



**HAL**  
open science

# Analyse structurale et fonctionnelle de la région des A-repeats de l'ARN Xist impliqué dans l'inactivation du chromosome X dans les mammifères femelles

Anne Savoye

► **To cite this version:**

Anne Savoye. Analyse structurale et fonctionnelle de la région des A-repeats de l'ARN Xist impliqué dans l'inactivation du chromosome X dans les mammifères femelles. Biochimie, Biologie Moléculaire. Université de Lorraine, 2012. Français. NNT : 2012LORR0354 . tel-01749664

**HAL Id: tel-01749664**

**<https://hal.univ-lorraine.fr/tel-01749664v1>**

Submitted on 29 Mar 2018

**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.



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



**UNIVERSITÉ  
DE LORRAINE**

**Ecole Doctorale BioSE (Biologie-Santé-Environnement)**

**Thèse**

**Présentée et soutenue publiquement pour l'obtention du titre de**

**DOCTEUR DE L'UNIVERSITE DE LORRAINE**

**Mention : « Sciences de la Vie et de la Santé »**

par **Anne SAVOYE**

---

**Analyse structurale et fonctionnelle de la région des A-repeats  
de l'ARN Xist impliqué dans l'inactivation du chromosome X  
chez les mammifères femelles**

---

**Le 14 décembre 2012**

**Membres du jury :**

<b>Rapporteurs :</b>	<b>Mme Joëlle MARIE</b>	<b>Directeur de Recherche, INSERM, Paris</b>
	<b>M. Guy CATHALA</b>	<b>Directeur de Recherche, CNRS, Montpellier</b>
<b>Examineurs :</b>	<b>M. Alain KROL</b>	<b>Directeur de Recherche, CNRS, Strasbourg</b>
	<b>Mme Christiane BRANLANT</b>	<b>Directeur de Recherche, CNRS, Nancy</b>
	<b>M. Athanase VISVIKIS</b>	<b>Professeur, Université de Lorraine, Nancy Directeur de Thèse</b>

---

Laboratoire AREMS : ARN, RNP, structure-fonction-maturation, Enzymologie moléculaire et Structurale  
UMR 7214-CNRS UL, Faculté de Médecine, Bâtiment Biopôle, 9 avenue de la Forêt de Haye  
BP 184, 54505 Vandœuvre-lès-Nancy







# Remerciements

---

Je tiens tout d'abord à remercier Madame Christiane Branlant, directrice de l'UMR CNRS-UL 7214, pour m'avoir accueillie au sein de son laboratoire, tout d'abord en stage de master puis en thèse. Je lui suis particulièrement reconnaissante pour m'avoir confié ce travail de recherches, ainsi que pour son aide et ses précieux conseils au cours de ces années.

Je remercie également mon directeur de thèse, le Pr Athanase Visvikis, pour la confiance et la liberté qu'il m'a accordées. Je le remercie aussi pour son aide très importante dans la rédaction de ce mémoire.

Mes remerciements s'adressent également aux Dr Guy Cathala et Dr Joëlle Marie, pour la rapidité avec laquelle ils ont lu mon manuscrit et l'intérêt qu'ils ont porté à mon travail. Merci également au Dr Alain Krol qui a accepté de juger ce travail.

Je tiens également à saluer ici les personnes qui, de près ou de loin, ont contribué à la concrétisation de ce travail de thèse. Je remercie les membres du laboratoire de Génétique Moléculaire Murine pour m'avoir accueillie à plusieurs reprises au sein de leur équipe, en particulier Philip Avner, Philippe Clerc et Agnès Dubois. Merci à tous les membres du laboratoire AREMS pour leur disponibilité et leur gentillesse.

Merci à Sylvain qui a eu la lourde tâche de m'initier à l'inactivation du chromosome X. Un immense MERCI pour m'avoir si bien conseillée et guidée, que ce soit en stage ou en début de thèse. C'est grâce à toi que le sujet a pris son envol, tu y as apporté ton enthousiasme, tes idées et ta détermination.

Un grand merci aux autres membres du box 4, Nath et Flo grâce à qui le jeu des boîtes marquera à jamais le souvenir du bureau, Maxime qui a du nous supporter quelques années, en espérant qu'elles ne lui ont pas paru trop longues.

Une pensée pour tous les étudiants avec qui j'ai partagé une salle de manipe, un plasmide ou un café. Aux anciens du « box 3 » pour leur bonne humeur contagieuse. Au « box 2 » où l'on avait l'impression de se retrouver au café du Commerce. Aux « petits

jeunes » (et aux moins jeunes) pour toutes ces soirées, apéros ou foot-barbecue qui ont ponctué ces six années au laboratoire de moments bien agréables, et de quelques hématomes par la même occasion...

