S.A., Rivera, A., Keilhauer, C.N., and Weber, B.H. (2008). Age-related macular degeneration is associated with an unstable ARMS2 (LOC387715) mRNA. *Nat Genet* 40, 892-896.
- Gallie, D.R. (1998). A tale of two termini: a functional interaction between the termini of an mRNA is a prerequisite for efficient translation initiation. *Gene* 216, 1-11.
- Garneau, D., Revil, T., Fiset, J.F., and Chabot, B. (2005). Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J Biol Chem* 280, 22641-22650.
- Garneau, N.L., Wilusz, J., and Wilusz, C.J. (2007). The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* 8, 113-126.
- Gassman, N.R., Clodfelter, J.E., McCauley, A.K., Bonin, K., Salsbury, F.R., Jr., and Scarpinato, K.D. (2011). Cooperative nuclear localization sequences lend a novel role to the N-terminal region of MSH6. *PLoS One* 6, e17907.
- Gehring, N.H., Frede, U., Neu-Yilik, G., Hundsdoerfer, P., Vetter, B., Hentze, M.W., and Kulozik, A.E. (2001). Increased efficiency of mRNA 3' end formation: a

REFERENCES

- new genetic mechanism contributing to hereditary thrombophilia. *Nat Genet* 28, 389-392.
- Gemignani, F., Moreno, V., Landi, S., Moullan, N., Chabrier, A., Gutierrez-Enriquez, S., Hall, J., Guino, E., Peinado, M.A., Capella, G., *et al.* (2004). A TP53 polymorphism is associated with increased risk of colorectal cancer and with reduced levels of TP53 mRNA. *Oncogene* 23, 1954-1956.
- Gil, A., and Proudfoot, N.J. (1987). Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation. *Cell* 49, 399-406.
- Gilmartin, G.M., McDevitt, M.A., and Nevins, J.R. (1988). Multiple factors are required for specific RNA cleavage at a poly(A) addition site. *Genes Dev* 2, 578-587.
- Giovannucci, E., Colditz, G.A., Stampfer, M.J., Hunter, D., Rosner, B.A., Willett, W.C., and Speizer, F.E. (1994). A prospective study of cigarette smoking and risk of colorectal adenoma and colorectal cancer in U.S. women. *J Natl Cancer Inst* 86, 192-199.
- Glorian, V., Maillot, G., Poles, S., Iacovoni, J.S., Favre, G., and Vagner, S. (2011). HuR-dependent loading of miRNA RISC to the mRNA encoding the Ras-related small GTPase RhoB controls its translation during UV-induced apoptosis. *Cell Death Differ*.
- Glover-Cutter, K., Kim, S., Espinosa, J., and Bentley, D.L. (2008). RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat Struct Mol Biol* 15, 71-78.
- Gomes, N.P., Bjerke, G., Llorente, B., Szostek, S.A., Emerson, B.M., and Espinosa, J.M. (2006). Gene-specific requirement for P-TEFb activity and RNA polymerase II phosphorylation within the p53 transcriptional program. *Genes Dev* 20, 601-612.
- Gomez, D., Lemarteleur, T., Lacroix, L., Mailliet, P., Mergny, J.L., and Riou, J.F. (2004). Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res* 32, 371-379.
- Gromak, N., West, S., and Proudfoot, N.J. (2006). Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol Cell Biol* 26, 3986-3996.
- Grosset, C., Boniface, R., Duchez, P., Solanilla, A., Cosson, B., and Ripoche, J. (2004). In vivo studies of translational repression mediated by the granulocyte-macrophage colony-stimulating factor AU-rich element. *J Biol Chem* 279, 13354-13362.
- Gu, H., and Schoenberg, D.R. (2003). U2AF modulates poly(A) length control by the poly(A)-limiting element. *Nucleic Acids Res* 31, 6264-6271.
- Guang, S., Felthouser, A.M., and Mertz, J.E. (2005). Binding of hnRNP L to the pre-mRNA processing enhancer of the herpes simplex virus thymidine kinase gene enhances both polyadenylation and nucleocytoplasmic export of intronless mRNAs. *Mol Cell Biol* 25, 6303-6313.
- Gudipati, R.K., Villa, T., Boulay, J., and Libri, D. (2008). Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nat Struct Mol Biol* 15, 786-794.
- Gunderson, S.I., Beyer, K., Martin, G., Keller, W., Boelens, W.C., and Mattaj, L.W. (1994). The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell* 76, 531-541.

REFERENCES

- Gunderson, S.I., Polycarpou-Schwarz, M., and Mattaj, I.W. (1998). U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol Cell* 1, 255-264.
- Gunderson, S.I., Vagner, S., Polycarpou-Schwarz, M., and Mattaj, I.W. (1997). Involvement of the carboxyl terminus of vertebrate poly(A) polymerase in U1A autoregulation and in the coupling of splicing and polyadenylation. *Genes Dev* 11, 761-773.
- Gyawali, S., Subaran, R., Weissman, M.M., Hershkowitz, D., McKenna, M.C., Talati, A., Fyer, A.J., Wickramaratne, P., Adams, P.B., Hodge, S.E., *et al.* (2010). Association of a polyadenylation polymorphism in the serotonin transporter and panic disorder. *Biol Psychiatry* 67, 331-338.
- Halder, K., Wieland, M., and Hartig, J.S. (2009). Predictable suppression of gene expression by 5'-UTR-based RNA quadruplexes. *Nucleic Acids Res* 37, 6811-6817.
- Hall-Pogar, T., Liang, S., Hague, L.K., and Lutz, C.S. (2007). Specific trans-acting proteins interact with auxiliary RNA polyadenylation elements in the COX-2 3'-UTR. *Rna* 13, 1103-1115.
- Hammell, C.M., Gross, S., Zenklusen, D., Heath, C.V., Stutz, F., Moore, C., and Cole, C.N. (2002). Coupling of termination, 3' processing, and mRNA export. *Mol Cell Biol* 22, 6441-6457.
- Hanakahi, L.A., Sun, H., and Maizels, N. (1999). High affinity interactions of nucleolin with G-G-paired rDNA. *J Biol Chem* 274, 15908-15912.
- Harfe, B.D., and Jinks-Robertson, S. (2000). DNA mismatch repair and genetic instability. *Annu Rev Genet* 34, 359-399.
- Hart, R.P., McDevitt, M.A., and Nevins, J.R. (1985). Poly(A) site cleavage in a HeLa nuclear extract is dependent on downstream sequences. *Cell* 43, 677-683.
- Harteveld, C.L., Losekoot, M., Haak, H., Heister, G.A., Giordano, P.C., and Bernini, L.F. (1994). A novel polyadenylation signal mutation in the alpha 2-globin gene causing alpha thalassaemia. *Br J Haematol* 87, 139-143.
- Hausmann, S., Koiwa, H., Krishnamurthy, S., Hampsey, M., and Shuman, S. (2005). Different strategies for carboxyl-terminal domain (CTD) recognition by serine 5-specific CTD phosphatases. *J Biol Chem* 280, 37681-37688.
- Hein, D.W., Doll, M.A., Rustan, T.D., Gray, K., Feng, Y., Ferguson, R.J., and Grant, D.M. (1993). Metabolic activation and deactivation of arylamine carcinogens by recombinant human NAT1 and polymorphic NAT2 acetyltransferases. *Carcinogenesis* 14, 1633-1638.
- Hellin, A.C., Calmant, P., Gielen, J., Bours, V., and Merville, M.P. (1998). Nuclear factor - kappaB-dependent regulation of p53 gene expression induced by daunomycin genotoxic drug. *Oncogene* 16, 1187-1195.
- Hennig, B.J., Hellier, S., Frodsham, A.J., Zhang, L., Klenerman, P., Knapp, S., Wright, M., Thomas, H.C., Thursz, M., and Hill, A.V. (2002). Association of low-density lipoprotein receptor polymorphisms and outcome of hepatitis C infection. *Genes Immun* 3, 359-367.
- Higgs, D.R., Goodbourn, S.E., Lamb, J., Clegg, J.B., Weatherall, D.J., and Proudfoot, N.J. (1983). Alpha-thalassaemia caused by a polyadenylation signal mutation. *Nature* 306, 398-400.
- Hilleren, P., McCarthy, T., Rosbash, M., Parker, R., and Jensen, T.H. (2001). Quality control of mRNA 3'-end processing is linked to the nuclear exosome. *Nature* 413, 538-542.

REFERENCES

- Hirose, Y., and Manley, J.L. (2000). RNA polymerase II and the integration of nuclear events. *Genes Dev* *14*, 1415-1429.
- Hollingworth, D., Noble, C.G., Taylor, I.A., and Ramos, A. (2006). RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11. *Rna* *12*, 555-560.
- Honore, B., Baandrup, U., and Vorum, H. (2004). Heterogeneous nuclear ribonucleoproteins F and H/H' show differential expression in normal and selected cancer tissues. *Exp Cell Res* *294*, 199-209.
- Hsieh, P. (2001). Molecular mechanisms of DNA mismatch repair. *Mutat Res* *486*, 71-87.
- Hu, J., Lutz, C.S., Wilusz, J., and Tian, B. (2005). Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *Rna* *11*, 1485-1493.
- Huang, Y., and Carmichael, G.G. (1996). Role of polyadenylation in nucleocytoplasmic transport of mRNA. *Mol Cell Biol* *16*, 1534-1542.
- Hulboy, D.L., and Lozano, G. (1994). Structural and functional analysis of p53: the acidic activation domain has transforming capability. *Cell Growth Differ* *5*, 1023-1031.
- Hung, L.H., Heiner, M., Hui, J., Schreiner, S., Benes, V., and Bindereif, A. (2008). Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *RNA* *14*, 284-296.
- Hung, R.J., Boffetta, P., Canzian, F., Moullan, N., Szeszenia-Dabrowska, N., Zaridze, D., Lissowska, J., Rudnai, P., Fabianova, E., Mates, D., *et al.* (2006). Sequence variants in cell cycle control pathway, X-ray exposure, and lung cancer risk: a multicenter case-control study in Central Europe. *Cancer Res* *66*, 8280-8286.
- Jacob, S.T., Terns, M.P., and Maguire, K.A. (1989). Polyadenylate polymerases from normal and cancer cells and their potential role in messenger RNA processing: a review. *Cancer Res* *49*, 2827-2833.
- Jankovic, L., Efremov, G.D., Petkov, G., Kattamis, C., George, E., Yang, K.G., Stoming, T.A., and Huisman, T.H. (1990). Two novel polyadenylation mutations leading to beta(+)-thalassemia. *Br J Haematol* *75*, 122-126.
- Janssens, S., and Tschopp, J. (2006). Signals from within: the DNA-damage-induced NF-kappaB response. *Cell Death Differ* *13*, 773-784.
- Jenny, A., Hauri, H.P., and Keller, W. (1994). Characterization of cleavage and polyadenylation specificity factor and cloning of its 100-kilodalton subunit. *Mol Cell Biol* *14*, 8183-8190.
- Ji, Z., Lee, J.Y., Pan, Z., Jiang, B., and Tian, B. (2009). Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci U S A* *106*, 7028-7033.
- Jin, Y., Xu, X.L., Yang, M.C., Wei, F., Ayi, T.C., Bowcock, A.M., and Baer, R. (1997). Cell cycle-dependent colocalization of BARD1 and BRCA1 proteins in discrete nuclear domains. *Proc Natl Acad Sci U S A* *94*, 12075-12080.
- Jiricny, J. (1998). Replication errors: cha(lle)nging the genome. *EMBO J* *17*, 6427-6436.
- Jung, J.H., Lee, S.M., Bae, S., Lee, S.J., Park, I.C., Jin, Y.W., Lee, J.H., and An, S. (2010). Triad 1 induces apoptosis by p53 activation. *FEBS Lett* *584*, 1565-1570.
- Kahvejian, A., Roy, G., and Sonenberg, N. (2001). The mRNA closed-loop model: the function of PABP and PABP-interacting proteins in mRNA translation. *Cold Spring Harb Symp Quant Biol* *66*, 293-300.

REFERENCES

- Kahvejian, A., Svitkin, Y.V., Sukarieh, R., M'Boutchou, M.N., and Sonenberg, N. (2005). Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes Dev* *19*, 104-113.
- Kameoka, S., Duque, P., and Konarska, M.M. (2004). p54(nrb) associates with the 5' splice site within large transcription/splicing complexes. *EMBO J* *23*, 1782-1791.
- Kaufmann, I., Martin, G., Friedlein, A., Langen, H., and Keller, W. (2004). Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *Embo J* *23*, 616-626.
- Kiefer, H., Mizutani, A., Iemura, S., Natsume, T., Ando, H., Kuroda, Y., and Mikoshiba, K. (2009). Inositol 1,4,5-triphosphate receptor-binding protein released with inositol 1,4,5-triphosphate (IRBIT) associates with components of the mRNA 3' processing machinery in a phosphorylation-dependent manner and inhibits polyadenylation. *J Biol Chem* *284*, 10694-10705.
- Kim, D.H., Langlois, M.A., Lee, K.B., Riggs, A.D., Puymirat, J., and Rossi, J.J. (2005). HnRNP H inhibits nuclear export of mRNA containing expanded CUG repeats and a distal branch point sequence. *Nucleic Acids Res* *33*, 3866-3874.
- Kim, H., Lee, J.H., and Lee, Y. (2003). Regulation of poly(A) polymerase by 14-3-3epsilon. *EMBO J* *22*, 5208-5219.
- Kim, H.S., Li, H., Cevher, M., Parmelee, A., Fonseca, D., Kleiman, F.E., and Lee, S.B. (2006). DNA damage-induced BARD1 phosphorylation is critical for the inhibition of messenger RNA processing by BRCA1/BARD1 complex. *Cancer Res* *66*, 4561-4565.
- Kim, M.Y., Vankayalapati, H., Shin-Ya, K., Wierzbza, K., and Hurley, L.H. (2002). Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular g-quadruplex. *J Am Chem Soc* *124*, 2098-2099.
- Kimble, J., and Crittenden, S.L. (2007). Controls of germline stem cells, entry into meiosis, and the sperm/oocyte decision in *Caenorhabditis elegans*. *Annu Rev Cell Dev Biol* *23*, 405-433.
- Kleiman, F.E., and Manley, J.L. (1999). Functional interaction of BRCA1-associated BARD1 with polyadenylation factor CstF-50. *Science* *285*, 1576-1579.
- Kleiman, F.E., and Manley, J.L. (2001). The BARD1-CstF-50 interaction links mRNA 3' end formation to DNA damage and tumor suppression. *Cell* *104*, 743-753.
- Kleiman, F.E., Wu-Baer, F., Fonseca, D., Kaneko, S., Baer, R., and Manley, J.L. (2005). BRCA1/BARD1 inhibition of mRNA 3' processing involves targeted degradation of RNA polymerase II. *Genes Dev* *19*, 1227-1237.
- Ko, B., and Gunderson, S.I. (2002). Identification of new poly(A) polymerase-inhibitory proteins capable of regulating pre-mRNA polyadenylation. *J Mol Biol* *318*, 1189-1206.
- Kolev, N.G., and Steitz, J.A. (2005). Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs. *Genes Dev* *19*, 2583-2592.
- Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* *14*, 2452-2460.
- Kostadinov, R., Malhotra, N., Viotti, M., Shine, R., D'Antonio, L., and Bagga, P. (2006). GRSDDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. *Nucleic Acids Res* *34*, D119-124.

REFERENCES

- Krecic, A.M., and Swanson, M.S. (1999). hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol* 11, 363-371.
- Kruse, J.P., and Gu, W. (2009). Modes of p53 regulation. *Cell* 137, 609-622.
- Kuhn, U., Gundel, M., Knoth, A., Kerwitz, Y., Rudel, S., and Wahle, E. (2009). Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. *J Biol Chem* 284, 22803-22814.
- Kumari, S., Bugaut, A., Huppert, J.L., and Balasubramanian, S. (2007). An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat Chem Biol* 3, 218-221.
- Kuroki, K., Tsuchiya, N., Tsao, B.P., Grossman, J.M., Fukazawa, T., Hagiwara, K., Kano, H., Takazoe, M., Iwata, T., Hashimoto, H., *et al.* (2002). Polymorphisms of human CD19 gene: possible association with susceptibility to systemic lupus erythematosus in Japanese. *Genes Immun* 3 *Suppl 1*, S21-30.
- Kyburz, A., Friedlein, A., Langen, H., and Keller, W. (2006). Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. *Mol Cell* 23, 195-205.
- Lambert, J.C., Luedeking-Zimmer, E., Merrot, S., Hayes, A., Thaker, U., Desai, P., Houzet, A., Hermant, X., Cotel, D., Pritchard, A., *et al.* (2003). Association of 3'-UTR polymorphisms of the oxidised LDL receptor 1 (OLR1) gene with Alzheimer's disease. *J Med Genet* 40, 424-430.
- Lazarowitz, S.G., Compans, R.W., and Choppin, P.W. (1971). Influenza virus structural and nonstructural proteins in infected cells and their plasma membranes. *Virology* 46, 830-843.
- Le, M.T., Teh, C., Shyh-Chang, N., Xie, H., Zhou, B., Korzh, V., Lodish, H.F., and Lim, B. (2009). MicroRNA-125b is a novel negative regulator of p53. *Genes Dev* 23, 862-876.
- Le Sommer, C., Lesimple, M., Mereau, A., Menoret, S., Allo, M.R., and Hardy, S. (2005). PTB regulates the processing of a 3'-terminal exon by repressing both splicing and polyadenylation. *Mol Cell Biol* 25, 9595-9607.
- Legendre, M., Lambert, A., and Gautheret, D. (2005). Profile-based detection of microRNA precursors in animal genomes. *Bioinformatics* 21, 841-845.
- Lemmers, R.J., van der Vliet, P.J., Klooster, R., Sacconi, S., Camano, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., van Ommen, G.J., Padberg, G.W., *et al.* (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* 329, 1650-1653.
- Lewis, J.D., and Izaurralde, E. (1997). The role of the cap structure in RNA processing and nuclear export. *Eur J Biochem* 247, 461-469.
- Liang, S., and Lutz, C.S. (2006). p54nrb is a component of the snRNP-free U1A (SF-A) complex that promotes pre-mRNA cleavage during polyadenylation. *RNA* 12, 111-121.
- Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., *et al.* (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456, 464-469.
- Listerman, I., Sapra, A.K., and Neugebauer, K.M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat Struct Mol Biol* 13, 815-822.

REFERENCES

- Liu, H., Hew, H.C., Lu, Z.G., Yamaguchi, T., Miki, Y., and Yoshida, K. (2009). DNA damage signalling recruits RREB-1 to the p53 tumour suppressor promoter. *Biochem J* 422, 543-551.
- Logan, J., Falck-Pedersen, E., Darnell, J.E., Jr., and Shenk, T. (1987). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc Natl Acad Sci U S A* 84, 8306-8310.
- Losekoot, M., Fodde, R., Harteveld, C.L., van Heeren, H., Giordano, P.C., Went, L.N., and Bernini, L.F. (1991). Homozygous beta+ thalassaemia owing to a mutation in the cleavage-polyadenylation sequence of the human beta globin gene. *J Med Genet* 28, 252-255.
- Lou, H., Gagel, R.F., and Berget, S.M. (1996). An intron enhancer recognized by splicing factors activates polyadenylation. *Genes Dev* 10, 208-219.
- Lou, H., Helfman, D.M., Gagel, R.F., and Berget, S.M. (1999). Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3'-terminal exon. *Mol Cell Biol* 19, 78-85.
- Lou, H., Neugebauer, K.M., Gagel, R.F., and Berget, S.M. (1998). Regulation of alternative polyadenylation by U1 snRNPs and SRp20. *Mol Cell Biol* 18, 4977-4985.
- Lou, H., Yang, Y., Cote, G.J., Berget, S.M., and Gagel, R.F. (1995). An intron enhancer containing a 5' splice site sequence in the human calcitonin/calcitonin gene-related peptide gene. *Mol Cell Biol* 15, 7135-7142.
- Lu, X., Errington, J., Curtin, N.J., Lunec, J., and Newell, D.R. (2001). The impact of p53 status on cellular sensitivity to antifolate drugs. *Clin Cancer Res* 7, 2114-2123.
- Luo, W., and Bentley, D. (2004). A ribonucleolytic rat torpedoed RNA polymerase II. *Cell* 119, 911-914.
- Luo, W., Johnson, A.W., and Bentley, D.L. (2006). The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev* 20, 954-965.
- Lutz, C.S., and Alwine, J.C. (1994). Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal. *Genes Dev* 8, 576-586.
- Lutz, C.S., Murthy, K.G., Schek, N., O'Connor, J.P., Manley, J.L., and Alwine, J.C. (1996). Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro. *Genes Dev* 10, 325-337.
- Maciolek, N.L., and McNally, M.T. (2007). Serine/arginine-rich proteins contribute to negative regulator of splicing element-stimulated polyadenylation in rous sarcoma virus. *J Virol* 81, 11208-11217.
- Malorni, L., Cacace, G., Cuccurullo, M., Pocsfalvi, G., Chambery, A., Farina, A., Di Maro, A., Parente, A., and Malorni, A. (2006). Proteomic analysis of MCF-7 breast cancer cell line exposed to mitogenic concentration of 17beta-estradiol. *Proteomics* 6, 5973-5982.
- Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci* 65, 1099-1122.
- Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* 444, 953-956.