Merci à Alan. Tu as été mon premier encadrant de stage et j'ai appris les rudiments de la biologie moléculaire avec toi. A la paillasse, je me rappelle encore quelques petites astuces de manipes que tu as m'as transmises. Merci pour ton soutien sans faille et tes nombreux conseils avisés pendant toutes ces années, quel que soit le domaine.

Et parce qu'une thèse n'est pas toujours facile pour les proches (si si, il faut bien le dire !), je voudrais remercier mes amis. Merci à Cédric pour ton amitié et tes encouragements, mais aussi pour les conversations biologico-informatiques très intéressantes que nous avons pu avoir. Merci à Franck, ton soutien moral (et logistique !) ainsi que ta patience pendant ces derniers mois pas tout à fait ordinaires me touchent énormément. Et un immense merci à toi ma petite n'Alex, pour ton amitié inébranlable et ta patience sans limite, mais aussi pour nos pauses chocolatées et nos très nombreux fous-rires incontrôlés. Des rapports de TP au manuscrit de thèse, nous aurons été inséparables. Sans toi cette thèse n'aurait pas été la même, et même si nos chemins vont bientôt nous éloigner, tu resteras mon binôme adoré.

Je voudrais enfin remercier mes parents et ma famille pour le soutien et la confiance qu'ils m'ont toujours accordés.

---

# ABBREVIATIONS

---



2n	espèce diploïde
Ac	acétylation
ACF	ATP-utilizing chromatin assembly and remodeling factor
ADAR	adenosine deaminase that act on RNA
ADN	acide désoxyribonucléique
ADNc	ADN complémentaire
ADP	adénosine di-phosphate
Air	antisens du gène Igf2r
Amp	ampicilline
ARN	acide ribonucléique
ARNi	voie d'interférence à ARN
ARNm	ARN messenger
ARNnc	ARN non-codant
ARNr	ARN ribosomique
ARNt	ARN de transfert
ATP	adénosine tri-phosphate
ATRX	alpha thalassemia/mental retardation X-linked
BAC	bacterial artificial chromosome
BF	blocking factor
BRCA1	breast cancer type 1 susceptibility protein
Brms11	breast cancer metastasis-suppressor 1
BSA	bovine serum albumine
CAF	facteur d'assemblage de la chromatine
CBP/p300	protéine de liaison à CREB, enzyme d'acétylation des histones
ccdB	control cell death B
Cdx4	caudal X-linked gene 4
CE	counting element
CES	Chromatin Entry Site
CF	competence factor
Chic1	cysteine rich hydrophobic 1
ChIP	immunoprécipitation de chromatine
CIP	calf intestine phosphatase
CMCT	1-cyclohexyl-3-(2-morpholinoéthyl) carbodiimide méthyl-p-toluène
Cnbp2	cellular nucleic acid binding protein 2
CpG	dinucléotide cytosine-guanine
CTCF	facteur de transcription fixant les motifs CCCTC
db/sb	double brin/simple brin
DCC	dosage compensation complex
Dicer	enzyme d'hydrolyse des ARNs double-brin
DIMT1	dimethyladenosine transferase 1-like
DMEM	dulbecco's modified eagle medium
DMS	diméthyl sulfate
DNase	enzyme d'hydrolyse de l'ADN
Dnmt	enzyme de méthylation des CpG
dox	dependent on X
dpy	dumpy
DTT	dithiothréitol

DXPas34	mini-satellite constitué d'un motif de 34 paires de bases répété,
EC	cône ectoplacentaire
EDTA	éthylène diamine tétraacétate
EE	ectoderme extraembryonnaire
Eed	protéine de développement de l'ectoderme embryonnaire
Enox	expressed neighbour of <i>Xist</i>
EPI	épiblaste
ES	cellules embryonnaires souches
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
Fig.	figure
FISH	Fluorescence in-situ hybridization
FRET	Förster Resonance Energy Transfer
Ftx	five prime to <i>Xist</i>
GFP	green fluorescent protein
GGT	gamma-glutamyltranspeptidase 1
H3K4	lysine 4 de l'histone H3
H3K9	lysine 9 de l'histone H3
H3K14	lysine 14 de l'histone H3
H3K27	lysine 27 de l'histone H3
H4K20	lysine 20 de l'histone H4
H4R3	arginine 3 de l'histone H8
HAT	enzyme d'acétylation des histones
HDAC	enzyme de désacétylation des histones
HDM	histone déméthylase
HEK	human embryonic kidney
HeLa	Henrietta Lacks
hES	cellules embryonnaires humaine
HFD	histone-fold domain
HIV	Human Immunodeficiency Virus
HMT	histone méthyl transférase
hnRNP	heterogeneous nuclear ribonucleoproteins
Hox	homeobox
HP1	protéine d'hétérochromatine 1
hprt	hypoxanthine-guanine phosphoribosyltransférase
ICM	masse cellulaire interne du blastocyste
ICR	Imprinting Control Region
Igf2	facteur de croissance dépendant de l'insuline
Igf2r	recepteur de Igf2
IRES	Internal Ribosome Entry Site
ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
ISWI	"imitation of SWI" complexe de remodelage ATP-dépendant
JIL	chromosomal serine/threonine-protein kinase JIL-1
kb	kilobase
Kcnq1	gène codant un canal au potassium voltage dépendant
Kcnq1ot1	antisens de Kcnq1ot1

kDa	kilodalton
KH	K Homology
KSRP	KH type Splicing Regulatory Protein
LCR	région de contrôle du locus
LIF	leukemia Inhibitory Factor
lincRNA	large intervening non-coding RNA
LINE	long interspersed nuclear element
LTR	long-terminal repeat
Luc	luciférase
Mb	mégabase
MBD	protéine à domaine de liaison aux CpG méthylés
MBP	maltose binding protein
me	méthyl
MeCP2	protéine de liaison au CpG méthylés
MEF	fibroblaste embryonnaire de souris
MEST	mesoderm specific transcript
Met	méthylation
MEX	motif enriched on X
miRNA	micro ARN
mix	mitosis and X associated
MLE	maleless
MOF	males absent on the first
MSL	male specific lethal
MT	marqueur de taille
NARs	Nucleoporin-Associated Regions
NCL	nucléoline
NLS	nuclear localization signal
nm	nanomètre
NMD	nonsense mediated mRNA decay
Nop56	nucleolar protein 56
nt	nucléotides
P/CAF	facteur associé à CBP et p300
PARN	poly(A)-ribonucléase
PARP-1	poly [ADP-ribose] polymerase 1
pb	paire de bases
PBS	Phosphate Buffer Saline (tampon phosphate)
Pc	polycomb
PcG	polycomb group
PCNA	proliferating cell nuclear antigen
PCR	réaction en chaîne de la polymérase
pcv	packed cell volume
PE	endoderme primitif
PGC	primordial germ cells
pH	potentiel hydrogène
PHD	Plant Homeo Domain
PNK	polynucléotide kinase
pnv	packed nuclear volume

PPTase	phosphopantetheinyl transferase
PRC1	complexe polycomb 1
PRC2	complexe polycomb 2
PRE	polycomb response element
PTB	polypyrimidine tract-binding protein
RbAp46-48	retinoblastoma-binding protein p46-p48
rex	recruitment element on X
RGG	boîte riche en arginines et glycines
Rif1	Rap1-interacting factor 1 homolog
Ring	ring-finger protein
RIP	RNA immuno-precipitation
RISC	RNA induced silencing complex
RITS	RNA induced transcriptional silencing
RMN	résonance magnétique nucléaire
RNase	ribonucléase
RNasin	inhibiteur de ribonucléases
RNP	riboNucléoParticule/complexe ribonucléoprotéique
roX	RNA on the X chromosome
RRM	RNA recognition motif
RSC	Remodel the Structure of Chromatin
Rsx	RNA-on-the-silent X
RT	transcription inverse
SAF-A	Scaffold-attachment factor A
Sam68	src-associated in mitosis 68 kDa protein
SCC	Sister Chromatid Cohesion
SDC	sex determination and dosage compensation defect
SDS	sodium dodécylsulfate
SELEX	Systematic Evolution of Ligands by EXponential enrichment
SET	domaine catalytique de méthylation des histones retrouvé dans les protéines
SF	Splicing Factor
Sin3A	SWI-independent 3A
siRNA	small interfering RNA
SMC	Structural maintenance of chromosomes
SMCA	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A
smcHD1	structural maintenance of chromosomes flexible hinge domain containing 1
SMN	Survival Motor Neuron protein
snoRNA	Small nucleolar RNA (petit ARN nucléolaire)
SV40	Simian Virus 40
SWI/SNF	mating type Switching-Sucrose Non-Fertile
SXL	protein sex-lethal
TADs	topologically associated domains
TBE	tampon Tris-borate-EDTA
TE	trophectoderme
TGS	transcriptional gene silencing
TIF1b	transcription intermediary factor 1-beta
Tris	tris (hydroxyméthyl) aminométhane