REFERENCES

- Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature* *416*, 499-506.
- Manley, J.L. (1988). Polyadenylation of mRNA precursors. *Biochim Biophys Acta* *950*, 1-12.
- Mapendano, C.K., Lykke-Andersen, S., Kjems, J., Bertrand, E., and Jensen, T.H. (2010). Crosstalk between mRNA 3' end processing and transcription initiation. *Mol Cell* *40*, 410-422.
- Marcel, V., Tran, P.L., Sagne, C., Martel-Planche, G., Vaslin, L., Teulade-Fichou, M.P., Hall, J., Mergny, J.L., Hainaut, P., and Van Dyck, E. (2010). G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis* *32*, 271-278.
- Martin, G., Moglich, A., Keller, W., and Doublié, S. (2004). Biochemical and structural insights into substrate binding and catalytic mechanism of mammalian poly(A) polymerase. *J Mol Biol* *341*, 911-925.
- Martincic, K., Alkan, S.A., Cheatle, A., Borghesi, L., and Milcarek, C. (2009). Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. *Nat Immunol* *10*, 1102-1109.
- Martincic, K., Campbell, R., Edwalds-Gilbert, G., Souan, L., Lotze, M.T., and Milcarek, C. (1998). Increase in the 64-kDa subunit of the polyadenylation/cleavage stimulatory factor during the G0 to S phase transition. *Proc Natl Acad Sci U S A* *95*, 11095-11100.
- Marzluff, W.F., Wagner, E.J., and Duronio, R.J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet* *9*, 843-854.
- Mason, P.J., Elkington, J.A., Lloyd, M.M., Jones, M.B., and Williams, J.G. (1986). Mutations downstream of the polyadenylation site of a *Xenopus* beta-globin mRNA affect the position but not the efficiency of 3' processing. *Cell* *46*, 263-270.
- Mayr, C., and Bartel, D.P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* *138*, 673-684.
- Mazan-Mamczarz, K., Galban, S., Lopez de Silanes, I., Martindale, J.L., Atasoy, U., Keene, J.D., and Gorospe, M. (2003). RNA-binding protein HuR enhances p53 translation in response to ultraviolet light irradiation. *Proc Natl Acad Sci U S A* *100*, 8354-8359.
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* *385*, 357-361.
- McCracken, S., Lambermon, M., and Blencowe, B.J. (2002). SRm160 splicing coactivator promotes transcript 3'-end cleavage. *Mol Cell Biol* *22*, 148-160.
- McCracken, S., Longman, D., Johnstone, I.L., Caceres, J.F., and Blencowe, B.J. (2003). An evolutionarily conserved role for SRm160 in 3'-end processing that functions independently of exon junction complex formation. *J Biol Chem* *278*, 44153-44160.
- McDevitt, M.A., Hart, R.P., Wong, W.W., and Nevins, J.R. (1986). Sequences capable of restoring poly(A) site function define two distinct downstream elements. *Embo J* *5*, 2907-2913.

REFERENCES

- McGregor, F., Phelan, A., Dunlop, J., and Clements, J.B. (1996). Regulation of herpes simplex virus poly (A) site usage and the action of immediate-early protein IE63 in the early-late switch. *J Virol* 70, 1931-1940.
- McLaren, M., Asai, K., and Cochrane, A. (2004). A novel function for Sam68: enhancement of HIV-1 RNA 3' end processing. *RNA* 10, 1119-1129.
- McLaren, M., and Cochrane, A. (2009). Mapping of determinants involved in the stimulation of HIV-1 expression by Sam68. *Virology* 385, 93-104.
- McLauchlan, J., Gaffney, D., Whitton, J.L., and Clements, J.B. (1985). The consensus sequence YGTGTTY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucleic Acids Res* 13, 1347-1368.
- McLauchlan, J., Phelan, A., Loney, C., Sandri-Goldin, R.M., and Clements, J.B. (1992). Herpes simplex virus IE63 acts at the posttranscriptional level to stimulate viral mRNA 3' processing. *J Virol* 66, 6939-6945.
- McLauchlan, J., Simpson, S., and Clements, J.B. (1989). Herpes simplex virus induces a processing factor that stimulates poly(A) site usage. *Cell* 59, 1093-1105.
- Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* 430, 223-226.
- Mellman, D.L., Gonzales, M.L., Song, C., Barlow, C.A., Wang, P., Kendziorski, C., and Anderson, R.A. (2008). A PtdIns4,5P2-regulated nuclear poly(A) polymerase controls expression of select mRNAs. *Nature* 451, 1013-1017.
- Millevoi, S., Decorsiere, A., Loulergue, C., Iacovoni, J., Bernat, S., Antoniou, M., and Vagner, S. (2009). A physical and functional link between splicing factors promotes pre-mRNA 3' end processing. *Nucleic Acids Res* 37, 4672-4683.
- Millevoi, S., Geraghty, F., Idowu, B., Tam, J.L., Antoniou, M., and Vagner, S. (2002). A novel function for the U2AF 65 splicing factor in promoting pre-mRNA 3'-end processing. *EMBO Rep* 3, 869-874.
- Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S.Z., Keller, W., Antoniou, M., and Vagner, S. (2006). An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. *Embo J* 25, 4854-4864.
- Millevoi, S., and Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res* 38, 2757-2774.
- Min, H., Chan, R.C., and Black, D.L. (1995). The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event. *Genes Dev* 9, 2659-2671.
- Mirkin, N., Fonseca, D., Mohammed, S., Cevher, M.A., Manley, J.L., and Kleiman, F.E. (2008). The 3' processing factor CstF functions in the DNA repair response. *Nucleic Acids Res* 36, 1792-1804.
- Moore, C.L., and Sharp, P.A. (1985). Accurate cleavage and polyadenylation of exogenous RNA substrate. *Cell* 41, 845-855.
- Moore, C.L., Skolnik-David, H., and Sharp, P.A. (1986). Analysis of RNA cleavage at the adenovirus-2 L3 polyadenylation site. *Embo J* 5, 1929-1938.
- Moreira, A., Takagaki, Y., Brackenridge, S., Wollerton, M., Manley, J.L., and Proudfoot, N.J. (1998). The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. *Genes Dev* 12, 2522-2534.
- Moreira, A., Wollerton, M., Monks, J., and Proudfoot, N.J. (1995). Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved. *Embo J* 14, 3809-3819.

REFERENCES

- Morris, M.J., Negishi, Y., Pázsint, C., Schonhoff, J.D., and Basu, S. (2010). An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. *J Am Chem Soc* *132*, 17831-17839.
- Moucadel, V., Lopez, F., Ara, T., Benech, P., and Gautheret, D. (2007). Beyond the 3' end: experimental validation of extended transcript isoforms. *Nucleic Acids Res* *35*, 1947-1957.
- Mowry, K.L., and Steitz, J.A. (1987). Both conserved signals on mammalian histone pre-mRNAs associate with small nuclear ribonucleoproteins during 3' end formation in vitro. *Mol Cell Biol* *7*, 1663-1672.
- Murthy, K.G., and Manley, J.L. (1992). Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. *J Biol Chem* *267*, 14804-14811.
- Murthy, K.G., and Manley, J.L. (1995). The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev* *9*, 2672-2683.
- Natalizio, B.J., Muniz, L.C., Arhin, G.K., Wilusz, J., and Lutz, C.S. (2002). Upstream elements present in the 3'-untranslated region of collagen genes influence the processing efficiency of overlapping polyadenylation signals. *J Biol Chem* *277*, 42733-42740.
- Navakauskiene, R., Treigyte, G., Gineitis, A., and Magnusson, K.E. (2004). Identification of apoptotic tyrosine-phosphorylated proteins after etoposide or retinoic acid treatment. *Proteomics* *4*, 1029-1041.
- Nazeer, F.I., Devany, E., Mohammed, S., Fonseca, D., Akukwe, B., Taveras, C., and Kleiman, F.E. (2011). p53 inhibits mRNA 3' processing through its interaction with the CstF/BARD1 complex. *Oncogene*.
- Neidle, S. (2009). The structures of quadruplex nucleic acids and their drug complexes. *Curr Opin Struct Biol* *19*, 239-250.
- Nemeroff, M.E., Barabino, S.M., Li, Y., Keller, W., and Krug, R.M. (1998). Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3' end formation of cellular pre-mRNAs. *Mol Cell* *1*, 991-1000.
- Neugebauer, K.M., and Roth, M.B. (1997). Transcription units as RNA processing units. *Genes Dev* *11*, 3279-3285.
- Newman, A.J., and Nagai, K. (2010). Structural studies of the spliceosome: blind men and an elephant. *Curr Opin Struct Biol* *20*, 82-89.
- Niwa, M., and Berget, S.M. (1991). Polyadenylation precedes splicing in vitro. *Gene Expr* *1*, 5-14.
- Noble, C.G., Hollingworth, D., Martin, S.R., Ennis-Adeniran, V., Smerdon, S.J., Kelly, G., Taylor, I.A., and Ramos, A. (2005). Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat Struct Mol Biol* *12*, 144-151.
- Nunes, N.M., Li, W., Tian, B., and Furger, A. (2010). A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence. *EMBO J* *29*, 1523-1536.
- Ofer, N., Weisman-Shomer, P., Shklover, J., and Fry, M. (2009). The quadruplex r(CG_n) destabilizing cationic porphyrin TMPyP4 cooperates with hnRNPs to increase the translation efficiency of fragile X premutation mRNA. *Nucleic Acids Res* *37*, 2712-2722.
- Offer, H., Erez, N., Zurer, I., Tang, X., Milyavsky, M., Goldfinger, N., and Rotter, V. (2002). The onset of p53-dependent DNA repair or apoptosis is determined by the level of accumulated damaged DNA. *Carcinogenesis* *23*, 1025-1032.

REFERENCES

- Oganesian, L., and Bryan, T.M. (2007). Physiological relevance of telomeric G-quadruplex formation: a potential drug target. *Bioessays* 29, 155-165.
- Orkin, S.H., Cheng, T.C., Antonarakis, S.E., and Kazazian, H.H., Jr. (1985). Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. *EMBO J* 4, 453-456.
- Osheim, Y.N., Sikes, M.L., and Beyer, A.L. (2002). EM visualization of Pol II genes in *Drosophila*: most genes terminate without prior 3' end cleavage of nascent transcripts. *Chromosoma* 111, 1-12.
- Ou, T.M., Lu, Y.J., Zhang, C., Huang, Z.S., Wang, X.D., Tan, J.H., Chen, Y., Ma, D.L., Wong, K.Y., Tang, J.C., *et al.* (2007). Stabilization of G-quadruplex DNA and down-regulation of oncogene c-myc by quindoline derivatives. *J Med Chem* 50, 1465-1474.
- Palmero, E.I., Achatz, M.I., Ashton-Prolla, P., Olivier, M., and Hainaut, P. (2010). Tumor protein 53 mutations and inherited cancer: beyond Li-Fraumeni syndrome. *Curr Opin Oncol* 22, 64-69.
- Pandey, N.B., Chodchoy, N., Liu, T.J., and Marzluff, W.F. (1990). Introns in histone genes alter the distribution of 3' ends. *Nucleic Acids Res* 18, 3161-3170.
- Pangalis, G.A., Trangas, T., Papanastasiou, C.J., Roussou, P.A., and Tsiapalis, C.M. (1985). Poly(A)-polymerase activity in chronic lymphocytic leukemia of the B cell type. *Acta Haematol* 74, 31-34.
- Parkinson, G.N., Lee, M.P., and Neidle, S. (2002). Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* 417, 876-880.
- Pei, X.H., Nakanishi, Y., Takayama, K., Bai, F., and Hara, N. (1999). Benzo[a]pyrene activates the human p53 gene through induction of nuclear factor kappaB activity. *J Biol Chem* 274, 35240-35246.
- Pendurthi, U.R., Alok, D., and Rao, L.V. (1997). Binding of factor VIIa to tissue factor induces alterations in gene expression in human fibroblast cells: up-regulation of poly(A) polymerase. *Proc Natl Acad Sci U S A* 94, 12598-12603.
- Peterson, M.L., Bingham, G.L., and Cowan, C. (2006). Multiple features contribute to the use of the immunoglobulin M secretion-specific poly(A) signal but are not required for developmental regulation. *Mol Cell Biol* 26, 6762-6771.
- Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* 20, 2922-2936.
- Phillips, C., Jung, S., and Gunderson, S.I. (2001). Regulation of nuclear poly(A) addition controls the expression of immunoglobulin M secretory mRNA. *EMBO J* 20, 6443-6452.
- Phillips, C., Pachikara, N., and Gunderson, S.I. (2004). U1A inhibits cleavage at the immunoglobulin M heavy-chain secretory poly(A) site by binding between the two downstream GU-rich regions. *Mol Cell Biol* 24, 6162-6171.
- Potter, J.D., Slattery, M.L., Bostick, R.M., and Gapstur, S.M. (1993). Colon cancer: a review of the epidemiology. *Epidemiol Rev* 15, 499-545.
- Prescott, J., and Falck-Pedersen, E. (1994). Sequence elements upstream of the 3' cleavage site confer substrate strength to the adenovirus L1 and L3 polyadenylation sites. *Mol Cell Biol* 14, 4682-4693.
- Probst, M.R., Blum, M., Fasshauer, I., D'Orazio, D., Meyer, U.A., and Wild, D. (1992). The role of the human acetylation polymorphism in the metabolic activation of the food carcinogen 2-amino-3-methylimidazo[4,5-f]quinoline (IQ). *Carcinogenesis* 13, 1713-1717.
- Proudfoot, N. (2000). Connecting transcription to messenger RNA processing. *Trends Biochem Sci* 25, 290-293.

REFERENCES

- Proudfoot, N.J., and Brownlee, G.G. (1976). 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* 263, 211-214.
- Radford, H.E., Meijer, H.A., and de Moor, C.H. (2008). Translational control by cytoplasmic polyadenylation in *Xenopus* oocytes. *Biochim Biophys Acta* 1779, 217-229.
- Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res* 12, 1231-1245.
- Rauch, J., Moran-Jones, K., Albrecht, V., Schwarzl, T., Hunter, K., Gires, O., and Kolch, W. (2011). c-Myc Regulates RNA Splicing of the A-Raf Kinase and Its Activation of the ERK Pathway. *Cancer Res* 71, 4664-4674.
- Rauch, J., O'Neill, E., Mack, B., Matthias, C., Munz, M., Kolch, W., and Gires, O. (2010). Heterogeneous nuclear ribonucleoprotein H blocks MST2-mediated apoptosis in cancer cells by regulating A-Raf transcription. *Cancer Res* 70, 1679-1688.
- Riou, J.F., Gomez, D., Lemarteleur, T., and Trentesaux, C. (2003). [G-quadruplex DNA: myth or reality?]. *Bull Cancer* 90, 305-313.
- Rosonina, E., Ip, J.Y., Calarco, J.A., Bakowski, M.A., Emili, A., McCracken, S., Tucker, P., Ingles, C.J., and Blencowe, B.J. (2005). Role for PSF in mediating transcriptional activator-dependent stimulation of pre-mRNA processing in vivo. *Mol Cell Biol* 25, 6734-6746.
- Rozenblatt-Rosen, O., Nagaike, T., Francis, J.M., Kaneko, S., Glatt, K.A., Hughes, C.M., LaFramboise, T., Manley, J.L., and Meyerson, M. (2009). The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors. *Proc Natl Acad Sci U S A* 106, 755-760.
- Ruegsegger, U., Beyer, K., and Keller, W. (1996). Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. *J Biol Chem* 271, 6107-6113.
- Ruegsegger, U., Blank, D., and Keller, W. (1998). Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. *Mol Cell* 1, 243-253.
- Ruepp, M.D., Schweingruber, C., Kleinschmidt, N., and Schumperli, D. (2011). Interactions of CstF-64, CstF-77, and symplekin: implications on localisation and function. *Mol Biol Cell* 22, 91-104.
- Rund, D., Dowling, C., Najjar, K., Rachmilewitz, E.A., Kazazian, H.H., Jr., and Oppenheim, A. (1992). Two mutations in the beta-globin polyadenylation signal reveal extended transcripts and new RNA polyadenylation sites. *Proc Natl Acad Sci U S A* 89, 4324-4328.
- Ryan, K., Calvo, O., and Manley, J.L. (2004). Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *Rna* 10, 565-573.
- Ryan, K., Murthy, K.G., Kaneko, S., and Manley, J.L. (2002). Requirements of the RNA polymerase II C-terminal domain for reconstituting pre-mRNA 3' cleavage. *Mol Cell Biol* 22, 1684-1692.
- Rzuczek, S.G., Pilch, D.S., Liu, A., Liu, L., LaVoie, E.J., and Rice, J.E. (2010). Macrocyclic pyridyl polyoxazoles: selective RNA and DNA G-quadruplex ligands as antitumor agents. *J Med Chem* 53, 3632-3644.
- Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A., and Burge, C.B. (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* 320, 1643-1647.

REFERENCES

- Sandri-Goldin, R.M., and Mendoza, G.E. (1992). A herpesvirus regulatory protein appears to act post-transcriptionally by affecting mRNA processing. *Genes Dev* 6, 848-863.
- Sasaki, R., Minowada, J., Bollum, F.J., and Miura, Y. (1990). Polyadenylic acid polymerase activity in chronic myelogenous leukemia. *Leuk Res* 14, 273-278.
- Schek, N., Cooke, C., and Alwine, J.C. (1992). Definition of the upstream efficiency element of the simian virus 40 late polyadenylation signal by using in vitro analyses. *Mol Cell Biol* 12, 5386-5393.
- Schrijver, I., Lenzi, T.J., Jones, C.D., Lay, M.J., Druzin, M.L., and Zehnder, J.L. (2003). Prothrombin gene variants in non-Caucasians with fetal loss and intrauterine growth retardation. *J Mol Diagn* 5, 250-253.
- Scorilas, A. (2002). Polyadenylate polymerase (PAP) and 3' end pre-mRNA processing: function, assays, and association with disease. *Crit Rev Clin Lab Sci* 39, 193-224.
- Scorilas, A., Talieri, M., Ardavanis, A., Courtis, N., Dimitriadis, E., Yotis, J., Tsiapalis, C.M., and Trangas, T. (2000). Polyadenylate polymerase enzymatic activity in mammary tumor cytosols: A new independent prognostic marker in primary breast cancer. *Cancer Res* 60, 5427-5433.
- Scott, J.M., and Imperiale, M.J. (1996). Reciprocal effects of splicing and polyadenylation on human immunodeficiency virus type 1 pre-mRNA processing. *Virology* 224, 498-509.
- Scully, R., Chen, J., Ochs, R.L., Keegan, K., Hoekstra, M., Feunteun, J., and Livingston, D.M. (1997). Dynamic changes of BRCA1 subnuclear location and phosphorylation state are initiated by DNA damage. *Cell* 90, 425-435.
- Seegers, D., Zwiers, A., Strober, W., Pena, A.S., and Bouma, G. (2002). A TaqI polymorphism in the 3'UTR of the IL-12 p40 gene correlates with increased IL-12 secretion. *Genes Immun* 3, 419-423.
- Sen, D., and Gilbert, W. (1990). A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* 344, 410-414.
- Shatkin, A.J. (1976). Capping of eucaryotic mRNAs. *Cell* 9, 645-653.
- Shaw, G., and Kamen, R. (1986). A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* 46, 659-667.
- Sheets, M.D., Ogg, S.C., and Wickens, M.P. (1990). Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res* 18, 5799-5805.
- Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* 33, 365-376.
- Shin-ya, K., Wierzba, K., Matsuo, K., Ohtani, T., Yamada, Y., Furihata, K., Hayakawa, Y., and Seto, H. (2001). Telomestatin, a novel telomerase inhibitor from *Streptomyces anulatus*. *J Am Chem Soc* 123, 1262-1263.
- Shin, H.D., Kim, L.H., Park, B.L., Jung, H.S., Cho, Y.M., Moon, M.K., Lee, H.K., and Park, K.S. (2003). Polymorphisms in fatty acid-binding protein-3 (FABP3) - putative association with type 2 diabetes mellitus. *Hum Mutat* 22, 180.
- Shin, J.H., Janer, M., McNeney, B., Blay, S., Deutsch, K., Sanjeevi, C.B., Kockum, I., Lernmark, A., Graham, J., Arnqvist, H., *et al.* (2007). IA-2 autoantibodies in incident type I diabetes patients are associated with a polyadenylation signal polymorphism in GIMAP5. *Genes Immun* 8, 503-512.