TS	cellules souches du trophoctoderme
TSC	transcriptional silencing complex
Tsix	antisens à Xist
Tsx	testis specific X-linked
Ub	ubiquitinylation
UsnRNA	U rich small nuclear RNA (petit ARN nucléaire riche en U)
UsnRNP	U rich Small Ribonucleoprotein Particle (petite particule ribonucléoprotéique)
UTR	Untranslated Region
UV	Ultra-Violet
VIH	virus de l'immunodéficience humaine
Xa	chromosome X actif
XCI	X chromosome inactivation
XEN	cellules souches de l'endoderme extra-embryonnaire
Xi	chromosome X inactif
Xic	centre d'inactivation du chromosome X
Xist	X-inactif specific transcript
Xite	X intergenic transcription element
Xm	chromosome X maternel
XO	cellule femelle ayant perdu un chromosome X
Xp	chromosome X paternel
Xpct	X-linked PEST-containing transporter
Xpr	X-pairing
XX	cellule femelle
XY	cellule mâle
YAC	yeast artificial chromosome
YY1	facteur de transcription Yin-Yang1

Ne sont pas incluses dans ce tableau (a) **les abréviations internationales utilisées pour désigner les acides aminés** (nomenclature à 3 lettres ou à une lettre), (b) **les abréviations internationales des unités de mesure** (mL, mg, s, h, etc.) pour lesquels l'abréviation internationalement acceptée a été gardée.



---

# LISTE DES FIGURES & TABLEAUX

---



## LISTE DES FIGURES

## INTRODUCTION

<b>Figure 1</b> : Structure d'un nucléosome et les différents niveaux de compaction de l'ADN .....	3
<b>Figure 2</b> : Visualisation de la chromatine sous ses formes euchromatine et hétérochromatine	4
<b>Figure 3</b> : Structure schématique des trois ADN-méthyltransférases chez les mammifères.....	6
<b>Figure 4</b> : Mécanismes de méthylation de l'ADN .....	6
<b>Figure 5</b> : Mécanisme par lequel la méthylation de l'ADN et la déacétylation des histones coopèrent pour réprimer la transcription .....	8
<b>Figure 6</b> : Les différentes modifications post-traductionnelles des histones.....	9
<b>Figure 7</b> : Hypothèse du "langage des histones" .....	10
<b>Figure 8</b> : Modifications de l'extrémité N-terminale de l'histone H3 .....	11
<b>Figure 9</b> : Diagramme représentant l'effet de l'ubiquitinylation de H2B sur la méthylation de H3 .....	12
<b>Figure 10</b> : Traduction du code histone.....	14
<b>Figure 11</b> : Modèle d'action de SWI/SNF.....	15
<b>Figure 12</b> : Effet de l'action de SWI/SNF sur l'accès des facteurs de fixation à l'ADN.....	16
<b>Figure 13</b> : Représentation schématique de variants d'histones .....	18
<b>Figure 14</b> : Régulation de clusters à empreinte par l'intermédiaire de longs ARN non-codants .....	20
<b>Figure 15</b> : Modèle de silence transcriptionnel dirigé par des petits ARN antisens .....	22
<b>Figure 16</b> : Modèles d'action des longs ARN non-codants en <i>cis</i> versus en <i>trans</i> .....	24
<b>Figure 17</b> : Différentes stratégies pour la compensation de dosage .....	26
<b>Figure 18</b> : Modèle à deux étapes du ciblage du complexe MSL dans la compensation de dosage chez la drosophile.....	28
<b>Figure 19</b> : Stratégie de compensation de dosage chez les hermaphrodites <i>C. elegans</i> .....	30
<b>Figure 20</b> : Modèle de recrutement du DCC sur le chromosome X.....	31
<b>Figure 21</b> : Corps de Barr visualisés par immunofluorescence.....	32
<b>Figure 22</b> : Localisation de H3K27me3 sur le chromosome X inactif.....	35
<b>Figure 23</b> : Marques épigénétiques associées au chromosome X inactif .....	36
<b>Figure 24</b> : Les étapes de l'inactivation au cours du développement .....	39
<b>Figure 25</b> : Le centre d'inactivation du chromosome X (XIC) murin .....	40
<b>Figure 26</b> : Modèles de mécanismes pour le comptage et la sélection du chromosome X à inactiver.....	42
<b>Figure 27</b> : Modèle de comptage des chromosomes X faisant intervenir <i>Tsix/Xite</i> et <i>Xpr</i> .....	44
<b>Figure 28</b> : Modèle stochastique de l'inactivation du chromosome X .....	45
<b>Figure 29</b> : Modèle proposé pour la formation du domaine Xist .....	47
<b>Figure 30</b> : Les étapes de la mise en place de l'inactivation aléatoire du chromosome X dans les cellules ES de souris .....	48
<b>Figure 31</b> : Comparaison de l'ARN Xist entre différentes espèces .....	50
<b>Figure 32</b> : Eléments fonctionnels de Xist impliqués dans la localisation et dans la répression transcriptionnelle du Xi .....	52
<b>Figure 33</b> : Localisation des différents promoteurs du gène <i>Xist</i> .....	54
<b>Figure 34</b> : Les acteurs de la régulation de l'expression de Xist.....	55
<b>Figure 35</b> : Mécanismes possibles de la régulation par Tsix de l'expression de <i>Xist</i> par interférence transcriptionnelle.....	56
<b>Figure 36</b> : Localisation des sites de fixation de CTCF et YY1 sur les gènes <i>Tsix</i> .....	57
<b>Figure 37</b> : Modèle de la régulation dynamique et multifactorielle de l'inactivation du chromosome X par Oct4 .....	58

<b>Figure 38</b> : Réseau transcriptionnel couplant les régulateurs de pluripotence à l'inactivation du chromosome X .....	59
<b>Figure 39</b> : Régulation de l'inactivation par RNF12 et Rex1 .....	61
<b>Figure 40</b> : Localisation de la région des A-repeats dans le gène <i>Xist</i> .....	63
<b>Figure 41</b> : Conservation de la région des A-repeats chez les vertébrés .....	64
<b>Figure 42</b> : Analyse de la région des A-repeats .....	67
<b>Figure 43</b> : Modèle de dimérisation des éléments répétés de la région des A-repeats proposé par Duszczuk <i>et al.</i> (2008) .....	67
Figure 44 : Modèle 2D de la région des A-repeats proposé par Duszczuk <i>et al.</i> (2011) .....	68
<b>Figure 45</b> : Rôle de la région des A-repeats dans la relocalisation des gènes à inactiver dans le domaine <i>Xist</i> .....	70
<b>Figure 46</b> : Les étapes de la réaction d'épissage des introns des pré-ARN messagers nucléaires .....	73
<b>Figure 47</b> : Séquences essentielles à la réaction d'épissage dans les introns majeurs des pré-ARNm de vertébrés .....	74
<b>Figure 48</b> : Assemblage séquentiel du spliceosome autour des introns majeurs .....	76
<b>Figure 49</b> : Les différents modes d'épissage alternatif .....	78
<b>Figure 50</b> : Organisation structurale des protéines SR .....	81
<b>Figure 51</b> : Représentation schématique de la protéine ASF/SF2 humaine et de la structure de son domaine RRM1 .....	83
<b>Figure 52</b> : Rôles de la protéine ASF/SF2 dans la sélection des sites d'épissage .....	85
<b>Figure 53</b> : Représentation schématique des cibles de PTB et de la régulation de l'épissage associée .....	88
<b>Figure 54</b> : Modèle structural de PTB et son interaction avec l'ARN .....	88
<b>Figure 55</b> : Modèles structuraux représentant des boucles dans l'ARN provoquées par PTB, exposant des points de branchement ou des exons .....	89

## MATERIEL ET METHODES

<b>Figure 56</b> : Stratégie de purification de complexes formés entre la région des A-repeats et les protéines d'un extrait nucléaire de cellules ES murines .....	101
<b>Figure 57</b> : Schéma représentant la position des oligonucléotides dans la construction pGL3-Arep-ex1/6 .....	110

## RESULTATS

<b>Figure 58</b> : Représentation de la structure d'un élément répété de la région des A-repeats selon A. Wutz .....	113
<b>Figure 59</b> : Modèles de structure de la région des A-repeats .....	114
<b>Figure 60</b> : Principe du FRET .....	115
<b>Figure 61</b> : Recouvrement des spectres d'émission et d'excitation, exemple du couple Cy3-Cy5 .....	116
<b>Figure 62</b> : Courbe d'efficacité de transfert en fonction de la distance entre les fluorophores .....	117
<b>Figure 63</b> : Structures chimiques des fluorophores cyanine3 (Cy3) et cyanine5 (Cy5) utilisés dans cette étude .....	117
<b>Figure 64</b> : Oligonucléotides utilisés pour évaluer la validité des 3 modèles possibles de la région des A-repeats .....	120
<b>Figure 65</b> : Tests de tampon d'hybridation des oligonucléotides à l'ARN .....	122
<b>Figure 66</b> : Tests de purification des complexes ARN-oligonucléotides .....	123
<b>Figure 67</b> : Protocole suivi pour les mesures de FRET .....	126