REFERENCES

- Simonsen, C.C., and Levinson, A.D. (1983). Analysis of processing and polyadenylation signals of the hepatitis B virus surface antigen gene by using simian virus 40-hepatitis B virus chimeric plasmids. *Mol Cell Biol* 3, 2250-2258.
- Singh, P., Alley, T.L., Wright, S.M., Kamdar, S., Schott, W., Wilpan, R.Y., Mills, K.D., and Graber, J.H. (2009). Global changes in processing of mRNA 3' untranslated regions characterize clinically distinct cancer subtypes. *Cancer Res* 69, 9422-9430.
- Sittler, A., Gallinaro, H., and Jacob, M. (1994). Upstream and downstream cis-acting elements for cleavage at the L4 polyadenylation site of adenovirus-2. *Nucleic Acids Res* 22, 222-231.
- Slee, E.A., O'Connor, D.J., and Lu, X. (2004). To die or not to die: how does p53 decide? *Oncogene* 23, 2809-2818.
- Soller, M., and White, K. (2003). ELAV inhibits 3'-end processing to promote neural splicing of ewg pre-mRNA. *Genes Dev* 17, 2526-2538.
- Sotiriou, S., Gibney, G., Baxevanis, A.D., and Nussbaum, R.L. (2009). A single nucleotide polymorphism in the 3'UTR of the SNCA gene encoding alpha-synuclein is a new potential susceptibility locus for Parkinson disease. *Neurosci Lett* 461, 196-201.
- Soussi, T., Ishioka, C., Claustres, M., and Beroud, C. (2006). Locus-specific mutation databases: pitfalls and good practice based on the p53 experience. *Nat Rev Cancer* 6, 83-90.
- Stetler, D.A., Reichlin, M., Berlin, C.M., and Jacob, S.T. (1987). Antibodies against nuclear poly(A) polymerases in rheumatic autoimmune diseases. *J Clin Immunol* 7, 24-28.
- Stetler, D.A., Rose, K.M., and Jacob, S.T. (1981). Anti-poly(A) polymerase antibodies in sera of tumor-bearing rats and human cancer patients. *Proc Natl Acad Sci U S A* 78, 7732-7736.
- Stutz, F., and Rosbash, M. (1998). Nuclear RNA export. *Genes Dev* 12, 3303-3319.
- Subramanian, M., Rage, F., Tabet, R., Flatter, E., Mandel, J.L., and Moine, H. (2011). G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Rep* 12, 697-704.
- Sun, D., Liu, W.J., Guo, K., Rusche, J.J., Ebbinghaus, S., Gokhale, V., and Hurley, L.H. (2008). The proximal promoter region of the human vascular endothelial growth factor gene has a G-quadruplex structure that can be targeted by G-quadruplex-interactive agents. *Mol Cancer Ther* 7, 880-889.
- Sun, X., Shimizu, H., and Yamamoto, K. (1995). Identification of a novel p53 promoter element involved in genotoxic stress-inducible p53 gene expression. *Mol Cell Biol* 15, 4489-4496.
- Sundquist, W.I., and Klug, A. (1989). Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* 342, 825-829.
- Szankasi, P., and Smith, G.R. (1995). A role for exonuclease I from *S. pombe* in mutation avoidance and mismatch correction. *Science* 267, 1166-1169.
- Szymczynska, B.R., Bowman, J., McCracken, S., Pineda-Lucena, A., Lu, Y., Cox, B., Lambermon, M., Graveley, B.R., Arrowsmith, C.H., and Blencowe, B.J. (2003). Structure and function of the PWI motif: a novel nucleic acid-binding domain that facilitates pre-mRNA processing. *Genes Dev* 17, 461-475.
- Takagaki, Y., and Manley, J.L. (1997). RNA recognition by the human polyadenylation factor CstF. *Mol Cell Biol* 17, 3907-3914.

REFERENCES

- Takagaki, Y., and Manley, J.L. (1998). Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. *Mol Cell* 2, 761-771.
- Takagaki, Y., and Manley, J.L. (2000). Complex protein interactions within the human polyadenylation machinery identify a novel component. *Mol Cell Biol* 20, 1515-1525.
- Takagaki, Y., Seipelt, R.L., Peterson, M.L., and Manley, J.L. (1996). The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* 87, 941-952.
- Takagi, M., Absalon, M.J., McLure, K.G., and Kastan, M.B. (2005). Regulation of p53 translation and induction after DNA damage by ribosomal protein L26 and nucleolin. *Cell* 123, 49-63.
- Teixeira, A., Tahiri-Alaoui, A., West, S., Thomas, B., Ramadass, A., Martianov, I., Dye, M., James, W., Proudfoot, N.J., and Akoulitchev, A. (2004). Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. *Nature* 432, 526-530.
- Thomas, J.E., Smith, M., Tonkinson, J.L., Rubinfeld, B., and Polakis, P. (1997). Induction of phosphorylation on BRCA1 during the cell cycle and after DNA damage. *Cell Growth Differ* 8, 801-809.
- Tian, B., Hu, J., Zhang, H., and Lutz, C.S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res* 33, 201-212.
- Tibbetts, R.S., Brumbaugh, K.M., Williams, J.M., Sarkaria, J.N., Cliby, W.A., Shieh, S.Y., Taya, Y., Prives, C., and Abraham, R.T. (1999). A role for ATR in the DNA damage-induced phosphorylation of p53. *Genes Dev* 13, 152-157.
- Tran, D.P., Kim, S.J., Park, N.J., Jew, T.M., and Martinson, H.G. (2001a). Mechanism of poly(A) signal transduction to RNA polymerase II in vitro. *Mol Cell Biol* 21, 7495-7508.
- Tran, P.T., Simon, J.A., and Liskay, R.M. (2001b). Interactions of Exo1p with components of MutLalpha in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 98, 9760-9765.
- Tran, Q., Coleman, T.P., and Roesser, J.R. (2003). Human transformer 2beta and SRp55 interact with a calcitonin-specific splice enhancer. *Biochim Biophys Acta* 1625, 141-152.
- Tran, Q., and Roesser, J.R. (2003). SRp55 is a regulator of calcitonin/CGRP alternative RNA splicing. *Biochemistry* 42, 951-957.
- Trangas, T., Courtis, N., Gounaris, A., Perez, S., Kokkinopoulos, D., Pangalis, G.A., and Tsiapalis, C.M. (1989). Patterns of adenosine deaminase, ecto-5'-nucleotidase, poly(A)polymerase and surface light chain expression in chronic lymphocytic leukemias. *Blut* 58, 187-193.
- Uitte de Willige, S., de Visser, M.C., Houwing-Duistermaat, J.J., Rosendaal, F.R., Vos, H.L., and Bertina, R.M. (2005). Genetic variation in the fibrinogen gamma gene increases the risk for deep venous thrombosis by reducing plasma fibrinogen gamma' levels. *Blood* 106, 4176-4183.
- Uitte de Willige, S., Rietveld, I.M., De Visser, M.C., Vos, H.L., and Bertina, R.M. (2007). Polymorphism 10034C>T is located in a region regulating polyadenylation of FGG transcripts and influences the fibrinogen gamma'/gammaA mRNA ratio. *J Thromb Haemost* 5, 1243-1249.
- Vagner, S., Ruegsegger, U., Gunderson, S.I., Keller, W., and Mattaj, I.W. (2000a). Position-dependent inhibition of the cleavage step of pre-mRNA 3'-end processing by U1 snRNP. *RNA* 6, 178-188.

REFERENCES

- Vagner, S., Vagner, C., and Mattaj, I.W. (2000b). The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes Dev* 14, 403-413.
- Valente, S.T., Gilmartin, G.M., Venkatarama, K., Arriagada, G., and Goff, S.P. (2009). HIV-1 mRNA 3' end processing is distinctively regulated by eIF3f, CDK11, and splice factor 9G8. *Mol Cell* 36, 279-289.
- Valsamakis, A., Zeichner, S., Carswell, S., and Alwine, J.C. (1991). The human immunodeficiency virus type 1 polyadenylation signal: a 3' long terminal repeat element upstream of the AAUAAA necessary for efficient polyadenylation. *Proc Natl Acad Sci U S A* 88, 2108-2112.
- Van Dusen, C.M., Yee, L., McNally, L.M., and McNally, M.T. (2010). A glycine-rich domain of hnRNP H/F promotes nucleocytoplasmic shuttling and nuclear import through an interaction with transportin 1. *Mol Cell Biol* 30, 2552-2562.
- van Solinge, W.W., Lind, B., van Wijk, R., Hart, H.C., and Kraaijenhagen, R.J. (1996). Clinical expression of a rare beta-globin gene mutation co-inherited with haemoglobin E-disease. *Eur J Clin Chem Clin Biochem* 34, 949-954.
- Varani, L., Gunderson, S.I., Mattaj, I.W., Kay, L.E., Neuhaus, D., and Varani, G. (2000). The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. *Nat Struct Biol* 7, 329-335.
- Venkataraman, K., Brown, K.M., and Gilmartin, G.M. (2005). Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev* 19, 1315-1327.
- Veraldi, K.L., Arhin, G.K., Martincic, K., Chung-Ganster, L.H., Wilusz, J., and Milcarek, C. (2001). hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells. *Mol Cell Biol* 21, 1228-1238.
- Vethantham, V., Rao, N., and Manley, J.L. (2008). Sumoylation regulates multiple aspects of mammalian poly(A) polymerase function. *Genes Dev* 22, 499-511.
- Villalba, A., Coll, O., and Gebauer, F. (2011). Cytoplasmic polyadenylation and translational control. *Curr Opin Genet Dev*.
- Vogelstein, B., Lane, D., and Levine, A.J. (2000). Surfing the p53 network. *Nature* 408, 307-310.
- Wagner, E.J., Burch, B.D., Godfrey, A.C., Salzler, H.R., Duronio, R.J., and Marzluff, W.F. (2007). A genome-wide RNA interference screen reveals that variant histones are necessary for replication-dependent histone pre-mRNA processing. *Mol Cell* 28, 692-699.
- Wagner, E.J., and Marzluff, W.F. (2006). ZFP100, a component of the active U7 snRNP limiting for histone pre-mRNA processing, is required for entry into S phase. *Mol Cell Biol* 26, 6702-6712.
- Wahle, E. (1991). Purification and characterization of a mammalian polyadenylate polymerase involved in the 3' end processing of messenger RNA precursors. *J Biol Chem* 266, 3131-3139.
- Wahle, E. (1995). Poly(A) tail length control is caused by termination of processive synthesis. *J Biol Chem* 270, 2800-2808.
- Wahle, E., and Kuhn, U. (1997). The mechanism of 3' cleavage and polyadenylation of eukaryotic pre-mRNA. *Prog Nucleic Acid Res Mol Biol* 57, 41-71.
- Wakiyama, M., Imataka, H., and Sonenberg, N. (2000). Interaction of eIF4G with poly(A)-binding protein stimulates translation and is critical for Xenopus oocyte maturation. *Curr Biol* 10, 1147-1150.

REFERENCES

- Wakiyama, M., Takimoto, K., Ohara, O., and Yokoyama, S. (2007). Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system. *Genes Dev* 21, 1857-1862.
- Walther, T.N., Wittop Koning, T.H., Schumperli, D., and Muller, B. (1998). A 5'-3' exonuclease activity involved in forming the 3' products of histone pre-mRNA processing in vitro. *RNA* 4, 1034-1046.
- Wang-Gohrke, S., Becher, H., Kreienberg, R., Runnebaum, I.B., and Chang-Claude, J. (2002). Intron 3 16 bp duplication polymorphism of p53 is associated with an increased risk for breast cancer by the age of 50 years. *Pharmacogenetics* 12, 269-272.
- Wang, Y., and Patel, D.J. (1993). Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure* 1, 263-282.
- Wassarman, K.M., and Steitz, J.A. (1993). Association with terminal exons in pre-mRNAs: a new role for the U1 snRNP? *Genes Dev* 7, 647-659.
- West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* 432, 522-525.
- Wickens, M., Anderson, P., and Jackson, R.J. (1997). Life and death in the cytoplasm: messages from the 3' end. *Curr Opin Genet Dev* 7, 220-232.
- Wiestner, A., Tehrani, M., Chiorazzi, M., Wright, G., Gibellini, F., Nakayama, K., Liu, H., Rosenwald, A., Muller-Hermelink, H.K., Ott, G., *et al.* (2007). Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival. *Blood* 109, 4599-4606.
- Wilusz, C.J., Wormington, M., and Peltz, S.W. (2001). The cap-to-tail guide to mRNA turnover. *Nat Rev Mol Cell Biol* 2, 237-246.
- Wilusz, J.E., and Beemon, K.L. (2006). The negative regulator of splicing element of Rous sarcoma virus promotes polyadenylation. *J Virol* 80, 9634-9640.
- Xing, H., Mayhew, C.N., Cullen, K.E., Park-Sarge, O.K., and Sarge, K.D. (2004). HSF1 modulation of Hsp70 mRNA polyadenylation via interaction with symplekin. *J Biol Chem* 279, 10551-10555.
- Yan, J., and Marr, T.G. (2005). Computational analysis of 3'-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat. *Genome Res* 15, 369-375.
- Yang, D., and Okamoto, K. (2010). Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Med Chem* 2, 619-646.
- Yang, D.Q., Halaby, M.J., and Zhang, Y. (2006). The identification of an internal ribosomal entry site in the 5'-untranslated region of p53 mRNA provides a novel mechanism for the regulation of its translation following DNA damage. *Oncogene* 25, 4613-4619.
- Yang, X.C., Burch, B.D., Yan, Y., Marzluff, W.F., and Dominski, Z. (2009). FLASH, a proapoptotic protein involved in activation of caspase-8, is essential for 3' end processing of histone pre-mRNAs. *Mol Cell* 36, 267-278.
- Yang, Y., Li, C.C., and Weissman, A.M. (2004). Regulating the p53 system through ubiquitination. *Oncogene* 23, 2096-2106.
- Yang, Z., and Kaye, D.M. (2009). Mechanistic insights into the link between a polymorphism of the 3'UTR of the SLC7A1 gene and hypertension. *Hum Mutat* 30, 328-333.

REFERENCES

- Yasuda, M., Shabbeer, J., Osawa, M., and Desnick, R.J. (2003). Fabry disease: novel alpha-galactosidase A 3'-terminal mutations result in multiple transcripts due to aberrant 3'-end formation. *Am J Hum Genet* 73, 162-173.
- Yonaha, M., and Proudfoot, N.J. (1999). Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. *Mol Cell* 3, 593-600.
- Yonaha, M., and Proudfoot, N.J. (2000). Transcriptional termination and coupled polyadenylation in vitro. *EMBO J* 19, 3770-3777.
- Yoshida, T., Makino, Y., and Tamura, T. (1999). Association of the rat heterogeneous nuclear RNA-ribonucleoprotein F with TATA-binding protein. *FEBS Lett* 457, 251-254.
- Zahler, A.M., Williamson, J.R., Cech, T.R., and Prescott, D.M. (1991). Inhibition of telomerase by G-quartet DNA structures. *Nature* 350, 718-720.
- Zarudnaya, M.I., Kolomiets, I.M., Potyahaylo, A.L., and Hovorun, D.M. (2003). Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res* 31, 1375-1386.
- Zhang, X., Virtanen, A., and Kleiman, F.E. (2010). To polyadenylate or to deadenylate: that is the question. *Cell Cycle* 9, 4437-4449.
- Zhang, Z., Fu, J., and Gilmour, D.S. (2005). CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes Dev* 19, 1572-1580.
- Zhang, Z., Shi, Q., Sturgis, E.M., Spitz, M.R., Hong, W.K., and Wei, Q. (2004). Thymidylate synthase 5'- and 3'-untranslated region polymorphisms associated with risk and progression of squamous cell carcinoma of the head and neck. *Clin Cancer Res* 10, 7903-7910.
- Zhao, J., Hyman, L., and Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev* 63, 405-445.
- Zhao, W., and Manley, J.L. (1996). Complex alternative RNA processing generates an unexpected diversity of poly(A) polymerase isoforms. *Mol Cell Biol* 16, 2378-2386.
- Zhao, W., and Manley, J.L. (1998). Deregulation of poly(A) polymerase interferes with cell growth. *Mol Cell Biol* 18, 5010-5020.
- Zheng, S.L., Augustsson-Balter, K., Chang, B., Hedelin, M., Li, L., Adami, H.O., Bensen, J., Li, G., Johnsson, J.E., Turner, A.R., *et al.* (2004). Sequence variants of toll-like receptor 4 are associated with prostate cancer risk: results from the CAncer Prostate in Sweden Study. *Cancer Res* 64, 2918-2922.
- Zhou, H.L., Baraniak, A.P., and Lou, H. (2007). Role for Fox-1/Fox-2 in mediating the neuronal pathway of calcitonin/calcitonin gene-related peptide alternative RNA processing. *Mol Cell Biol* 27, 830-841.
- Zhou, H.L., and Lou, H. (2008). Repression of prespliceosome complex formation at two distinct steps by Fox-1/Fox-2 proteins. *Mol Cell Biol* 28, 5507-5516.
- Zhu, H., Hasman, R.A., Barron, V.A., Luo, G., and Lou, H. (2006). A nuclear function of Hu proteins as neuron-specific alternative RNA processing regulators. *Mol Biol Cell* 17, 5105-5114.
- Zhu, H., Hasman, R.A., Young, K.M., Kedersha, N.L., and Lou, H. (2003). U1 snRNP-dependent function of TIAR in the regulation of alternative RNA processing of the human calcitonin/CGRP pre-mRNA. *Mol Cell Biol* 23, 5959-5971.

REFERENCES

- Zhu, H., Zhou, H.L., Hasman, R.A., and Lou, H. (2007). Hu proteins regulate polyadenylation by blocking sites containing U-rich sequences. *J Biol Chem* 282, 2203-2210.
- Zhu, Z.H., Yu, Y.P., Shi, Y.K., Nelson, J.B., and Luo, J.H. (2009). CSR1 induces cell death through inactivation of CPSF3. *Oncogene* 28, 41-51.

ANNEXES

A physical and functional link between splicing factors promotes pre-mRNA 3' end processing

Stefania Millevoi^{1,2,3,*}, Adrien Decorsière^{1,2,3}, Clarisse Loulergue^{1,2,3}, Jason Iacovoni^{2,4}, Sandra Bernat^{1,2,3}, Michael Antoniou⁵ and Stéphan Vagner^{1,2,3,*}

¹INSERM, U563, Toulouse, ²Université de Toulouse, UPS, Centre de Physiopathologie de Toulouse Purpan, Toulouse, F-31300, ³Institut Claudius Regaud, Toulouse, F-31052, ⁴INSERM, IFR31, Bioinformatic Plateau, 1, Avenue Jean Poulhes, BP 84225, 31432 Toulouse Cedex 4, France and ⁵Nuclear Biology Group, Department of Medical and Molecular Genetics, King's College London School of Medicine, 8th Floor Tower Wing, Guy's Hospital, London, SE1 9RT, UK

Received April 22, 2009; Revised and Accepted May 15, 2009

ABSTRACT

Polypyrimidine tract-binding protein (PTB) is a splicing regulator that also plays a positive role in pre-mRNA 3' end processing when bound upstream of the polyadenylation signal (pA signal). Here, we address the mechanism of PTB stimulatory function in mRNA 3' end formation. We identify PTB as the protein factor whose binding to the human β -globin (HBB) 3' UTR is abrogated by a 3' end processing-inactivating mutation. We show that PTB promotes both *in vitro* 3' end cleavage and polyadenylation and recruits directly the splicing factor hnRNP H to G-rich sequences associated with several pA signals. Increased binding of hnRNP H results in stimulation of polyadenylation through a direct interaction with poly(A) polymerase. Therefore, our results provide evidence of a concerted regulation of pA signal recognition by splicing factors bound to auxiliary polyadenylation sequence elements.

INTRODUCTION

The 3' ends of most RNA polymerase II transcripts are generated by a nuclear co-transcriptional process that involves a site-specific endonucleolytic cleavage event followed by the addition of 15–200 adenylate residues (1–4).

In mammals, cleavage and polyadenylation requires the co-transcriptional assembly of a large complex consisting of at least six multimeric factors onto a bipartite core signal (pA signal), composed of the highly conserved A AUAAA hexamer situated 10–30 nucleotides upstream of the cleavage site and a more degenerate U- or GU-rich downstream sequence element (DSE). The pre-mRNA

forms a stable complex through a cooperative interaction between AAUAAA hexamer-bound cleavage and polyadenylation specificity factor (CPSF) and DSE-bound cleavage stimulatory factor (CstF). Additional factors are also required such as cleavage factor I (CF I), cleavage factor II (CF II), poly(A) polymerase (PAP) and nuclear poly(A)-binding protein (PABPN1). These factors are both necessary and sufficient to reconstitute the 3' end processing reaction *in vitro* (5,6). However, optimal *in vivo* processing not only depends upon the concerted recognition of the hexamer and DSE by the core 3' end machinery but also utilizes additional auxiliary *cis*-acting sequence elements and *trans*-acting factors that have been reported to modulate processing efficiency.

Upstream sequence elements (USE) are U-rich regions identified in association with several cellular pA sites (7–10). A number of factors have been identified as USE-binding proteins, including splicing factors of the hnRNP and SR protein families and polyadenylation factors. Among the splicing factors, polypyrimidine tract-binding protein (PTB), a major hnRNP protein that plays multiple roles in mRNA metabolism, has been frequently found to associate with USEs. In particular, PTB has been shown to interact with the USE of complement C2 (C2), prothrombin (F2) and cyclooxygenase-2 (COX-2) pre-mRNAs where it enhances 3' end processing efficiency (7,10,11). Contrastingly, PTB binding to the DSEs of the human α -globin and β -globin (HBB) pre-mRNA pA signals decreases the efficiency of 3' end cleavage (12). A similar inhibitory function of DSE-bound PTB has been observed for the C2 pA signal, thus revealing that PTB may play a role in both activation and repression of polyadenylation (12). While the competition between PTB and CstF 64 for binding to the DSE can account for the general inhibitory function of PTB (12), the molecular

*To whom correspondence should be addressed. Tel: +33 5 61 42 42 42. Ext: 4431; Fax: +33 5 61 42 46 31; Email: stefania.millevoi@inserm.fr
Correspondence may also be addressed to Stéphan Vagner. Tel: +33 5 67 69 63 11; Fax: +33 5 61 42 46 31; Email: stephan.vagner@inserm.fr

The authors wish it to be known that, in their opinion, the second and third authors should be regarded as joint Second Authors.

© 2009 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

mechanism of 3' end processing stimulation by PTB bound to upstream enhancer elements remains unclear.

In contrast to USEs, auxiliary sequence elements located downstream of the pA site (AUX-DSEs) are less frequently characterized. The best described AUX-DSE is the G-rich sequence element (GRS) found downstream of the SV40 late (SVL) pA signal (13–16). The SVL GRS which serves as a binding site for the splicing and polyadenylation factor hnRNP H (13, 15), has been proposed to stimulate 3' end processing by stabilizing the binding of CstF to the DSE (15,17–19). Similarly, GRSs which exhibit a stimulatory function on 3' end cleavage through the binding of hnRNP H have been found associated with other pA signals (13,15,20,21).