<b>Figure 68</b> : Normalisation de l'efficacité de transfert par rapport à la quantité d'oligonucléotide-Cy3 hybridé à l'ARN - exemple du couple P7-P6.....	127
<b>Figure 69</b> : Efficacité de transfert d'énergie obtenue pour les différents couples d'oligonucléotides.....	129
<b>Figure 70</b> : Structure de la protéine KSRP .....	164
<b>Figure 71</b> : Représentation des différents fragments de la protéine KSRP clonés dans le plasmide d'expression pET28 .....	166
<b>Figure 72</b> : Production, purification et analyse par dichroïsme circulaire de la structuration de la protéine KH <sub>14</sub> -6His.....	167
<b>Figure 73</b> : Mise en évidence de l'interaction entre la région des A-repeats et la protéine KH1-4.....	169
<b>Figure 74</b> : Identification des séquences de la région des A-repeats murine fixant la protéine KH1-4.....	171
<b>Figure 75</b> : Identification des séquences de la région des A-repeats humaine fixant la protéine KH1-4.....	172
<b>Figure 76</b> : Localisation des zones de la région des A-repeats murine et humaine protégées par la protéine KH1-4.....	173
<b>Figure 77</b> : Compétition entre les protéines KH1-4 et PTB pour la région des A-repeats....	175
<b>Figure 78</b> : Expériences de chromatographie d'affinité MBP/MS2 réalisées avec la région des A-repeats et un extrait nucléaire complétement en protéine KH1-4-6His .....	176
<b>Figure 79</b> : Mise en évidence d'une interaction directe entre la région des A-repeats et la protéine ASF/SF2.....	178
<b>Figure 80</b> : Prédiction de la localisation des sites de fixation de la protéine ASF/SF2 sur la région des A-repeats murine .....	179
<b>Figure 81</b> : Identification des séquences de la région des A-repeats murine fixant la protéine ASF/SF2.....	180
<b>Figure 82</b> : Présence de la protéine Sin3A dans les complexes purifiés à partir de la région des A-repeats.....	182
<b>Figure 83</b> : Mise en évidence de la protéine Rif1 parmi les protéines en complexe avec la région des A-repeats.....	183
<b>Figure 84</b> : Modèle pour le complexe cohésine .....	184
<b>Figure 85</b> : Présence de protéines du complexe cohésine parmi les protéines purifiées avec la région des A-repeats.....	185
<b>Figure 86</b> : Mise en évidence de sites d'épissage cryptiques dans l'ARN Xist (Royce-Tolland <i>et al.</i> , 2010).....	186
<b>Figure 87</b> : Schéma du clonage par Gateway des exons 2 à 6 du gène <i>Xist</i> murin dans le plasmide pSpliceExpress.....	188
<b>Figure 88</b> : Profils d'épissage entre les exons 2 et 6 de l'ARN Xist.....	190
<b>Figure 89</b> : Schéma représentant les étapes de clonages de la région 5' du gène <i>Xist</i> dans le plasmide pGL3 .....	192
<b>Figure 90</b> : Etude de l'épissage de l'ARN C2 sous l'effet des A-repeats .....	194
<b>Figure 91</b> : Variation du taux des transcrits épissés par rapport à celui des transcrits totaux (épissés et non épissés) selon le nombre de répétitions de la région des A-repeats.....	195
<b>Figure 92</b> : Schéma des différentes étapes de clonage du gène <i>Xist</i> murin tronqué dans le plasmide pGL3 .....	197
<b>Figure 93</b> : Variation du taux d'ARN Xist selon le nombre de répétitions de la région des A-repeats.....	198
<b>Figure 94</b> : Quantification du taux de transcrit Xist en fonction du nombre de répétitions de la région des A-repeats.....	199

<b>Figure 95</b> : Variation l'activité de la luciférase Firefly selon le nombre de répétitions de la région des A-repeats.....	200
<b>Figure 96</b> : Estimation du temps de demi-vie de l'ARN Xist selon le nombre de répétitions de la région des A-repeats.....	202
<b>Figure 97</b> : Profil d'épissage de l'ARN Xist selon le nombre de répétitions de la région des A-repeats.....	204
<b>Figure 98</b> : Effet de la surexpression de différentes protéines sur l'épissage entre les exons 1 et 2 de l'ARN Xist.....	206
<b>Figure 99</b> : Effet de la surproduction de différentes protéines sur l'épissage entre les exons 2 et 6 de l'ARN Xist.....	207

## DISCUSSION

<b>Figure 100</b> : Modèle de structuration de l'ARN Xist faisant intervenir la protéine PTB et son implication dans la régulation de l'épissage.....	217
---	-----

## LISTE DES TABLEAUX

### MATERIEL ET METHODES

<b>Tableau 1</b> : Liste de plasmides pour la production <i>in vitro</i> d'ARN.....	93
<b>Tableau 2</b> : Listes des oligonucléotides utilisés pour la construction des plasmides d'expression dans des cellules eucaryotes.....	96
<b>Tableau 3</b> : Liste des anticorps utilisés au cours de ce travail.....	97
<b>Tableau 4</b> : Liste des oligonucléotides utilisés pour les expériences de FRET.....	100
<b>Tableau 5</b> : Liste des oligonucléotides utilisés pour les expériences de RT-PCR.....	110

---

# SOMMAIRE

---



**REMERCIEMENTS****ABREVIATIONS****LISTE DES FIGURES ET TABLEAUX****SOMMAIRE**

<b>INTRODUCTION .....</b>	<b>1</b>
<b>I. La régulation épigénétique et les ARN non-codants.....</b>	<b>1</b>
I.1 La notion d'épigénétique.....	1
I.1.1 Généralités.....	1
I.1.2 Les régulations de l'expression génique par l'état de compaction de la chromatine .....	5
a. La méthylation de l'ADN.....	5
b. Les modifications post-traductionnelles des histones .....	9
c. Les complexes de remodelage modifiant la conformation de la chromatine .....	15
d. Les variants d'histones .....	17
I.2 L'implication des ARN dans la régulation épigénétique.....	18
I.2.1 La régulation par de longs ARN non-codants .....	19
I.2.2 La régulation par de petits ARN non-codants .....	21
I.2.3 Les mécanismes possibles .....	23
<b>II. La compensation du dosage des gènes .....</b>	<b>25</b>
II.1 Les principes généraux.....	25
II.2 La compensation du dosage chez différentes espèces .....	26
II.2.1 Chez <i>Drosophila melanogaster</i> .....	26
II.2.2 Chez <i>Caenorhabditis elegans</i> .....	29
II.2.3 Chez les mammifères .....	32
<b>III. L'inactivation du chromosome X chez les mammifères .....</b>	<b>34</b>
III.1 Les caractéristiques du chromosome X inactif.....	34
III.1.1 Les marques épigénétiques du Xi.....	34
III.1.2 La réplication tardive.....	36
III.1.3 La méthylation de l'ADN.....	37
III.1.4 Les gènes échappant à l'inactivation.....	37
III.2 Les mécanismes de l'inactivation.....	38
III.2.1 Les étapes de l'inactivation au cours de l'embryogénèse chez la souris .....	38
III.2.2 Le centre d'inactivation du chromosome X (XIC).....	39
III.2.3 Les étapes de l'inactivation aléatoire.....	41
a. Le comptage et le choix du chromosome X à inactiver .....	41
b. L'initiation de l'inactivation.....	46
c. La propagation de l'ARN Xist .....	46
d. La mise en place des marques épigénétiques et la maintenance de l'état réprimé .....	47
<b>IV. Le gène <i>Xist</i> et son transcrit.....</b>	<b>49</b>
IV.1 Description du gène <i>Xist</i> .....	49
IV.2 Les fonctions de l'ARN Xist .....	51
IV.3 Les régulations de l'expression du gène <i>Xist</i> .....	52