The importance of auxiliary regulatory sequence elements for 3' end processing is highlighted by the increasing number of human diseases in which this process is deregulated by mutations in *cis*-regulatory sequences (4). The HBB 3' end processing reaction is highly susceptible to β -thalassemia causing mutations within both the core polyadenylation elements and the flanking regulatory sequences (4). We previously reported that β -thalassemia mutations within the last splice acceptor site of HBB pre-mRNA reduce the binding of the splicing factor U2AF65 and interfere with the efficiency of 3' end cleavage (22).

In this report, we have investigated the mechanism by which the naturally occurring β -thalassemia term +6 (C to G) mutation, located 6 bases after the stop codon of the HBB 3' UTR, reduces 3' end cleavage efficiency (23). We show that the mutation resides in an evolutionary conserved pyrimidine tract and that it impairs the binding of PTB to this element. In agreement with the enhancing function of PTB in 3' end processing, PTB tethered upstream of the HBB pA signal stimulates *in vitro* cleavage/polyadenylation reactions and enhances the RNA binding activity of hnRNP H. Increased binding of hnRNP H results in stimulation of polyadenylation through a direct interaction with PAP. Importantly, PTB enhances hnRNP H recruitment to other PTB-regulated pA signals, suggesting that the interaction between the two splicing factors plays a general role in pA signal recognition.

MATERIALS AND METHODS

RNA substrates

DNA templates for *in vitro* transcription were obtained by two rounds of PCR. The forward primers containing the r17 sequence upstream of the indicated gene-specific sequence and reverse primers used in the first PCR were the following: HBB pA signal, forward: 5'-AATTTCTAT TAAAGGTTCCCTT-3' and reverse: 5'-GTTTGAAGCTAG CTCTTCATTTCTTTATG-3'; C2 pA signal, forward: 5'-ATGGAATTTCCAGTTAT-3' and reverse: 5'-GCT CTTGGAGTCATTCTGGC-3'; F2 pA signal, forward: 5'-CTAAAATATGGTTCCCAAT-3' and reverse: 5'-T CCCACCTCAGCCTCCCGAG-3'.

In the second round of PCR, the three PCR products were amplified using a forward primer containing the T7 promoter sequence upstream of the r17 sequence and

reverse primers as in the first round of PCR. Capped, uniformly ^{32}P -labeled RNAs used for cleavage, polyadenylation and UV cross-linking assays were obtained by *in vitro* transcription of these PCR products. The L3 pA signal was obtained by *in vitro* transcription of the R17-L3 linearized plasmid as previously described (22).

Cleavage and polyadenylation reactions

In vitro cleavage reactions were performed by incubating ^{32}P -labeled RNA substrates with nuclear extracts (NE) for 90 min at 30°C in the presence of purified recombinant GST-R17 or GST-R17-PTB fusion proteins as previously described (24). The *in vitro* polyadenylation assays using NE were performed for 15 min at 30°C as for the cleavage reaction except that it was done without cordycepin and in the presence of 0.7 mM ATP and Mn^{2+} . hnRNP H/F depleted NE were performed by three consecutive rounds of incubation of NE with 1 μg of streptavidine/agarose-bound SVL GRS for 1 h at 4°C. Alternatively, sequestration of hnRNP H/F was performed by addition of 1 μg of SVL GRS (or a control RNA) in the cleavage assay. Reconstituted polyadenylation reactions were performed for 15 min at 37°C as described in ref. (24). Analysis and quantification of cleavage/polyadenylation reactions after RNA extraction and resolution on a denaturing 6% polyacrylamide gel was done by PhosphorImager (Molecular Dynamics) analysis. Cleavage activity was calculated by dividing the amount of upstream cleavage product by the sum of cleavage plus precursor products. Polyadenylation efficiency was calculated by dividing the pre-cleaved product by the polyadenylated product.

UV crosslinking/IP

Purified GST-tagged R17, R17-PTB or PTB proteins were incubated with ^{32}P -labeled transcripts corresponding to the HBB, C2, F2 or L3 pA signals under cleavage conditions for 30 min at room temperature. The reaction mixtures were then irradiated on ice with UV light (254 nm) in a Stratalinker (Stratagene) at 0.4 J/cm² at 10 cm distance. Then, 50 units of RNase ONE (Promega) was added and the reaction mixtures were incubated for 30 min at 37°C. SDS gel loading buffer was added and the samples were boiled for 2 min before fractionation on a 10% SDS-PAGE. For IP analysis, protein G sepharose was incubated with antibodies prior to addition of UV-crosslinked complexes in NETN buffer (20 mM Tris at pH 8.0, 100 mM NaCl, 0.5% NP-40, 0.5 mM EDTA). After 1 h at 4°C, bound proteins were washed six times with NETN buffer and eluted by adding SDS loading buffer to the beads. Antibodies used were hnRNP H/F (mAb 1G11, Abcam), PTB (mAb Bb7, Abcam), U2AF65 (mAb MC3; a gift of M. Carmo-Fonseca), CFIm (polyclonal; a gift of W. Keller), CstF 64 (mAb 3A1; a gift of C. McDonald).

Bioinformatical sequence motif search

In order to select genes with the PYR2 consensus motif [C/T]TTTC[C/T]TGCT in their 3' ends, we developed three custom bioinformatics programs that extract

genomic sequences, search these sequences for regular expressions and then partition the sequences based on their motif content/topology. The extraction used the UCSC human genome sequence (hg18) in conjunction with the transcript coordinates provided by the refFlat.txt SQL dump and the polyA prediction track file polyApredict.txt. Coordinates of transcripts and polyA sites were used to extract genomic sequences encompassing all 3' UTR exons and downstream genomic sequences. A set of regular expressions were generated to match the sequence motifs of interest and to search the above-mentioned 3' end sequence database. Finally, a program was developed to partition all sequences with matching motifs based on the order and proximity of the motifs in order to select the final set of genes that contained the desired motif topology.

Recombinant proteins

GST fusion proteins were cloned in the pGEX-2T *E. coli* expression vector. His-tagged hnRNP H and hnRNP F encoding plasmids were kindly provided by Douglas Black. All fusion proteins were expressed in *E. coli* at 20°C overnight and purified to homogeneity by glutathione agarose chromatography. The His-tagged bovine PAP (residues 1–694), hnRNP H and hnRNP F were purified to homogeneity by Ni²⁺-NTA chromatography as described in ref. (25). Recombinant purified PABP was kindly provided by Maria Carmo-Fonseca.

GST pull-down assays

GST pull-down assays were performed by incubating 1 µg of purified GST-R17 or GST-R17-PTB bound to 20 µl of glutathione agarose beads with 200 µg of HeLa NEs (Figure 3a) in NETN buffer (20 mM Tris at pH 8.0, 100 mM NaCl, 0.5% NP-40, 0.5 mM EDTA) for 60 min at 4°C. Beads were then washed five times, treated with 10 µg/ml RNase A at room temperature for 30 min and washed again. Protein elution was performed by adding SDS loading buffer to the beads. Eluted proteins were resolved by 10% SDS-polyacrylamide gel electrophoresis (PAGE) and analyzed by western blot. GST pull-down assays shown in Figures 3b and 6g and h were performed as described above except that it was done with 1 µg of His-tagged hnRNP H, hnRNP F or PAP, and bound proteins were visualized by Coomassie blue staining.

RESULTS

PTB binds to a pyrimidine tract upstream of the HBB pA signal

Previous data have shown that the term+6 mutation in the HBB 3' UTR causes β-thalassemia by interfering with correct 3' end formation (23). Analysis of the sequence surrounding the mutation revealed that it resides within a region containing a long series of pyrimidine residues, which represents a potential PYR tract (named PYR2, Figure 1a). This pyrimidine-rich sequence closely

resembles the HBB IVSII PYR tract (named PYR1; Figure 1a) at the IVSII 3' splicing site that binds U2AF65 and is involved in 3' end processing (22,24).

In order to determine the protein factor(s) that bind to PYR2 and that may be involved in the mechanism by which the term+6 mutation affects 3' end processing, we performed UV crosslinking and immunoprecipitation (IP) experiments. UV crosslinking using a ³²P-end labeled RNA oligonucleotide corresponding to PYR2 (PYR2 RNA) in the presence of HeLa cell NEs produced three major bands (Figure 1b). The term+6 mutation within PYR2 (term+6 RNA) abolished the binding of a protein of 55 kDa. This band was identified as PTB by IP of the UV-crosslinked complexes with the PTB antibody (Figure 1c). No IP bands were detectable using either the mutated RNA or a mock IP (Figure 1c), indicating that binding of PTB to PYR2 RNA was eliminated by the term+6 mutation.

Next, we performed UV crosslinking experiments using recombinant, purified PTB. Even the highest concentrations of PTB were only able to bind to the wild-type PYR2 RNA (Figure 1d). As U2AF65 has been found by us (22) and others (11) to regulate 3' end processing by binding to upstream pyrimidine-rich regulatory elements, we next determined whether this splicing factor was able to bind to PYR2 RNA. UV crosslinking/IP experiments with the U2AF65 antibody gave a strong immunoprecipitated complex with PYR1 RNA, while only a faint band was detectable with PYR2 RNA (Figure 1e). This suggests that U2AF65 more closely associates with the PYR1 compared to the PYR2 RNA.

Taken together, these results demonstrate that the HBB term+6 mutation impairs the binding of PTB to the PYR2 element and that PTB, but not U2AF65, may be the factor involved in the mechanism of 3' end processing deregulation induced by this genetic defect.

PTB bound upstream of the pA signal stimulates 3'-end cleavage

PTB has been shown to play an inhibitory function in the HBB 3' end processing by competing with the cleavage factor CstF 64 for the U-rich DSE (12). We therefore investigated the function of PTB bound upstream of the HBB pA signal by tethering a R17/MS2-PTB fusion protein to a ³²P-labeled HBB pre-mRNA containing a high-affinity R17 binding site (named r17) in place of PYR2 (Figure 2a) and using NEs. No cleavage product was observed with an RNA substrate containing a AAUAA A to AAGAAA mutation in the HBB pA signal (Figure 2b). Addition of R17-PTB, but not R17 alone, increased cleavage efficiency in a dose-dependent manner up to 2-fold (Figure 2b). Similar results were obtained using the r17/adenovirus L3 (L3) pA signal substrate (Figure 2c), suggesting a more general positive function of PTB in 3' end cleavage when bound upstream of a pA signal.

PTB interacts directly with hnRNP H

In order to dissect the molecular mechanism whereby PTB stimulates 3' end processing, we wished to identify

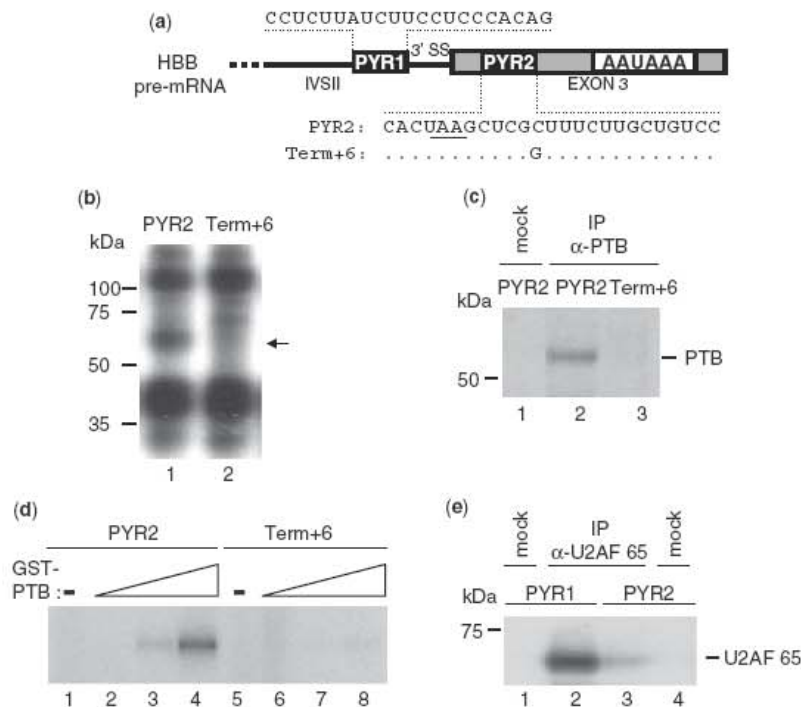


Figure 1. PTB binding to a pyrimidine tract within the HBB 3' UTR is impaired by the β -thalassemia term+6 mutation. (a) Schematic representation of the 3' terminal region of the HBB pre-mRNA showing the sequence of two pyrimidines tracts, PYR1 and PYR2, located at the last intron (IVSII) 3' splice site (3'SS) and in the terminal exon (exon 3), respectively. The sequence difference between the PYR2 and term+6, which contains the β -thalassemia C to G mutation at 6 nucleotides after the stop codon (underlined), is shown. (b) UV crosslinking using 5' 32 P-labeled RNA oligonucleotides corresponding to the PYR2 or term+6 RNAs in the presence of NEs and resolution by SDS-PAGE. The 55-kDa protein UV crosslinked to the PYR2 but not to the term+6 RNA is indicated by the arrow. (c) IP of the UV-crosslinked complexes described in (b) with the PTB antibody (Bb7) followed by SDS-PAGE analysis. (d) UV crosslinking of GST-tagged PTB (2.5, 1.25 or 0.6 pmol) to the PYR2 or term+6 RNAs. Lanes 1 and 5: UV crosslinking in the absence of PTB. (e) IP following UV crosslinking of NEs to 5' 32 P-labeled PYR1 or PYR2 with the U2AF65 antibody followed by SDS-PAGE analysis.

proteins with which PTB interacts. As PTB and hnRNP H form a cooperative assembly on the splicing downstream control sequence element of the c-src pre-mRNA (26), we tested the interaction between PTB and hnRNP H family regulators of 3' end processing, including hnRNP H/H' and hnRNP F. To this end, we performed a glutathione S-transferase (GST) pull-down experiment using GST-R17-PTB and NEs. Western blot analysis with an hnRNP H/F antibody detected both hnRNP proteins, suggesting an interaction between PTB and both hnRNP H and F (Figure 3a). No protein-protein interactions were found between PTB and two other 3' cleavage/polyadenylation factors, CstF 64 and CFI_m, which are often targeted by 3' end processing regulators (Figure 3a). GST pull-down assays using only recombinant proteins revealed that hnRNP H was able to bind to GST-R17-PTB but not to the control GST-R17 fusion protein; no binding was observed for hnRNP F (Figure 3b).

We therefore conclude that PTB binds directly to hnRNP H. The interaction between PTB and hnRNP F shown in Figure 3a is probably indirect and mediated by the ability of hnRNP H to heterodimerize with hnRNP F (27).

HnRNP H regulates 3' end formation of HBB pre-mRNA by binding to a G-rich AUX-DSE

The results shown in Figure 3 raise the possibility that the interaction between PTB and hnRNP H may be involved in the stimulatory effect of PTB on 3' end processing. If so, we would expect that hnRNP H is able to modulate HBB pre-mRNA 3' end processing efficiency by binding to a GRS flanking the core pA signal. Analysis of the HBB region downstream of the DSE revealed the presence of a putative GRS located 45 nucleotides downstream of the cleavage site (named DGRS, Figure 4a). To ascertain whether hnRNP H binds to this sequence element, we performed UV crosslinking assays using 32 P-labeled RNA substrates corresponding to the HBB 3' flanking region containing either wild-type (WT) or mutant (Mut) DGRS in the presence of NEs. A comparison of the UV crosslinking patterns of WT and Mut substrate RNAs revealed a specific band at 55 kDa that disappeared with the G-to-C/A mutation (Figure 4b). IP of the cross-linked complexes with the hnRNP H/F antibody showed that the 55-kDa band corresponds to hnRNP H/F and that the G-to-C/A mutation mostly abolished hnRNP H/F binding to the DGRS (Figure 4c).

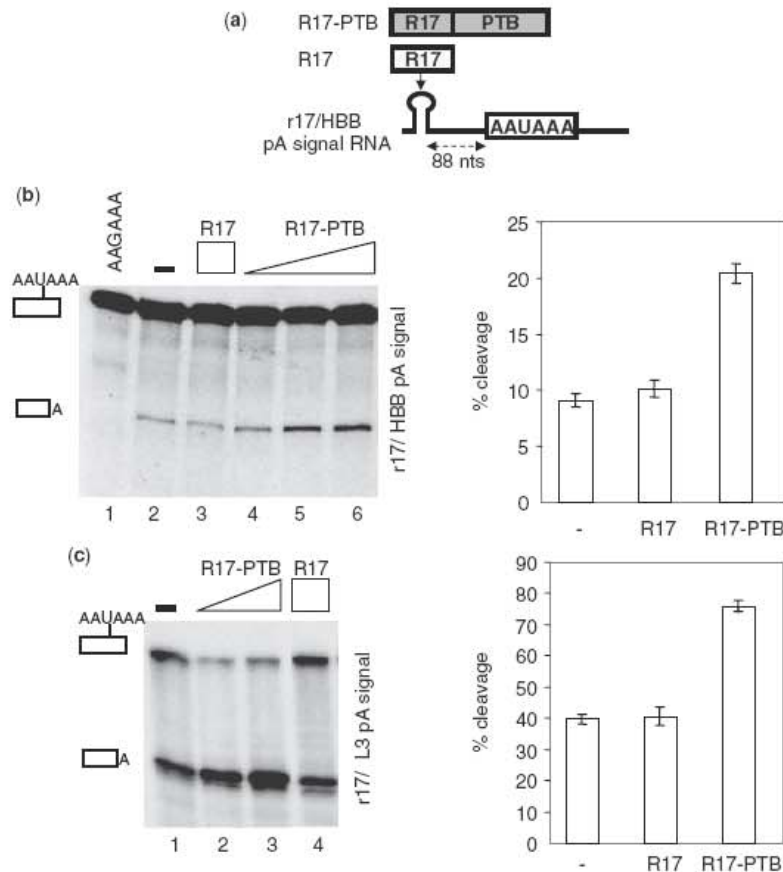


Figure 2. PTB stimulates *in vitro* 3' end cleavage in NEs. (a) Illustration of the 'tethering' assay consisting in binding R17-PTB or R17 fusion proteins to the high-affinity R17 binding site (r17) inserted in place of the PYR2 element located 88 nucleotides upstream of the HBB pA signal. (b) *In vitro* cleavage reactions using the r17/HBB pA signal substrate containing the wild-type (AAUAAA) or the mutated (AAGAAA) hexamer. 32 P-labeled transcript was incubated with NE in the presence of an increasing amount (3.5, 7 or 14 pmol) of R17-PTB or R17 (14 pmol). (c) Same as (b), with the 32 P-labeled RNA containing the R17 binding site upstream of the L3 pA signal and R17-PTB (7 or 14 pmol) or R17 (14 pmol) proteins. Identities of the uncleaved and cleaved products are shown on the left. The percentage of cleavage in the absence or in the presence of 14 pmol of R17-PTB or R17 proteins is shown on the right of (c) and (d), respectively.

We next investigated the function of hnRNP H in HBB 3' end cleavage by performing *in vitro* cleavage assays using NEs that had been either mock or depleted of endogenous hnRNP H/F and add-back experiments using recombinant hnRNP H or F. The endogenous hnRNP H/F was efficiently depleted by incubation of NEs with a biotinylated RNA containing the SVL GRS and streptavidin-agarose (Figure 4d). As shown in Figure 4e, depletion of hnRNP H/F reduced 3' end cleavage compared to mock-depleted extracts. Adding back recombinant hnRNP H restored 3' end cleavage in depleted extracts (Figure 4e), while raising the level of hnRNP H in mock extracts increased the formation of the 3' end cleavage product (Figure 4e). Unlike hnRNP H, add-back experiments with hnRNP F did not restore 3' end processing in depleted extracts nor did increasing the amount of hnRNP F reduce 3' end cleavage in mock NEs (Supplementary Figure S1). Similar results were obtained by performing 3' end cleavage assays in the presence of SVL GRS RNAs that sequester hnRNP H/F from the 3' end processing complex (Supplementary Figure S2).

Collectively, our results show that the 3' end processing of the HBB pA signal can be modulated by hnRNP H binding to the GRS downstream of the cleavage site.

PTB bound upstream of the pA signal recruits hnRNP H to the pre-mRNA

To investigate whether PTB was able to recruit hnRNP H to HBB transcripts, we analyzed hnRNP H RNA binding activity upon addition of R17-PTB using UV crosslinking/IP experiments. The 32 P-labeled r17/HBB pA signal RNA was incubated in NEs followed by the addition of the recombinant R17-PTB. The UV crosslinking pattern showed an increase in the binding of a 55-kDa species upon addition of R17-PTB (Figure 5a). This band was identified as hnRNP H/F by IP of the crosslinked proteins with the hnRNP H/F antibody (Figure 5b). Addition of R17-PTB (Figure 5b), but not of R17 (Figure 5b) or of PTB alone (Figure 5c), resulted in an enhanced hnRNP H crosslinking only in the presence of a WT DGRS (Figure 5d), suggesting that the physical interaction

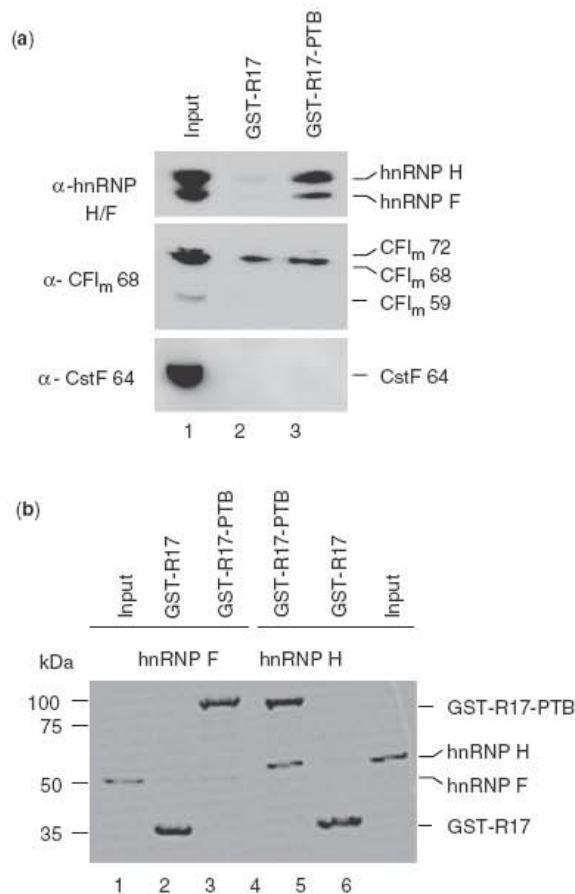


Figure 3. PTB interacts directly with hnRNP H, but not with hnRNP F. (a) *In vitro* GST pull-down assay using GST-tagged R17-PTB or R17 proteins and NEs, visualized by SDS-PAGE and followed by western blot analysis using antibodies against hnRNP H/F, CFI_m 68 or CstF 64. A 1/20th equivalent of the input NE is shown in lane 1. The experiments were carried out in the presence of RNase A, suggesting that the interaction was not mediated through RNA binding. (b) *In vitro* GST pull-down assay using GST-tagged R17-PTB or R17 proteins and recombinant His-tagged hnRNP F or hnRNP H, and this was analyzed by Coomassie staining of SDS-PAGE. The input lanes account for 10% of hnRNP H and hnRNP F used in the assay, respectively.

between PTB and hnRNP H creates a molecular link across the HBB pA signal.

In order to better characterize the effect of PTB on the RNA-binding activity of both hnRNP H and F, we performed UV crosslinking assays using only the r17/HBB pA signal substrate and recombinant proteins. Addition of R17-PTB strongly increased the binding of hnRNP H to the HBB substrate (Figure 5e). The effect of R17-PTB on hnRNP F binding to the RNA is less pronounced with hnRNP F alone but stronger in the presence of hnRNP H (Figure 5e). We therefore conclude that PTB bound upstream of the HBB pre-mRNA pA signal directly stimulates the binding of hnRNP H and indirectly helps recruit hnRNP F through heterodimerization with hnRNP H.

We next asked whether PTB was able to recruit hnRNP H on other pA signals that are modulated by PTB and possess a GRS downstream of the cleavage site. We performed UV crosslinking/IP assays using NEs and ³²P-labeled RNA substrates corresponding to the C2 and F2 pA signals containing the r17 moiety in place of the natural PTB-binding site (Figure 5f). Since the L3 pA signal can be regulated by PTB (Figure 2c), we also tested the r17/L3 pA substrate. As shown in Figure 5f (left panel), R17-PTB bound to sequences upstream of several pA signals stimulated hnRNP H binding to the RNA. This stimulatory effect was stronger when performing UV crosslinking of the r17/C2 or the r17/F2 substrates using only recombinant hnRNP H and R17-PTB (Figure 5f, right panel). Overall, these results suggest a more general stimulatory effect of PTB on the hnRNP H-RNA interaction.

PTB simulates polyadenylation by recruiting hnRNP H on a GRS located upstream of the pA signal

We next analyzed whether PTB was able to influence polyA addition through its recruitment of hnRNP H to GRSs upstream of the pA signal. This hypothesis is based on previous studies that suggest an implication of PTB or hnRNP H in polyadenylation (7,10,11). In addition, we identified a putative hnRNP H binding site between PYR2 and the pA signal of the HBB pre-mRNA (named UGRS, Figure 6a).

To verify this hypothesis, we performed *in vitro* polyadenylation reactions with NEs using a ³²P-labeled r17/HBB pre-cleaved substrate (Figure 6a). We first tested whether PTB functions as a regulator of the polyadenylation reaction. Addition of recombinant PTB stimulated addition of adenosine residues in a dose-dependent fashion (Figure 6b) to an extent similar to that obtained in the presence of the 3' end processing regulator U2AF65 (Figure 6b) (24).

To investigate the possible involvement of the PTB-hnRNP H interaction in the stimulatory effect of PTB on polyadenylation, we checked if PTB stimulated the interaction between hnRNP H and the pre-cleaved substrate using UV crosslinking/IP experiments with NEs. As expected, the UV crosslinking pattern of the pre-cleaved substrate was remarkably distinct in the 50-kDa region (Figure 6c), from that generated with the substrate including the HBB 3' flanking region (Figure 5a). However, the addition of R17-PTB still resulted in an increased crosslinked protein at this molecular size (Figure 6c). IP of the crosslinked complexes with the hnRNP H/F antibody confirmed that hnRNP H is the protein factor whose binding is potentiated by addition of R17 linked to PTB (Figure 6d).

If the increased binding of hnRNP H to the RNA is responsible for the stimulation of pA addition by PTB, then it should be possible to reproduce the stimulatory effect by increasing the amount of recombinant hnRNP H in a polyadenylation assay. To test this possibility, we performed reconstituted polyadenylation assays using only recombinant PAP, PABPN1, hnRNP H or F and R17-PTB. As shown in Figure 6e, the polyadenylation

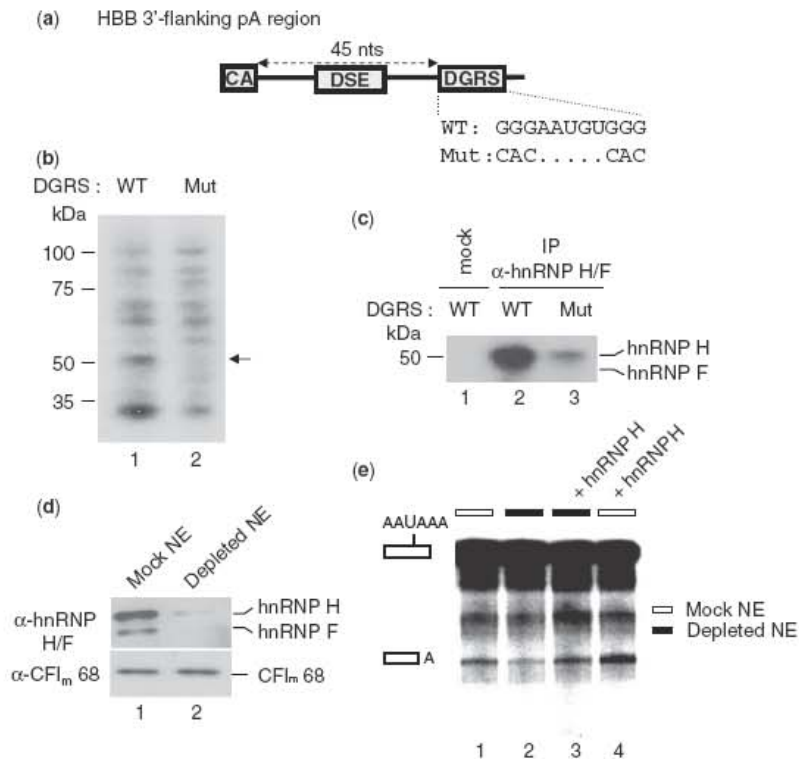


Figure 4. HnRNP H stimulates 3' end cleavage at the HBB pA signal by binding to a downstream GRS element. (a) Illustration of the HBB region downstream of the cleavage site (CA) and including a GRS located 45 nucleotides downstream of the cleavage site (DGRS). The sequence difference between the wild-type (WT) and mutated DGRS (Mut) is shown. (b) UV crosslinking of NE proteins to the uniformly ^{32}P -labeled r17/HBB pA signal RNA substrate containing the WT or Mut DGRS sequence. The 50-kDa protein crosslinked to the WT but not to the Mut RNA substrate is indicated by the arrow. (c) IP of UV-crosslinked complexes described in panel (b) with the hnRNP H/F antibody followed by SDS-PAGE analysis. (d) NEs depleted of endogenous hnRNP H/F without (mock) or with (depleted) a biotinylated RNA oligonucleotide corresponding to the SVL GRS. NEs were analyzed by SDS-PAGE followed by western blot analysis. (e) *In vitro* cleavage assays using the ^{32}P -labeled r17/HBB substrate in the presence of mock-depleted NEs (white boxes) or NEs depleted of hnRNP H/F (black boxes), supplemented or not with recombinant hnRNP H (1 pmol). The low cleavage efficiency is due to the depletion procedure. Identities of the uncleaved and cleaved products are shown on the left.

activity of PAP alone or in the presence of PABPN1 is increased by the addition of hnRNP H. Conversely, when the polyadenylation reaction occurred in the presence of hnRNP F, the polyadenylation activity of PAP was significantly reduced (Figure 6e). While additional experiments are needed to further investigate the inhibitory activity of hnRNP F on polyadenylation, our results confirm that hnRNP H plays a role in stimulating pA addition. We next tested whether this stimulatory effect can be further potentiated in the presence of R17-PTB. R17-PTB was able to modify PAP activity only slightly (Figure 6f) suggesting that the stimulation of pA addition by R17-PTB in NEs does not rely on a direct effect of PTB on PAP. Interestingly, when both R17-PTB and hnRNP H were included in the reconstituted polyadenylation assay, the stimulatory activity of hnRNP H on poly(A) tail elongation was further increased compared to hnRNP H alone (Figure 6f). A possible explanation for these results is that PTB recruits PAP indirectly by means of its direct interaction with hnRNP H. This hypothesis implies that hnRNP H and PAP interact directly. We tested this possibility by performing GST

pull-down assays using recombinant GST-PTB, hnRNP H and PAP and found that PTB was able to interact with hnRNP H but not with PAP alone (Figure 6g). Conversely, when both PAP and hnRNP H were present in the assay, PAP was pulled down by PTB (Figure 6g), suggesting that the binding of PTB to PAP is mediated by hnRNP H. To confirm this possibility using the same proteins as in the functional and pull-down assays, we immunoprecipitated hnRNP H and checked whether PAP was found in the IP fraction. As shown in Figure 6h, PAP was pulled down only in the presence of hnRNP H.

Taken together, these results show that PTB bound upstream of the HBB pA signal facilitates the association of hnRNP H with a GRS that in turn recruits PAP thereby stimulating pA addition.

The HBB PYR2 element is conserved and can be found in other pA signals

Our results highlight the importance of the PYR2 element for efficient HBB pre-mRNA 3' end processing. The term +6 β -thalassemia causing mutation within this

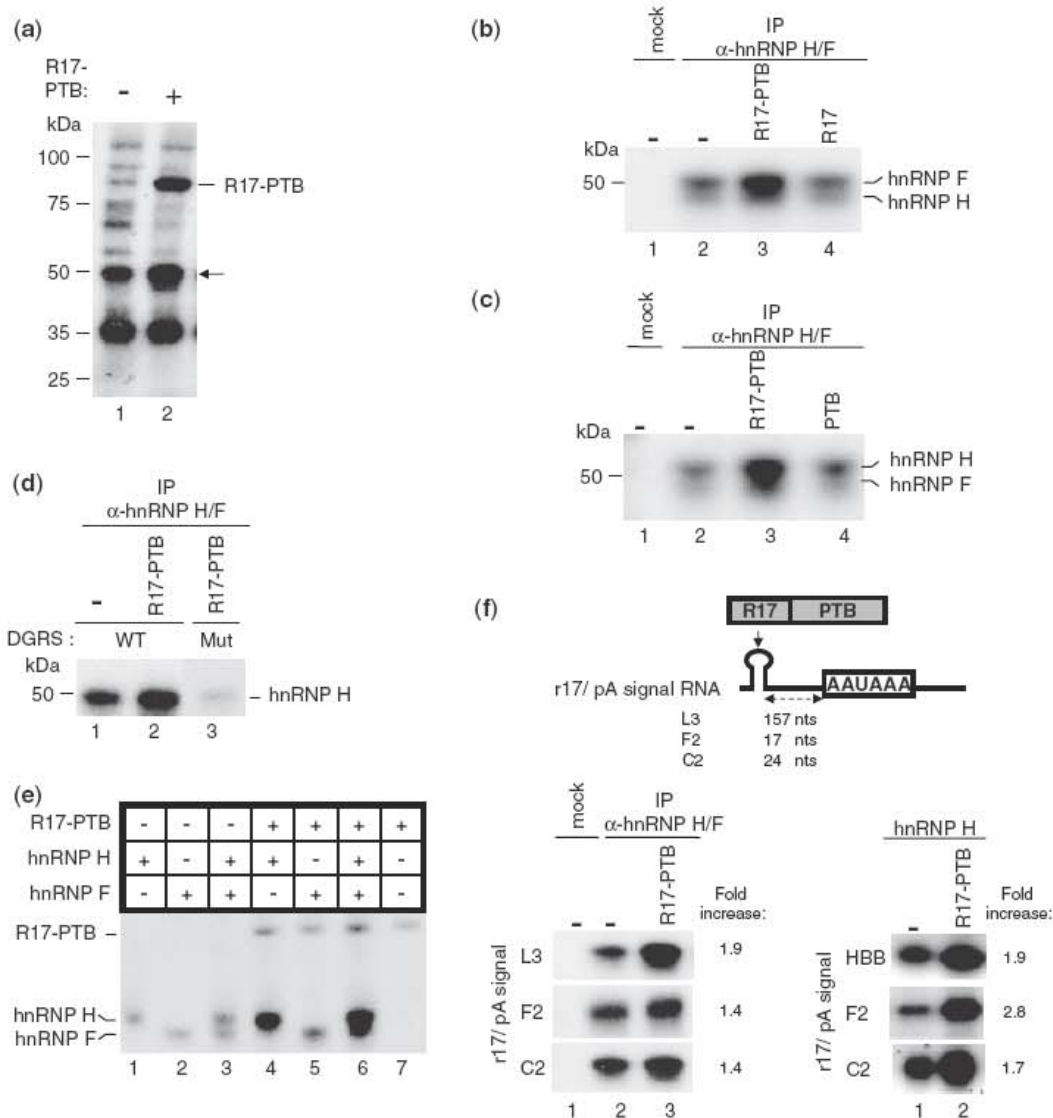


Figure 5. PTB facilitates hnRNP H binding to the HBB pA signal. (a) UV crosslinking using NEs with the ³²P-labeled r17/HBB substrate in the absence or presence of R17-PTB. The identity of the 50-kDa protein whose binding to the RNA is increased upon R17-PTB addition (indicated by an arrow) was tested in (b) by IP of the UV-crosslinked complexes with the antibody against hnRNP H/F. SDS-PAGE analysis of the immunoprecipitated complexes UV crosslinked to the ³²P-labeled r17/HBB substrate in the absence or presence of R17-PTB or R17. (c) UV crosslinking/IP as described in (b) except for lane 3 in which PTB replaced R17. The poor quality of the migration of the crosslinking reaction is due to the presence of polyvinylalcohol in the reaction. (d) UV crosslinking/IP as described in (b) but using a ³²P-labeled r17/HBB substrate containing the DGRS WT or Mut (as described in Figure 4). (e) UV crosslinking of recombinant hnRNP H and/or hnRNP F to the ³²P-labeled r17/HBB substrate in the absence or presence of the R17-PTB protein as indicated (upper panel). (f) (Left panel) UV crosslinking/IP of hnRNP H/F from NEs to ³²P-labeled RNA substrates containing the r17 moiety upstream of the L3, F2 or C2 pA signals (illustrated in the upper part) in the absence or presence of R17-PTB. (Right panel) UV crosslinking of hnRNP H with or without R17-PTB and using the ³²P-labeled r17/HBB, C2 or F2 pA signals.

element reduces both the binding of PTB and 3' end processing efficiency. A sequence alignment of the 3' ends of vertebrate *HBB* genes revealed that the nucleotide affected by the β -thalassemia mutation and the sequence surrounding this base are highly conserved among higher vertebrates (Figure 7). The sequence surrounding the mutation is the pyrimidine-rich region (9 pyrimidines out of 10 nucleotides of the most strongly conserved element)

with only two pyrimidine transitions. The upstream, conserved element is also positionally preserved. To gain insight into the possible involvement of this sequence in the control of 3' end formation of other mRNAs, we analyzed if other human mRNAs contain the sequence [C/T]TTTC[C/T]IGCT upstream of the pA signal. We identified 60 genes that contain this element in their pA signal region. Interestingly, 56 out of the 60 identified

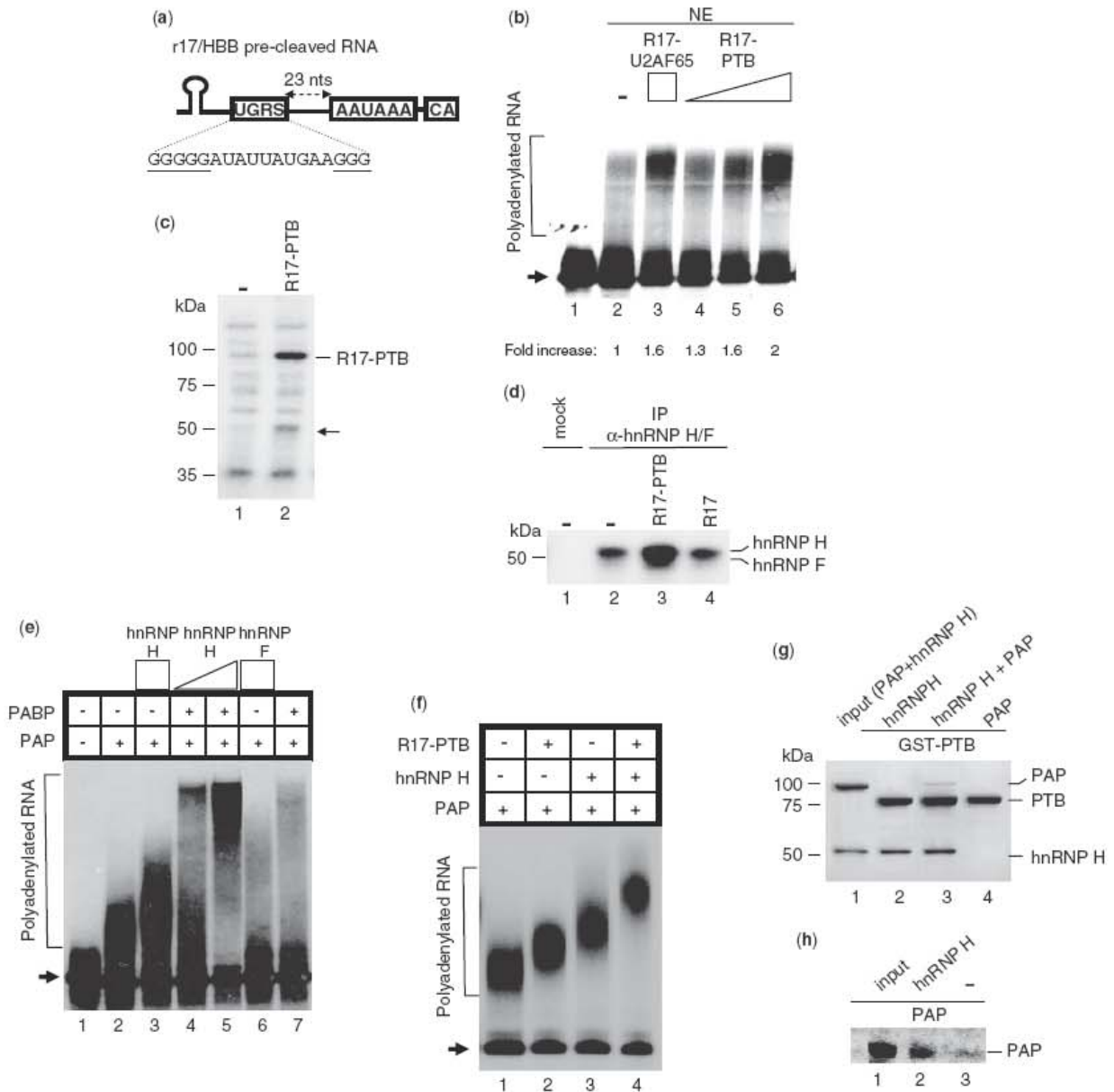


Figure 6. PTB promotes polyadenylation by increasing the binding of hnRNP H to a GRS located upstream of the HBB pA signal. (a) Illustration of the r17/HBB pre-cleaved substrate containing a GRS element upstream of the pA signal (named UGRS). (b) *In vitro* polyadenylation assays using the ³²P-labeled r17/HBB pre-cleaved substrate and NEs in the absence or presence of R17-U2AF65 (10 pmol) or increasing amounts of R17-PTB (2.5, 5 and 10 pmol). Normalized pA efficiency is shown. (c) UV crosslinking of NE proteins using the ³²P-labeled r17/HBB pre-cleaved substrate with (5 pmol) or without R17-PTB. The arrow indicates the 50-kDa protein whose binding to the HBB pre-cleaved RNA is increased upon R17-PTB addition. (d) IP of NE proteins UV crosslinked to the ³²P-labeled r17/HBB pre-cleaved substrate in the absence or presence of R17-PTB (5 pmol) or R17 (5 pmol) using the hnRNP H/F antibody. (e) Reconstituted polyadenylation assays using the ³²P-labeled r17/HBB pre-cleaved substrate, PAP (0.1 pmol) and PABPN1 (1.2 pmol), in the presence of increasing amounts of hnRNP H (2 and 4 pmol) or hnRNP F (4 pmol). Lane 1: input RNA; arrow: non-polyadenylated RNA substrate. (f) Reconstituted polyadenylation assays using the ³²P-labeled r17/HBB pre-cleaved substrate and PAP (0.1 pmol), in the presence of R17-PTB (4 pmol) and/or hnRNP H (4 pmol). Arrow: non-polyadenylated RNA substrate. (g) GST pull-down assay to test the interaction between GST-tagged PTB and PAP in the absence or presence of hnRNP H. The input lane accounts for 10% of hnRNP H and PAP used in the assay. (h) IP pull down assay with hnRNP H and PAP using the hnRNP H/F antibody. The input lane accounts for 10% of PAP used in the assay.

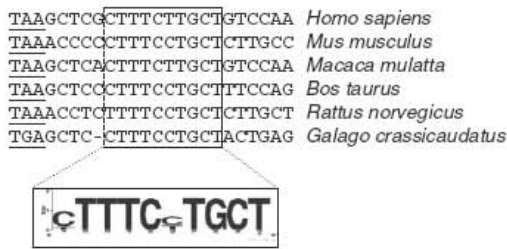


Figure 7. Eukaryotic conservation of the PTB-bound pyrimidine-rich element within the HBB 3' UTR as demonstrated by multiple sequence alignment.

genes contained at least one GGG downstream/upstream of the pA signal (Supplementary Table S1).

DISCUSSION

Bioinformatic analyses of the sequence surrounding the pA signal identified, in addition to the core elements, auxiliary over-represented *cis*-acting sequences, including U-rich upstream motifs and GRSs positioned both upstream and downstream of the pA signal (28). Both *cis*-elements have an enhancing function in 3' end formation through the binding of *trans*-acting proteins, including two splicing regulators, PTB and hnRNP H. PTB interacts with the U-rich USE and stimulates 3'-end cleavage by an unknown mechanism (7,10,11), whereas hnRNP H binds to the GRS located at the AUX-DSE and promotes 3' end formation by increasing the binding of CstF64 to the core pA signal (15,17–19). Our *in vitro* functional studies show that PTB and hnRNP H positively regulate both steps of HBB pre-mRNA 3' end processing by binding to an upstream U-rich element and two GRS sequences positioned on either side of the pA signal, respectively. More importantly, we demonstrate that PTB bound to the upstream element facilitates the interaction between hnRNP H and the two GRSs proximal to the HBB pA signal and that this physical interaction plays an enhancing function in both steps of 3' end formation. Thus, we provide evidence of a concerted regulation of pA signal recognition by splicing factors bound to auxiliary polyadenylation sequence elements.

On the basis of three main findings, we propose that this is a general mechanism of 3' end processing regulation. First, we have found that the HBB pyrimidine U-rich element is associated with 60 pA signals and that 92% of these pA signals also possess a GRS upstream and/or downstream of the core element (Supplementary Table S1). This result suggests that PTB bound to the pyrimidine U-rich element may function to modulate the 3' end processing efficiency of other pA signals and that hnRNP H may be involved in this stimulatory effect. Second, the sequence analysis of several PTB-regulated pA signals (including the C2, F2, proximal COX-2 and CT/CGRP exon 4 pA signals) has revealed that all of these pA signals contain a GRS in the proximity of the core pA element. Moreover, in the case of the C2 and F2 pA signals, we show that PTB stimulates hnRNP H

binding to RNA (Figure 5f). Finally, we show that the strong L3 pA signal can also be positively regulated by PTB bound upstream of the AAUAAA hexamer and that this stimulatory effect is associated with increased binding of hnRNP H, presumably to a G-rich motif present in the functional AUX-DSE (17; our observation).

One clue to the mechanism by which PTB stimulates 3' end processing via recruitment of hnRNP H comes from the observation that all of the pA signals considered in this study are associated with weak hnRNP H binding sites. Indeed, hnRNP H was identified as the protein factor that strongly binds to the SVL GRS. This element has the propensity to form G-quadruplex structures (29), and RNA secondary structures have been shown to play a role in polyadenylation (30,31). In addition, mutational analysis of the SVL GRS revealed that mutations disrupting the G-quadruplex structure not only affect the function of the GRS in the 3' end processing of the SVL late pA signal but also reduce the GRS-hnRNP H interaction (19). Both GRS elements of the HBB pA signal do not form G-quadruplex structures according to bioinformatic prediction (QGRS mapper, (32) and our experimental observations (A.D., unpublished data), indicating that they may represent weak binding sites for hnRNP H. However, previous studies have shown that hnRNP H is able to stimulate 3' end processing even when bound to weak binding sites and suggest that an accessory factor is needed to strengthen hnRNP H association and potentiate its stimulatory function (13). In agreement with this hypothesis and with the general view that auxiliary *cis*-acting motifs aid in the use of suboptimal polyadenylation sequence elements, we propose that PTB, bound to upstream auxiliary elements, exerts its stimulatory function by associating hnRNP H with suboptimal GRS elements and facilitating the assembly of the polyadenylation machinery (Figure 8).

The occurrence of a molecular and functional link between PTB and hnRNP H bound to their *cis*-regulatory elements may be critical for pA signals that inefficiently recruit the polyadenylation machinery due to U/GU-rich DSEs that are suboptimal for CstF binding, such as the DSEs from C2 and F2. For both pA signals, PTB bound to the USE stimulates both 3' end processing efficiency (7,11) and hnRNP H binding (Figure 5). Increased association of hnRNP H stimulates the association of a stable complex between CstF64 and the DSE (14,15,17–19). Unlike the F2 and C2 pA signals, the HBB pA signal consists of a consensus CPSF binding site and a U-rich DSE and is therefore very efficient in terms of 3' end processing. For this strong pA signal, the function of *cis*-auxiliary sequences is to further improve the cleavage/polyadenylation reaction, as shown for the pyrimidine tract near the last intron 3' splice site (24) or for the heterologous F2 USE inserted upstream of the HBB pA signal (11). The β -thalassemia term +6 mutation reduces 3' end efficiency by 22–30% (23), in agreement with the moderate ability of *cis*-regulatory sequences to modulate the efficiency of pA signals. As expected for a pA signal with a strong DSE, the PTB-induced increase in hnRNP H binding does not result in an enhanced binding of CstF 64 to the HBB cleavage substrate (data not shown).

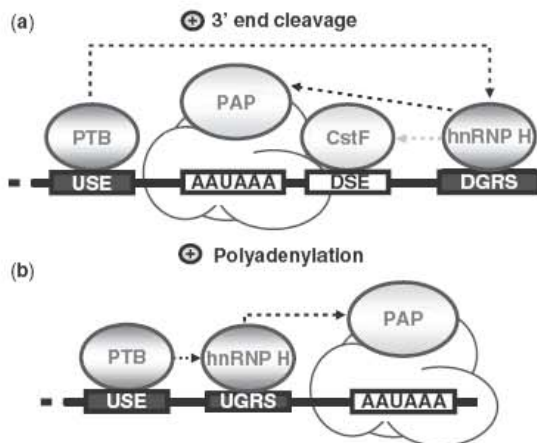


Figure 8. Illustration of mechanistic models for the stimulatory functions of PTB in 3' end processing. (a) PTB and 3' end cleavage. We propose that PTB facilitates the binding of hnRNP H to the DGRS located at the AUX-DSE, which in turn, stimulates the assembly of the 3' end cleavage machinery through the recruitment of CstF or PAP. (b) PTB and polyadenylation. We propose that PTB stabilizes the association of hnRNP H with a UGRS located upstream of the pA signal where it can increase the processivity of PAP. Black arrows indicate the functional interactions shown in this study while the grey arrow refers to the previously reported ability of hnRNP H to increase CstF binding to the DSE element (14,15,17–19).

Consequently, the stimulatory function of hnRNP H might involve other components of the polyadenylation machinery. Our results show that hnRNP H stimulates pA addition in reconstituted polyadenylation assays (Figure 6e and f) and that it interacts with PAP (Figure 6g and h), implying that a PAP–hnRNP H interaction plays a role in the stimulatory function of PTB. Since PAP is often involved in both 3' end cleavage and polyadenylation processes (33), a possible mechanism is that PTB stimulates the binding of hnRNP H to GRSs and that hnRNP H, in turn, enhances PAP activity in both steps of mRNA 3' end formation (Figure 8). According to this model and considering that the recruitment of PAP to pA signals is a late event in the assembly of the pA machinery, modulation of PAP association through PTB–hnRNP H concerted regulation can only moderately modify HBB cleavage efficiency. Since CstF can also enhance polyadenylation when bound to upstream *cis*-elements (7), it is possible that hnRNP H stimulates CstF binding to the upstream region and promotes pA addition. Although it will be important to study interactions between hnRNP H and other components of the polyadenylation machinery, we propose that hnRNP H primarily influences 3' end processing activity through PAP and CstF and that these interactions constitute the main determinant of PTB-mediated 3' end processing regulation.

We have previously shown that β -thalassemia causing mutations within the 3' splice site pyrimidine tract interfere with correct HBB pre-mRNA 3' end processing by reducing the ability of U2AF65 to recruit CF_Im to the pA signal (22, 24). In this study, we demonstrate that the

term+6 mutation in the HBB 3' UTR, which reduces 3' end processing, impairs the binding of PTB to a pyrimidine U-rich element. Our results show that a PTB-induced increase in binding of hnRNP H to GRS elements can account for the positive function of PTB in 3' end processing. Based on our results, we propose that disease-causing mutations not only disrupt the interactions between an mRNA and a trans-acting factor, but also disable the formation of synergistic complexes that regulate and perform 3' end processing reactions.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank D. Black for hnRNP H and F plasmids, M. Carmo-Fonseca for purified PABP and MC3 antibody, C. McDonald for CstF 64 antibody and W. Keller for CFI_m 68 antibody.

FUNDING

French National Institute for Health and Medical Research, the Institut Claudius Regaud, the Fondation de France; and by the Fondation pour la recherche médicale (Equipe FRM, soutenue par la Fondation Recherche Médicale) for work in the laboratory of S.V. Funding for open access charge: INSERM.

Conflict of interest statement. None declared.

REFERENCES

- Proudfoot,N.J., Furger,A. and Dye,M.J. (2002) Integrating mRNA processing with transcription. *Cell*, **108**, 501–512.
- Rosonina,E., Kaneko,S. and Manley,J.L. (2006) Terminating the transcript: breaking up is hard to do. *Genes Dev.*, **20**, 1050–1056.
- Mandel,C.R., Bai,Y. and Tong,L. (2008) Protein factors in pre-mRNA 3'-end processing. *Cell Mol. Life Sci.*, **65**, 1099–1122.
- Danckwardt,S., Hentze,M.W. and Kulozik,A.E. (2008) 3' end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J.*, **27**, 482–498.
- Takagaki,Y., Ryner,L.C. and Manley,J.L. (1989) Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes Dev.*, **3**, 1711–1724.
- Wahle,E. (1991) Purification and characterization of a mammalian polyadenylate polymerase involved in the 3' end processing of messenger RNA precursors. *J. Biol. Chem.*, **266**, 3131–3139.
- Moreira,A., Takagaki,Y., Brackenridge,S., Wollerton,M., Manley,J.L. and Proudfoot,N.J. (1998) The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. *Genes Dev.*, **12**, 2522–2534.
- Brackenridge,S. and Proudfoot,N.J. (2000) Recruitment of a basal polyadenylation factor by the upstream sequence element of the human lamin B2 polyadenylation signal. *Mol. Cell Biol.*, **20**, 2660–2669.
- Natalizio,B.J., Muniz,L.C., Arhin,G.K., Wilusz,J. and Lutz,C.S. (2002) Upstream elements present in the 3'-untranslated region of collagen genes influence the processing efficiency of overlapping polyadenylation signals. *J. Biol. Chem.*, **277**, 42733–42740.
- Hall-Pogar,T., Liang,S., Hague,L.K. and Lutz,C.S. (2007) Specific trans-acting proteins interact with auxiliary RNA polyadenylation elements in the COX-2 3'-UTR. *RNA*, **13**, 1103–1115.

12 *Nucleic Acids Research, 2009*

11. Danckwardt,S., Kaufmann,I., Gentzel,M., Foerstner,K.U., Gantzert,A.S., Gehring,N.H., Neu-Yilik,G., Bork,P., Keller,W., Wilm,M. *et al.* (2007) Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals. *EMBO J.*, **26**, 2658–2669.
12. Castelo-Branco,P., Furger,A., Wollerton,M., Smith,C., Moreira,A. and Proudfoot,N. (2004) Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol. Cell Biol.*, **24**, 4174–4183.
13. Arhin,G.K., Boots,M., Bagga,P.S., Milcarek,C. and Wilusz,J. (2002) Downstream sequence elements with different affinities for the hnRNP H/H' protein influence the processing efficiency of mammalian polyadenylation signals. *Nucleic Acids Res.*, **30**, 1842–1850.
14. Bagga,P.S., Ford,L.P., Chen,F. and Wilusz,J. (1995) The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3' end pre-mRNA processing through a trans-acting factor. *Nucleic Acids Res.*, **23**, 1625–1631.
15. Bagga,P.S., Arhin,G.K. and Wilusz,J. (1998) DSEF-1 is a member of the hnRNP H family of RNA-binding proteins and stimulates pre-mRNA cleavage and polyadenylation in vitro. *Nucleic Acids Res.*, **26**, 5343–5350.
16. Qian,Z.W. and Wilusz,J. (1991) An RNA-binding protein specifically interacts with a functionally important domain of the downstream element of the simian virus 40 late polyadenylation signal. *Mol. Cell Biol.*, **11**, 5312–5320.
17. Chen,F. and Wilusz,J. (1998) Auxiliary downstream elements are required for efficient polyadenylation of mammalian pre-mRNAs. *Nucleic Acids Res.*, **26**, 2891–2898.
18. Veraldi,K.L., Arhin,G.K., Martincic,K., Chung-Ganster,L.H., Wilusz,J. and Milcarek,C. (2001) hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells. *Mol. Cell Biol.*, **21**, 1228–1238.
19. Alkan,S.A., Martincic,K. and Milcarek,C. (2006) The hnRNPs F and H2 bind to similar sequences to influence gene expression. *Biochem. J.*, **393**, 361–371.
20. Dalziel,M., Nunes,N.M. and Furger,A. (2007) Two G-rich regulatory elements located adjacent to and 440 nucleotides downstream of the core poly(A) site of the intronless melanocortin receptor 1 gene are critical for efficient 3' end processing. *Mol. Cell Biol.*, **27**, 1568–1580.
21. Oberg,D., Fay,J., Lambkin,H. and Schwartz,S. (2005) A downstream polyadenylation element in human papillomavirus type 16 L2 encodes multiple GGG motifs and interacts with hnRNP H. *J. Virol.*, **79**, 9254–9269.
22. Millevoi,S., Geraghty,F., Idowu,B., Tam,J.L., Antoniou,M. and Vagner,S. (2002) A novel function for the U2AF 65 splicing factor in promoting pre-mRNA 3'-end processing. *EMBO Rep.*, **3**, 869–874.
23. Sgourou,A., Papachatzopoulou,A., Psiouri,L., Antoniou,M., Zombos,N., Gibbs,R. and Athanassiadou,A. (2002) The beta-globin C->G mutation at 6bp 3' to the termination codon causes beta-thalassaemia by decreasing the mRNA level. *Br. J. Haematol.*, **118**, 671–676.
24. Millevoi,S., Loulergue,C., Dettwiler,S., Karaa,S.Z., Keller,W., Antoniou,M. and Vagner,S. (2006) An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. *EMBO J.*, **25**, 4854–4864.
25. Gunderson,S.I., Beyer,K., Martin,G., Keller,W., Boelens,W.C. and Mattaj,L.W. (1994) The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell*, **76**, 531–541.
26. Markovtsov,V., Nikolic,J.M., Goldman,J.A., Turck,C.W., Chou,M.Y. and Black,D.L. (2000) Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. *Mol. Cell Biol.*, **20**, 7463–7479.
27. Chou,M.Y., Rooke,N., Turck,C.W. and Black,D.L. (1999) hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol. Cell Biol.*, **19**, 69–77.
28. Hu,J., Lutz,C.S., Wilusz,J. and Tian,B. (2005) Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *RNA*, **11**, 1485–1493.
29. Zarudnaya,M.I., Kolomiets,I.M., Potyahaylo,A.L. and Hovorun,D.M. (2003) Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res.*, **31**, 1375–1386.
30. Graveley,B.R., Fleming,E.S. and Gilmartin,G.M. (1996) RNA structure is a critical determinant of poly(A) site recognition by cleavage and polyadenylation specificity factor. *Mol. Cell Biol.*, **16**, 4942–4951.
31. Wu,C. and Alwine,J.C. (2004) Secondary structure as a functional feature in the downstream region of mammalian polyadenylation signals. *Mol. Cell Biol.*, **24**, 2789–2796.
32. Kikin,O., Zappala,Z., D'Antonio,L. and Bagga,P.S. (2008) GRSDB2 and GRS_UTRdb: databases of quadruplex forming G-rich sequences in pre-mRNAs and mRNAs. *Nucleic Acids Res.*, **36**, D141–D148.
33. Takagaki,Y., Ryner,L.C. and Manley,J.L. (1988) Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. *Cell*, **52**, 731–742.

Supplementary information

Materials and methods

RNA substrates

The HBB, C2 and F2 pA signals were amplified by PCR using a forward primer containing the r17 sequence upstream of the specific pA signal sequence (primer sequences are available in Supplementary data). These PCR products were used for a second round of PCR using a forward primer containing the T7 promoter upstream of the r17 sequence. Capped, uniformly ³²P-labelled RNAs used for cleavage, polyadenylation and UV cross-linking assays, were obtained by *in vitro* transcription of these PCR products. The L3 pA signal was obtained by *in vitro* transcription of the R17-L3 linearised plasmid previously described¹.

Cleavage and polyadenylation reactions

In vitro cleavage reactions were performed by incubating ³²-P labelled RNA substrates with NE for 90 min at 30°C in the presence of purified recombinant GST-R17 or GST-R17-PTB fusion proteins as previously described². The *in vitro* polyadenylation assays using NE were performed for 15 min at 30°C as for the cleavage reaction except that it was done without cordycepin and in the presence of 0.7 mM ATP and Mn²⁺. hnRNP H/F depleted NE were performed by three consecutive rounds of incubation of NE with 1 µg of streptavidine/agarose-bound SVL GRS for 1 h at 4°C. Alternatively, depletion of hnRNP H/F was performed by addition of 1 µg of SVL GRS (or a control RNA) in the cleavage assay. Reconstituted polyadenylation reactions were performed for 15 min at 37°C as described (Millevoi 2006). Analysis and quantification of cleavage/polyadenylation reactions after RNA extraction and resolution on a denaturing 6% polyacrylamide gel was done by PhosphorImager (Molecular Dynamics) analysis. Cleavage activity was calculated by dividing the amount of upstream cleavage product by the sum of cleavage plus precursor products. Polyadenylation efficiency was calculated by dividing the precleaved product by the polyadenylated product.

UV crosslinking/IP

Purified GST-tagged R17, R17-PTB or PTB proteins were incubated with ³²-P labelled transcripts corresponding to the HBB, C2, F2 or L3 pA signals under

cleavage conditions for 30 min at room temperature. The reaction mixtures were then irradiated on ice with UV light (254 nm) in a Stratalinker (Stratagene) at 0.4 J/cm² at 10 cm distance. Then, 50 units of RNase ONE (Promega) was added and the reaction mixtures incubated for 30 min at 37°C. SDS gel loading buffer was added and the samples boiled for 2 min before fractionation on a 10% SDS-PAGE. For IP analysis, protein G sepharose was incubated with antibodies prior to addition of UV cross-linked complexes in NETN buffer (20 mM Tris at pH 8.0, 100 mM NaCl, 0.5 % NP-40, 0.5 mM EDTA). After 1 h at 4°C, bound proteins were washed 6 times with NETN buffer and eluted by adding SDS loading buffer to the beads. Antibodies: hnRNP H/F (mAb 1G11, Abcam), PTB (mAb Bb7, Abcam), U2AF65 (mAb MC3; a gift of M. Carmo-Fonseca), CFIm (polyclonal; a gift of W. Keller), CstF 64 (mAb 3A1; a gift of C. McDonald)

PCR template for in vitro transcription

DNA templates for in vitro transcription were obtained by two rounds of PCR. The forward primers containing the r17 sequence upstream of the indicates gene specific sequence and reverse primers used in the first PCR were the following:

pA signal	forward	reverse
HBB	AATTTCTATTTAAAGGTTTCCTT	GTTTGAAGCTAGCTCTTCATTTCTTTATG
C2	TTTCACATGGAATTTCCCAG	GCTCTTGGAGTCATTCTGGC
F2	CTAAAACCTATGGTTCCAAT	TCCCACCTCAGCCTCCCGAG

In the second round of PCR the three PCR product were amplified using a forward primer containing the T7 promoter sequence upstream of the r17 sequence and reverse primers as in the first round PCR.

Bioinformatical sequence motif search

In order to select genes with the PYR2 consensus motif [C/T]TTTC[C/T]TGCT in their 3' ends, we developed 3 custom bioinformatics programs that extract genomic sequences, search these sequences for regular expressions and then partition the sequences based on their motif content/topology. The extraction used the UCSC

human genome sequence (hg18) in conjunction with the transcript coordinates provided by the refFlat.txt SQL dump and the polyA prediction track file polyApredict.txt. Coordinates of transcripts and polyA sites were used to extract genomic sequences encompassing all 3' UTR exons and downstream genomic sequence. A set of regular expressions were generated to match the sequence motifs of interest and to search the above described 3' end sequence database. Finally, a program was developed to partition all sequences with matching motifs based on the order and proximity of the motifs in order to select the final set of genes that contained the desired motif topology.

Recombinant proteins

GST fusion proteins were cloned in the pGEX 2T *E. coli* expression vector. His-tagged hnRNP H and hnRNP F encoding plasmids were kindly provided by Douglas Black. All fusion proteins were expressed in *E. coli* at 20°C overnight and purified to homogeneity by glutathione agarose chromatography. The His-tagged bovine PAP (residues 1-694), hnRNP H and hnRNP F were purified to homogeneity by Ni²⁺-NTA chromatography as described³. Recombinant purified PABP was kindly provided by Maria Carmo-Fonseca.

GST pull-down assays

GST pull-down assays were performed by incubating 1 µg of purified GST-R17 or GST-R17-PTB bound to 20 µl of glutathione agarose beads with 200 µg of HeLa nuclear extracts (Fig. 3A) in NETN buffer (20 mM Tris at pH 8.0, 100 mM NaCl, 0.5 % NP-40, 0.5 mM EDTA) for 60 min at 4°C. Beads were then washed 5 times, treated with 10 µg/ml RNase A at room temperature for 30 min and washed again. Protein elution was performed by adding SDS loading buffer to the beads. Eluted proteins were resolved by 10 % SDS-polyacrylamide gel electrophoresis (PAGE) and analyzed by western-blot. GST pull-down assays in Fig. 3B and Fig. 6H were performed as described above except that it was done with 1 µg of His-tagged hnRNP H, hnRNP F or PAP and bound proteins were visualized by Coomassie-blue staining.

1. Millevoi, S. et al. A novel function for the U2AF 65 splicing factor in promoting pre-mRNA 3'-end processing. *EMBO Rep* **3**, 869-874 (2002).
2. Millevoi, S. et al. An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. *Embo J* **25**, 4854-4864 (2006).
3. Gunderson, S.I. et al. The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell* **76**, 531-541 (1994).

Supplementary Table I PYR2 motif containing transcripts

Acc. No.	Symbol	Sequence
NM_001159	AOX1	TCCTTTCTGCT-45-GGG-317-AATAAA-10-GGG-183-GGG-6-GGG
NM_024769	ASAM	GACTTTCTGCT-73-GGG-53-AGAAAA-23-GGG-11-GGG-42-GGG-79-GGG-163-GGG
NM_198188	ASTN2	CTCTTTTTGCT-18-GGG-24-GGG-32-ATTAAA-26-GGG-16-GGG-100-GGG
NM_030660	ATXN3	AACTTTCTGCT-80-GGG-62-AATAAA-552-GGG-213-GGG
NM_174978	C14orf39	TTCTTTTTGCT-199-AATAAAGGG-143-GGG
NM_152448	C15orf43	CTCTTCTTGCT-252-AATAAA-167-GGGGG-7-GGGGG
NM_032266	C2orf16	TTCTTCTGCT-59-GGGGGGG-4-GGG-7-GGG-20-AATAAAGGG-39-GGGCGGG-5-GGG-13-GGG-37-GGG-64-GGG-20-GGGCGGGG
NM_004642	CDK2AP1	TTCTTTTTGCT-226-AATAAA-33-GGG-40-GGG-66-GGG
NM_018223	CHFR	AACTTTTTGCT-102-AATAAA-31-GGG-9-GGG-21-GGG-121-GGG
NM_024591	CHMP6	GGCTTCTTGCT-58-GGG-33-GGG-41-GGGG-11-GGG-20-GGG-15-GGG-35-GGG-23-GGG-27-GGG-34-GGG-35-ATTAAA-33-GGG-6-GGG-23-GGG-5-GGGTGGG-59-GGG-12-GGG-24-GGG-26-GGG
NM_001324	CSTF1	AACTTCTTGCT-356-ATTAAA
NM_024090	ELOVL6	CTCTTCTTGCT-19-GGG-91-GGG-38-AATAAA-77-GGGAGTGGGG-8-GGG
NM_001974	EMR1	TCCTTCTTGCT-115-GGG-9-GGGGG-236-AATAAA-91-GGG-70-GGGGG-35-GGG
NM_000569	FCGR3A	ACCTTTCTGCT-141-GGGG-130-GGG-11-AATAAA-154-GGG-17-GGG-6-GGG-6-GGG
NM_000570	FCGR3B	ACCTTTCTGCT-141-GGGG-144-AATAAA-154-GGG-4-GGGGG-8-GGG-6-GGG-6-GGG
NM_001042729	FGR	CCCTTCTGCTGGG-21-GGGGTCTGGG-11-GGG-69-AGTAA-48-GGGAGCGGGG-9-GGGGG-32-GGG-10-GGG-20-GGGG
NM_152440	FLJ32549	CACCTTCTGCT-217-GGG-90-GGG-27-GGG-35-AATAAA-119-GGG-79-GGG
NM_024707	GEMIN7	CACCTTCTGCT-54-GGG-72-GGG-92-GGG-23-GGG-4-GGGGTGGG-23-GGGGG-5-ATTAAA-114-GGGG-15-GGG-6-GGG-11-GGG-17-GGG-32-GGGG-47-GGGATTGGG-23-GGGCGGG-14-GGG-152-GGG-385-GGG
NM_001495	GFRA2	AGCTTCTGCT-74-GGG-165-GGG-46-GGG-7-GGGGG-34-AATAAA-37-GGG-58-GGGCTGGG
NM_147148	GSTM4	GGCTTCTGCT-26-GGG-57-AATAAA-191-GGGGG
NM_000518	HBB	CGCTTCTTGCT-51-GGGGG-10-GGG-23-AATAAA-59-GGG-5-GGG-49-GGG
NM_002125	HLA-DRB5	AGCTTCTGCT-228-ATTAAA-74-GGG
NM_001548	IFIT1	ATCTTTCTGCT-180-AATAAA
NM_006391	IPO7	TTCTTCTGCT-75-AATAAT-37-GGG-38-GGG-5-GGGGGTGGGGGGGATGGG-12-GGG-78-GGG-45-GGG-105-GGG
NM_000238	KCNH2	GACTTTCTGCT-76-AATAAA-30-GGGAGGG-68-GGG-24-GGG-20-GGG-10-GGGGG-8-GGGGG-19-GGG

Supplementary Table I (continue)

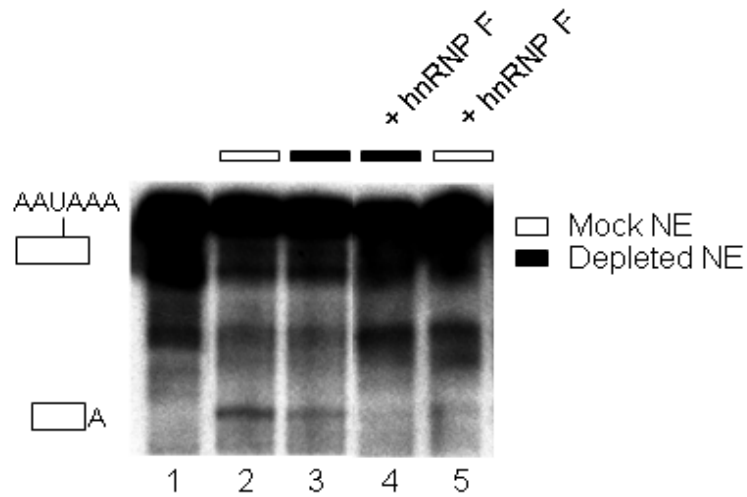
Acc. No.	Symbol	Sequence
NM_020818	KIAA1409	AGCTTTCTGCT-23-GGG-97-GGG-60-GGG-50-AATAAA-55-GGG
NM_006121	KRT1	TTCTTTCTGCT-157-GGG-118-AATAAA-60-GGG-43-GGG-98-GGG
NM_014572	LATS2	TTCTTTTCTGCT-369-ATTA-111-GGG-16-GGG-18-GGG
NM_207338	LCTL	CCCTTTCTGCT-4-GGG-71-GGG-32-ATTA-225-GGG-33-GGG
NM_001004432	LINGO4	TTCTTTCTGCT-69-GGG-9-GGG-71-ATTA-12-GGG-166-GGG
NM_006726	LRBA	GCCTTTTCTGCT-20-AATAA
NM_016383	LUZP4	AACTTTCTGCT-113-GGG-47-GGG-53-AATAA-129-GGG-33-GGG
NM_144634	LYZL4	TGCTTTCTGCT-79-GGG-13-AATAA-20-GGG
NM_002405	MFNG	GCCTTTCTGCT-4-GGG-7-GGG-16-GGG-29-GGG-41-GGG-16-GGG-79-ATTA-64-GGG-104-GGG
NM_032895	MGC14376	TCCTTTCTGCT-19-GGG-106-ATTA-53-GGG-199-GGG-12-GGG-33-GGG
NM_021107	MRPS12	ACCTTTCTGCTGGG-25-GGGAGGGG-4-AATAA-177-GGG
NM_032853	MUM1	AACTTTCTGCT-61-ATTA-56-GGG-18-GGG-21-GGG-15-GGG-20-GGG-4-GGG-26-GGG
NM_016143	NSFL1C	CTCTTCTGCT-53-GGG-33-GGG-52-GGG-57-GGG-34-ATTA-38-GGG
NM_006451	PAIP1	CACCTTTCTGCT-35-AATAA
NM_032439	PHYHIPL	TGCTTTCTGCT-107-AATAA
NM_002666	PLIN	TTCTTTTCTGCT-17-AATAA-103-GGG
NM_022361	POPDC3	CACCTTTTCTGCT-119-AATAA-146-GGG-6-GGG-6-GGG-27-GGG
NM_000307	POU3F4	GCCTTTCTGCT-4-AATAA-58-GGG-7-GGGGCGGGGTGGG
NM_002736	PRKAR2B	TACTTTTCTGCT-85-GGG-71-GGG-92-ATTA-128-GGG
NM_003675	PRPF18	TTCTTTCTGCT-70-GGG-47-ATTA-107-GGG-276-GGG
NM_006711	RNPS1	TTCTTTTCTGCT-111-AATAA-28-GGG-95-GGG-12-GGG-15-GGG
NM_002946	RPA2	CACCTTCTGCT-72-GGG-58-AATAA-206-GGG
NM_018115	SDAD1	GGCTTCTGCT-28-GGG-251-AATAA-84-GGG-106-GGG
NM_012248	SEPHS2	CTCTTCTGCT-33-GGG-72-GGG-106-AATAA-138-GGG-20-GGG
NM_013368	SERTAD3	ACCTTCTGCT-24-GGG-5-GGG-15-GGG-17-GGG-51-GGGGTGGAGGG-96-GGG-23-GGG-69-AATAA-15-GGG-28-GGG-5-GGGAGGG-45-GGG-80-GGG
NM_003027	SH3GL3	TTCTTTTCTGCT-218-AATAA-143-GGG-51-GGG
NM_148918	SHMT1	GGCTTCTGCT-205-AATAA-150-GGG-17-GGGACGGG
NM_018593	SLC16A10	TGCTTCTGCT-48-GGG-7-GGG-179-AATAA-152-GGGAGGG
NM_003795	SNX3	CTCTTCTGCT-301-GGG-85-ATTA-82-GGG-87-GGG

Supplementary Table I (continue)

Acc. No.	Symbol	Sequence
NM_001017371	SP3	TC TTTTTTGCT -303- <u>AATAAA</u> -63-GGGG
NM_032508	TMEM185A	G ACTTTCCTGCT -54- <u>AATAAA</u> -48-GGG-91-GGGG-39-GGG-5-GGGG-4-GGG-8-GGG
NM_022553	VPS52	TC TTTCCTGCT -120-GGGG-132- <u>AATAAA</u> -105-GGG-17-GGG-75-GGG
NM_198285	WDR86	GC TTTTCCTGCT -64-GGG-4-GGGGGG-6- <u>AATATA</u> -75-GGG-14-GGG-4-GGG-31-GGG-7-GGGGG-8-GGG-8-GGGG-33-GGGG-8-GGGG
NM_004625	WNT7A	GT TTTTCTGCT -5-GGG-10-GGG-18-GGGGG-128-GGG-57- <u>ATTAAA</u> -20-GGG
NM_194320	ZNF169	TT TTTTTTGCT -161-GGG-78- <u>ATTAAA</u> -48-GGG

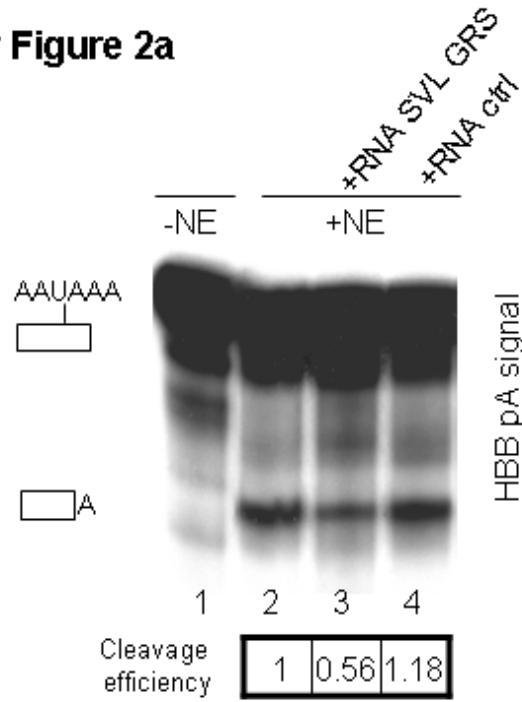
Supplementary Table I. Human transcripts containing the [C/T]TTTC[C/T]TGCT sequence (in bold) upstream of the poly(A) signal (underlined). The nucleotide distance between each motif is indicated.

Supplementary Figure 1

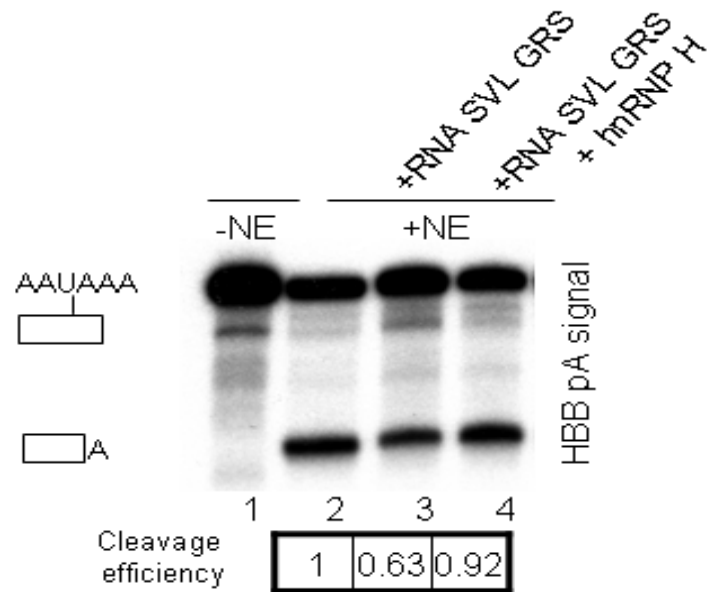


Supplementary Figure 1. Function of hnRNP F in HBB 3' end cleavage. *In vitro* cleavage assay using the ^{32}P labelled r17/HBB substrate in the presence of mock-depleted NE (white boxes) (lanes 2 and 5) or NE depleted of hnRNP H/F (black boxes) (lanes 3 and 4), supplemented (lanes 4 and 5) or not (lanes 2 and 3) with recombinant hnRNP H (1 pmol). The low cleavage efficiency is due to the depletion procedure. Identities of the uncleaved and cleaved products are shown on the left.

Supplementary Figure 2a



Supplementary Figure 2b



Supplementary Figure 2. HnRNP H sequestration and add back. (a) *In vitro* cleavage assay using the ^{32}P labelled r17/HBB substrate in the absence (lane 1) or in the presence of a SVL GRS (lane 2) or a RNA control (lane 3). (b) *In vitro* cleavage assay using the ^{32}P labelled r17/HBB substrate in the absence (lane 1) or in the presence of the SVL GRS (lanes 2 and 3) and with recombinant hnRNP H (1 pmol, lane 3) in NE. Identities of the uncleaved and cleaved products are shown on the left.

PROTOCOLES

Culture cellulaire

Lignées cellulaires

Les lignées cellulaires issues d'adénocarcinome pulmonaire humain (A549), de cancers du colon (HCT116) et de carcinome ovarien (OVCAR et SKOV) sont cultivées dans du milieu RPMI-1640 (Roswell Park Memorial Institute) (Lonza) supplémenté avec 10% de sérum de veau foetal (SVF) (Lonza). Les lignées cellulaires cancéreuses mammaires (MCF7), ou issues de métastase du col de l'utérus (Hela) sont cultivées dans du milieu DMEM (Dulbecco's modified Eagle's), supplémenté avec 5 ou 10% de SVF.

Les lignées sont maintenues sous une atmosphère contenant 5% de CO₂ et 95 % d'humidité, à une température de 37°C. Les différents types cellulaires ci-dessus sont cultivés avec deux passages par semaine. Les cellules HaCat n'ont jamais été utilisées à plus de 20 passages cellulaires pour les différentes expériences. Pour obtenir à chaque passage, environ 4 millions de cellules dans une flasque T75 (75 cm²), ensemercer 500000 cellules le lundi pour le vendredi.

Congélation et décongélation de lignées humaines

Congélation :

A partir d'une flask ou une boîte de pétri :

- Trypsiner les cellules.
- Reprendre avec du milieu de culture contenant du sérum, compter (Coulter ou Mallassez).
- Centrifuger 5' à 1000 rpm.
- Récupérer le culot cellulaire dans du milieu de culture contenant du sérum.
- Centrifuger 5' à 1000 rpm.
- Reprendre le culot dans x ml de milieu de culture avec sérum et 10% DMSO (pour avoir ~1,5 million de cellules par ml).
- Répartir par 1 ml dans des ampoules de congélation, marquer les ampoules (passage, date, manipulateur, cellules).
- Laisser les ampoules 1h à 4°C puis une nuit à -80°C.
- Conserver les ampoules dans de l'azote liquide.

Décongélation :

- Dans une boîte de pétri de diamètre 60 mettre 3 ml de milieu de culture.
- Amener la boîte à 37°C.
- Plonger rapidement l'ampoule de l'azote liquide vers 37°C.
- Mettre le contenu de l'ampoule dans la boîte de culture.
- Laisser adhérer (entre 4 à 12h) et changer le milieu car le DMSO est toxique.

Irradiation aux UV

Pour l'irradiation, le milieu est remplacé par du PBS (tampon phosphate), les cellules sont exposées aux UVC (50 J/m²) en utilisant le système de lampes UV stratalinker (254nm ; Stratagene). Après irradiation, le PBS est remplacé par le milieu adéquat et les cellules sont incubées à 37°C pendant 16h.

Transfections transitoire de siARN

Pour transfecter les siARN, nous utilisons le protocole de reverse transfection en utilisant l'agent transfectant Lipofectamine TM RNAimax (Invitogen), qui ne nécessite pas de faire adhérer au préalable les cellules. Les siARN sont conservés dans de petits aliquots de 25 µl à -20°C à la concentration de 20 µM. Le protocole est adapté en fonction du nombre de cellules transfectées et de la concentration finale voulue de siARN.

Pour des plaques 6 puits [si]=2.5nM

Pour un puit

-A additionner dans cette ordre

500µl Opti MEM

0.25µl de siARN a 20µM

3.5µl Lipofectamine TM RNAimax

-Incuber 15 min à température ambiante.

-Déposer 500µl de ce mélange par puit puis ajouter 1.5 ml de milieu de culture contenant entre 100 000 et 300 000 cellules (fonction du type cellulaire).

Transfection transitoire de plasmide et mesure de l'activité luciférase

Les constructions plasmidiques sont transfectées transitoirement dans de boîtes 12 puits,ensemencées la veille à la densité de 75 000 cellules par puits dans 650µl de milieu, en utilisant comme agent transfectant le jetPeI (Polyplus-transfection).

-Préparer un premier mélange de NaCl/jetPeI:

37.5 µl de NaCl/puits avec 1.5µl de jetPeI/puit

-Préparer un deuxième mélange de NaCl/solution plasmidique

37.5 µl de NaCl/puits avec 0.6 ug (plasmide d'intérêt) et 0.15 µg (plasmide contrôle)/puit.

-Ajouter au deuxième mélange de NaCl/solution plasmidique le premier mélange de NaCl/jetPeI (Toujours le jet dans l'ADN).

-Incuber 15 min à température ambiante.

-Distribuer 75 µl de ce mix par puit.

Les cellules sont lysées 24h après avec 100µl par puit de Passive Lysis Buffer 1X (Promega) pendant 30 min à température ambiante sous agitation. Les activités luciférase *renilla* et *firefly* de 10µl de lysat sont mesurées simultanément avec un luminomètre (MicroLumat Plus LB 96V, Berthold), à l'aide du kit Dual Luciferase Assay (Promega).

Immunoprécipitation

Immunoprécipitation d'ARN avec couplage covalent aux UV *in vivo*

Couplage covalent aux UV :

Pour cette expérience, 5 points sont réalisées pour chaque échantillon pour le couplage covalent au UV puis sont mises en commun pour la partie immunoprécipitation.

Assembler l'ARN et les protéines dans un volume de 10 µl réaction dans des plaques de microtitration.

Tampon D (20mM Hepes pH 7,9 ; 20% glycerol ; 100mM KCl ; 0,2mM EDTA ; 3 mM DTT ; 0,2mM PMSF).

EN 10 µg/µl (Extrait Nucléaire resuspendu dans du Tampon D).

ARN 80.000 cpm

-Ordre d'addition: Tampon D XµL Vf=10µl

RNA 1µL

EN . XµL Vf=10µl

H2O 4µL(-1µl tmpyp4 au 1/750 [0.1µM] ou - au 1/150 [0.5µM])

-Incuber 30 min à température ambiante

-Exposer aux UVC (4000 J/m²) en utilisant le système de lampe UV stratalinker.

-Ajouter 1µl de RNase ONE (Rnase One Ribonuclease, 5 000u, 10u/µL, Promega).

-Incuber 45 min à 37°C.

Préparation des billes : NETN

-Prélever le nombre de billes nécessaire pour l'ensemble de l'expérience (IP et préhybridation) en utilisant 20µl de billes A/G+ (TEBU) par condition.

-Centrifuger une minute à 5000 rpm et retirer le surnageant.

-Faire un premier lavage avec de l'eau, centrifuger et retirer le surnageant puis faire 2 autre lavage dans un tampon NETN de lavage (20mM Tris/Cl pH8 ; 150mM NaCl ; 0,5% NP40) et retirer le surnageant.

-Ajouter 200µl de tampon NETN +10µg d'anticorps pour 20µl de billes.

-Incuber pendant 1h sur une roue à 4°C.

-Faire deux lavage avec du tampon NETN puis resuspendre les billes dans 250µl de NETN.

Immunoprecipitation :

- Ajouter les échantillons aux billes et incuber 3h sur une roue à 4°C.
- Faire cinq lavages avec du tampon NETN et retirer tout le surnageant.
- Ajouter 4µl de bleu protéine (5X) et faire chauffer une minute à 90°C.
- déposer sur un gel proteique acrylamide 12%.

Immunoprécipitation d'ARN avec couplage covalent au formaldéhyde *in vivo*Couplage covalent *in vivo*

- Faire 2 lavages avec 10ml de PBS froid pour chaque boîte de cellules à confluence de diamètre 140.
- Gratter les cellules dans le PBS et les transférer dans un tube 50ml.
- Centrifuger à 2000 rpm pendant 5 minutes à 4°C et reprendre le culot cellulaire avec 10 ml de PBS froid par condition.
- Ajouter le formaldéhyde (Sigma F1635-500 ml 37 %) 0,5 %V/V (soit 137 µl de formaldéhyde à 37 % dans 10 ml de PBS), incuber 10 min à température ambiante en agitant doucement.
- Arrêter la réaction en ajoutant de la glycine (pH 7; 1,25 M) pour obtenir une concentration finale de 0,25 M et incuber 5 minutes à température ambiante.
- Centrifuger à 2000 rpm pendant 5 minutes à 4° C.
- Faire 4 lavages au PBS froid. Faire un dernier lavage, en ajoutant 500 µl de PBS et transférer alors le culot cellulaire repris dans le PBS dans tube de 1.5ml.
- Centrifuger à 2000 rpm pendant 5 minutes à 4 °C, aspirer le PBS et garder les culots cellulaires (après cette étape, il est possible de conserver les culots cellulaires à -80°C, après congélation rapide dans l'azote liquide ou de faire immédiatement la lyse cellulaire).

Préparation des billes :

- Prélever le nombre de billes nécessaire pour l'ensemble de l'expérience (IP et préhybridation) en utilisant 20µl de billes A/G+ (TEBU) par condition.
- Centrifuger une minute à 5000 rpm et retirer le surnageant.

- Faire un premier lavage avec de l'eau, centrifuger et retirer le surnageant puis faire 2 autres lavages dans un tampon RIPA (50mM Tris-HCl pH 7.4; 1% NP40; 0.5% NaDeoxycolate; 0.05% SDS; 1mM EDTA; 150mM NaCl)
- Bloquer les billes avec du tRNA (5µg/µl) et de la BSA (1µg/µl) dans du tampon de lavage pendant 5h sur une roue à 4°C.
- Après 3 lavages, ajouter un volume de tampon de lavage équivalent au volume de billes.

Lyse des cellules

- Reprendre le culot cellulaire dans 1ml de tampon RIPA et ajouter extemporanément des Inhibiteurs de protéase (1X).
- Incuber 30 minutes en chambre froide puis centrifuger à 10000 g pendant 10 min à 4 °C.
- Récupérer le surnageant et ajouter 5µl RNAsin (40 U/ml) par échantillon (à cette étape il est possible de congeler rapidement dans l'azote liquide les échantillons puis de les stocker à -80°C).

Préhybridation

- Incuber les extraits cellulaire pendant 30 minutes sur roue à 4°C avec 20µl de billes bloquées.
- Centrifuger ensuite les tubes à 10000rpm 5minutes puis récupérer le surnageants et doser les extraits protéiques.
- Calculer le volume nécessaire pour 1mg d'extrait par condition.
- Prélever 1/10 du volume pour l'input.

Incubation des anticorps sur les billes

- 20µg de billes par condition sont préincubées avec 10µg d'anticorps dans du tampon RIPA 16 heures sur roue à 4°C.
- Faire deux lavages avec du tampon RIPA.

Immunoprecipitation :

- Ajouter 1mg d'échantillons aux billes contenant les anticorps.
- Compléter avec du RIPA pour obtenir un volume final de 500µl.

- Incuber 16 heures sur roue à 4°C.
- Faire 7 lavages avec le tampon RIPA et un dernier lavage à l'eau puis resuspendre dans 100µl d'eau.

Reverser le couplage covalent

- Ajouter 100µl de tampon 2X RB (50mM Tris-Cl pH7 ; 5mM EDTA ; 10mM DTT ; 1% SDS).
- Incuber 45 minutes à 70°C.

Récupérer l'ARN immunoprécipité

- Extraction au trizol LS.
- Traitement à la turboDNase.
- Extraction au trizol LS.
- Quantifier l'ARN par RT-PCRq.

Immunoprécipitation d'ADN avec couplage covalent aux UV *in vitro*

Couplage covalent *in vivo*

- Faire 2 lavages avec 10ml de PBS froid pour chaque boîte de cellules à confluence diamètre 140.
- Gratter les cellules dans le PBS et les transférer dans un tube 50ml.
- Centrifuger à 2000 rpm pendant 5 minutes à 4°C et reprendre le culot cellulaire avec 10 ml de PBS froid par condition.
- Ajouter le formaldéhyde (Sigma F1635-500 ml 37 %) 0,5 %V/V (soit 137 µl de formaldéhyde à 37 % dans 10 ml de PBS), incuber 10 min à température ambiante en agitant doucement.
- Arrêter la réaction en ajoutant de la glycine (pH 7; 1,25 M) pour obtenir une concentration finale de 0,25 M et incuber 5 minutes à température ambiante.
- Centrifuger à 2000 rpm pendant 5 minutes à 4° C.
- Faire 4 lavages au PBS froid. Faire un dernier lavage, en ajoutant 500 µl de PBS et transférer alors le culot cellulaire repris dans le PBS dans tube de 1.5ml.

-Centrifuger à 2000 rpm pendant 5 minutes à 4 °C, aspirer le PBS et garder les culots cellulaires (après cette étape, il est possible de conserver les culots cellulaires à -80°C, après congélation rapide dans l'azote liquide ou de faire immédiatement la lyse cellulaire).

Préparation des billes :

-Prélever le nombre de billes nécessaire pour l'ensemble de l'expérience (IP et préhybridation) en utilisant 50µl de billes A/G+ (TEBU) par condition.

-Centrifuger une minute à 5000 rpm et retirer le surnageant.

-Faire un premier lavage avec de l'eau, centrifuger et retirer le surnageant puis faire 2 autres lavages dans un tampon RIPA (50mM Tris-HCl pH 7.4; 1% NP40; 0.5% NaDeoxycolate; 0.05% SDS; 1mM EDTA; 150mM NaCl).

-Bloquer les billes avec du 30 µl de l'ADN de sperme de Saumon à 10 mg/ml et 35,5 µl BSA à 20 mg/ml.

-Après 3 lavages, ajouter un volume de tampon de lavage équivalent au volume de billes.

Lyse des cellules

-Resuspendre les culots cellulaires (200µL pour 1 million de cellules) 5mM Pipes pH8 ; 85mM KCl ; 0.5% NP-40 et inhibiteurs de protéase).

-Incuber 10 minutes dans la glace.

-Homogénéiser 20 fois à l'aide d'un dounce (piston B loose).

-Centrifuger à 4000 rpm pendant 5 minutes à 4 °C.

-Resuspendre les culots nucléaires dans 100 µL (pour 1 million de cellules) dans du tampon de lyses.

-Incuber 10 minutes dans la glace.

-Sonifier les cellules (Branson sonifier) 10 fois 10 sec.

-Doser l'ADN avec un nanodrop et diluer 10 fois.

Préhybridation

-Incuber 200µg d'ADN pendant 1 heure sur roue à 4°C avec 40 µl de billes bloquées.

-Centrifuger ensuite les tubes à 10000rpm 5minutes.

Immunoprecipitation :

-Incuber 200µg d'ADN préhybrider avec 3µg d'anticorps 16 heures sur roue à 4°C.

-Ajouter 40µl de billes bloquées 2 heures sur roue à 4°C.

-Faire 5 lavages au tampon Wash (100mM Tris pH8,8 ; 500mM LiCl ; 1%NP-40 ; 1% NaDoc).

-Compléter avec du RIPA pour obtenir un volume final de 500µl.

-Incuber 16 heures sur roue à 4°C.

-Faire 7 lavages avec le tampon RIPA et un dernier lavage à l'eau puis resuspendre dans 100µl d'eau.

-Resuspendre dans du 200µl de tampon TE (10 mM Tris-Cl, pH 7.5. 1 mM EDTA).

Reverser le couplage covalent

-Ajouter 0.3 µL de RNase (30mg/mL) incuber 20min à 37°C.

-Ajouter 4µL de SDS10% incuber 16 heures à 70°C.

-Ajouter 4µL de proteinase K (10mg/mL) incuber 2 heures à 45°C.

-Purifier l'ADN par extraction au phénol chloroforme.

- Quantifier l'ADN par PCRq.

Extraction d'ARN

Extraction d'ARN au trizol (Invitrogen)

- Ajouter 1 ml de trizol pour 10 cm² de surface de culture et homogénéiser.
- Laisser 5 min à température ambiante (stockage au -80°C possible).
- Ajouter du chloroforme (200 µl par ml de trizol).
- Agitation très forte à la main pendant 15 secondes.
- Laisser 2 à 3 minutes à température ambiante.
- Centrifugation à 4°C pendant 15 min à 12 000 g maximum.
- Récupérer délicatement la phase aqueuse (~600µl).
- Rajouter un volume 600µl d'isopropanol.
- Précipiter 10 minutes à température ambiante.
- Centrifugation à 4°C pendant 15 min à 12 000 g.
- Enlever le surnageant.
- Ajouter 1 ml d'EtOH 75%.
- Centrifugation à 4°C pendant 5 min à 12000 rpm.
- Sécher le culot d'ARN.

Traitement à la Turbo DNase (Ambion)

- Resuspendre le culot dans 41µl d'eau.
- Ajouter un 9µl d'un mixe contenant : 0.5µl de DTT 0.1M ; 5µl de tampon Turbo DNase 10X buffer ; 1µl d'RNAsin (40U/ml Invitrogen).
- Incuber 20 minutes à 37°C.

Extraction d'ARN au trizol LS (Invitrogen)

- Ajouter 3 volumes de trizol LS pour 1 volume de liquide (150µl pour 50µl d'ARN)
- Laisser 5 min à température ambiante (stockage au -80°C possible)
- Ajouter du chloroforme (40 µl pour 150 µl de trizol LS).
- Agitation très forte à la main pendant 15 secondes.
- Laisser 3 à 5 minutes à température ambiante.

- Centrifugation à 4°C pendant 15 min à 12 000 g maximum.
 - Récupérer délicatement la phase aqueuse, non colorée (~100µl).
 - Rajouter un volume 100µl d'isopropanol.
 - Précipiter 10 minutes à température ambiante.
 - Centrifugation à 4°C pendant 30 min à 12 000 g.
 - Enlever le surnageant.
 - Ajouter 1 ml d'EtOH 75%.
- Centrifugation à 4°C pendant 5 min à 12000 rpm.
- Sécher le culot d'ARN.
 - Resuspendre le culot dans 30µl d'eau.

RT PCR sur ARN

Reverse transcription des ARN totaux

Kit SuperScript III reverse transcriptase

H2O qsp 13 µl

ARN x µl (5µg)

250 ng de random primer

1 µl de 10 mM dNTP

Chauffer à 65°C pendant 5 minutes

Ajouter 4 µl de 5X first Strand Buffer

1 µl de 0,1M DTT

1 µl de RNaseOUT ou RNAsin (40U/µl)

1 µl de SuperScript III reverse transcriptase (200U/µl)

Incuber à 25°C pendant 5 minutes

50°C pendant 1 heure

70C pendant 15 minutes

Conserver à -20°C jusqu'à utilisation.

PCR en temps réel sur ARNm

Kit LightCycler® FastStart DNA MasterPLUS SYBR Green I

Pour une PCR de 10 µl à réaliser dans des capillaires (LightCycler 32 capillaires, Roche)

2,6 µl H₂O

2 µl de 5X MasterMix

0,2 µl de qPCR Primer Forward (à 10 µM)

0,2 µl de qPCR Primer Reverse (à 10 µM)

5 µl d'ADNc dilué (dilution de la reverse transcription au 1/5)

Thermocyclage :

Activation de l'enzyme :

10 minutes à 95°C

Cyclage, 45 cycles :

Dénaturation : 95°C pendant 10 secondes

Hybridation : 60°C en général pendant 10 secondes

Elongation : 72°C pendant 10 secondes (amplicon (pb) / 25)

Courbe de dissociation :

Dénaturation : 95°C pendant 0 seconde

Hybridation : 65°C pendant 1 minute

Variation de température : 0,1°C par seconde

Refroidissement :

40°C pendant 30 secondes

Déterminer le Ct (cycle threshold) qui se définit comme le nombre de cycle où la fluorescence franchit le seuil du bruit de fond. En général, utiliser les paramètres de base de l'appareil pour déterminer le Ct.

Analyser les données en utilisant un contrôle avec la formule $2^{(Ct \text{ gène contrôle} - Ct \text{ gène à étudier})}$.

Fractionnement nucléo-cytoplasmique

- A partir de 30 million de cellules.
- Lavage des cellules avec 1 ml de PBS froid.
- Resuspendre dans 400 µl de Tampon A (10 mM Hepes KOH, pH=7,5 ; 10 mM KCl et à ajouter extemporanément 1mM DTT ; 1mM PMSF + inhibiteur de protéase si on veut extraire des protéines ou RNAsine si on veut extraire l'ARN).
- Incuber 15 min sur glace.
- Ajouter NP-40 à une concentration finale de 0,3%.
- Incuber 10 min dans la glace.
- Culoter le noyau par centrifugation à vitesse maximum (1500 g) pendant 5 min.
- Récupérer la fraction cytoplasmique dans un nouveau tube et la nettoyer par centrifugation à 13000 g pendant 15 min.
- Le culot nucléaire est lavé deux fois dans 1 ml de tampon A et est resuspendu dans 50 µl de tampon B (20 mM Hepes KOH, pH=7,5 ; 400 mM NaCl ; et à ajouter extemporanément 1mM DTT ; 1mM PMSF + inhibiteur de protéase) si on veut extraire des protéines ou directement dans 1 ml de trizol si on veut extraire l'ARN.
- Incuber 30 min la fraction cytoplasmique dans la glace en mélangeant chaque 5 min.
- Les débris nucléaires sont culotés à 13000 g pendant 5 min et la fraction cytoplasmique est récupérée (dans 1 ml de trizol LS si on veut extraire l'ARN).

Protection aux ribonucléases

Hybridation

- ARN sonde (de haute qualité) entre 20 000 et 80 000 CPM.
- ARN nucléaire 5 à 20 µg (fonction de l'expression de l'ARN).
- Ajouter de l'eau pour arriver à un volume final de 10µl.
- Ajouter 1µl de NH₄OAc 5M.
- Ajouter 25µl d'EtOH 100% (Bien homogénéisé).
- Précipitation à -20°C 15 min minimum.
- Centrif 15 min 13 000 rpm 4°C.
- Laver le culot à l'EtOH 70%.
- Resuspendre le culot dans 30µl R-loop (80% formamide ; 20% hybridation 5X).
Hybridation 5X (200 mM Pipes pH 7.4 ; 5 mM EDTA pH8 ; 2M NaCl).
- Dénaturer 20min à 90°C.
- Hybrider 16 heures à 56°C.

Digestion

- Ajouter 1.2µl d'RNase A (10mg/ml Roche) , 2µl d'RNase T1 (550U/ml Roche) et 300µl de tampon d'RNase (300 mM NaCl, 10mM Tris.Cl pH 7.4, 5mM EDTA, pH 8.0)
- Incuber 45 minutes à 30°C.
- Ajouter 1.5µl protéinase K et 7,5µl SDS 20% pendant 30 minutes à 37°C.
- Purifier l'ARN par extraction au phénol chloroforme.
- Précipiter et resuspendre dans du bleu ARN.
- Déposé sur un gel de polyacrylamide urée.

Pause de reverse transcription dépendante de cations

Hybridation

Dans un tube de 1.5µl ajouter :

-2 µl H₂O.

-2 µl ARN transcrit *in vitro*.

-1µl à 100 000 cpm/µl d'oligonucléotide antisens radiomarqué.

-1µl de tampon 6HB Na⁺ ou K⁺ ou Li⁺ (300mM de Hepes ; 600mM de (Na⁺/K⁺/Li⁺) Cl⁻ et 1mM d'EDTA).

-Dénaturer 2min à 90°C.

-Hybrider 5min à 20°C.

Reverse transcription

Ajouter 9 µl de pré-mix contenant:

-10,2 µl d'H₂O.

-7 µl de buff RTB Na⁺ ou K⁺ ou Li⁺ (250mM de trisHCl pH8 ; 200mM de (Na⁺/K⁺/Li⁺) Cl⁻ et 30mM de MgCl).

-5 µl de dNTP 2,5mM.

-0,3 µl de reverse transcriptase AMV (Promega) 20U/µl.

-Mettre 30 min à 37°C

-Précipiter l'ADN reverse transcrit et resuspendre dans du bleu de dépôt.

-Déposé sur un gel de polyacrylamide urée.

AUTEUR : Adrien DECORSIERE

TITRE : Régulation de la maturation en 3' des pré-ARNm de gènes de susceptibilité aux cancers p53 et msh6

DIRECTEURS DE THESE : Dr. Stefania MILLEVOI et Dr. Stéphan VAGNER

RESUME : La maturation en 3' des pré-ARNm constitue une étape de régulation post-transcriptionnelle indispensable à l'expression d'un gène. Elle est composée de deux réactions : le clivage de l'extrémité 3' qui permet de libérer l'ARNm du site de transcription et l'ajout de la queue poly(A) qui contrôle à la fois l'export nucléaire, la stabilité et la traduction du transcrit mature. Plusieurs études montrent, d'une part, que la maturation en 3' est inhibée de manière globale lors de dommages à l'ADN et d'autre part, que des dérégulations de cette maturation peuvent contribuer au développement tumoral. Nos travaux se sont axés sur l'étude de la régulation de la maturation en 3' (i) du pré-ARNm de p53 lors de dommages à l'ADN (ii) du pré-ARNm de MSH6 impliqué dans le syndrome de Lynch. (i) Nos travaux ont mis en évidence que le pré-ARNm de p53 résiste à l'inhibition de la maturation en 3' due aux dommages à l'ADN, grâce au recrutement des protéines hnRNP H/F sur une structure appelée G-quadruplexe située en aval du site de clivage. (ii) Nous avons ensuite identifié une duplication des 20 nucléotides précédant le signal de polyadénylation du pré-ARNm de MSH6 dans deux familles distinctes atteintes du syndrome de Lynch. Cette duplication entraîne une diminution de l'efficacité de la maturation en 3' du pré-ARNm de MSH6 qui semble être à l'origine de la pathologie.

SUMMARY: Pre-mRNA 3'-end processing is an essential post-transcriptional step. It is composed of two reactions: cleavage at the pre-mRNA 3' end that allows release of mRNA from the transcription site and addition of the poly(A) tail that controls nuclear export, stability and the translation of the mature transcript. Several studies show that 3'-end processing is globally inhibited during DNA damage and that deregulation of this process may contribute to tumor development. Our work focused on 3'-end processing regulation of (i) p53 pre-mRNA during DNA damage (ii) MSH6 pre-mRNA involved in Lynch syndrome. (i) We have shown that p53 pre-mRNA resists to DNA damage-dependent inhibition of 3'-end processing through the recruitment of the protein factor hnRNP H/F on a RNA structure called G-quadruplex located downstream of the cleavage site. (ii) We then identified a duplication of 20 nucleotides upstream of the polyadenylation signal at the MSH6 pre-mRNA in two distinct families with Lynch syndrome. This duplication causes a reduction of MSH6 pre-mRNA 3'-end processing efficiency and in consequence, it may be causal of Lynch syndrome.

MOTS CLES : Maturation en 3' des pré-ARNm, G-quadruplexe, P53, MSH6, hnRNP H/F

DISCIPLINE : Biologie moléculaire

LABORATOIRE : INSERM 1037, Institut Claudius Regaud, 20-24 Rue du Pont Saint-Pierre, 31052 Toulouse cedex