IV.3.1	La régulation de la localisation et de la demi-vie de l'ARN Xist.....	53
IV.3.2	Les régulations transcriptionnelles de l'expression de <i>Xist</i> .....	55
a.	Les inhibiteurs transcriptionnels de <i>Xist</i> .....	55
b.	Les activateurs transcriptionnels de <i>Xist</i> .....	59
c.	Régulation par la région des A-repeats de l'ARN Xist lui-même.....	62
<b>V.</b>	<b>La région des A-repeats .....</b>	<b>63</b>
V.1	Présentation de la région des A-repeats.....	63
V.1.1	Caractérisation de la région des A-repeats .....	63
V.1.2	Régulation de l'expression des A-repeats .....	65
V.2	Structure de la région des A-repeats.....	65
V.2.1	Prédiction informatique de Wutz <i>et al.</i> .....	65
V.2.2	Modèle proposé par Duszczuk <i>et al.</i> .....	67
V.3	Rôles de la région des A-repeats .....	69
V.3.1	Formation du domaine Xist.....	69
V.3.2	Recrutement du complexe PRC2.....	70
V.3.3	Rôle dans l'épissage de Xist.....	70
<b>VI.</b>	<b>L'épissage .....</b>	<b>71</b>
VI.1	Description générale de l'épissage des ARN .....	72
VI.1.1	Les étapes de la réaction d'épissage.....	72
VI.1.2	Les séquences définissant les introns .....	73
a.	Le site 5' donneur d'épissage.....	74
b.	Le site 3' accepteur d'épissage.....	75
VI.1.3	Les composants et l'assemblage du spliceosome .....	75
VI.2	L'épissage alternatif .....	77
VI.2.1	Les différents types d'épissage alternatif .....	77
VI.2.2	La régulation de l'épissage alternatif.....	79
VI.3	Les régulateurs de l'épissage constitutif et alternatif .....	80
VI.3.1	Les protéines SR.....	80
a.	Les caractéristiques communes des protéines SR .....	80
b.	Les rôles des protéines SR dans l'épissage .....	82
c.	La protéine ASF/SF2 (SRSF1).....	83
VI.3.2	Les protéines hnRNP.....	85
a.	Les caractéristiques communes des protéines hnRNP .....	86
b.	Les rôles des protéines hnRNP dans l'épissage.....	86
c.	La protéine PTB (hnRNP I).....	87
	<b>OBJECTIFS DE LA THESE .....</b>	<b>91</b>
	<b>MATERIEL ET METHODES .....</b>	<b>93</b>
<b>I.</b>	<b>Matériel.....</b>	<b>93</b>
I.1	Souches et plasmides utilisés.....	93
I.1.1	Les souches d' <i>E. Coli</i> .....	93
I.1.2	Les lignées de cellules de mammifères en culture .....	93
I.1.3	Les plasmides utilisés.....	93
a.	Plasmides recombinants utilisés pour la production <i>in vitro</i> d'ARN .....	93

b.	Vecteurs utilisés pour la production de protéines recombinantes .....	94
c.	Plasmides utilisés pour la transfection et l'expression dans des cellules eucaryotes .....	94
I.2	Milieux de culture, tampons et solutions d'usage courant .....	96
I.3	Anticorps .....	97
<b>II.</b>	<b>Méthodes .....</b>	<b>98</b>
II.1	Méthode utilisée pour l'analyse de la structure secondaire d'un ARN par la technique de FRET .....	98
II.1.1	Préparation des ARN.....	98
II.1.2	Technique de FRET permettant d'apprécier la distance relative entre 2 zones ARN d'intérêt .....	98
II.2	Purification des complexes RNP par chromatographie d'affinité MS2/MBP.....	100
II.3	Production et purification de la protéine KH1-4-6His chez <i>E. Coli</i> .....	102
II.3.1	Transformation et culture bactérienne.....	102
II.3.2	Préparation des extraits cellulaires .....	102
II.3.3	Purification de la protéine KH1-4-6His .....	102
II.3.4	Estimation de la quantité et de la qualité de l'échantillon protéique.....	102
II.4	Analyses des interactions ARN/protéines in vitro.....	103
II.4.1	Analyses des complexes ARN/Protéines par expérience de retard sur gel .....	103
a.	Préparation des ARN.....	103
b.	Formation des complexes ARN/protéines.....	103
II.4.2	Analyses des complexes ARN/Protéines par expérience d'empreinte aux RNases ....	104
a.	Préparation des ARN.....	104
b.	Analyse par extension d'amorce .....	104
II.4.3	Analyses des complexes ARN/Protéines par expérience de pontage covalent aux UV ....	105
a.	Préparation des ARN.....	105
b.	Préparation des puits .....	106
c.	Préparation du mélange réactionnel .....	106
d.	Pontage covalent aux UV entre les protéines et les ARN .....	106
II.5	Méthodes utilisées pour la culture cellulaire .....	107
II.5.1	Culture cellulaire .....	107
a.	Lignées HEK293 et HeLa .....	107
b.	Lignée cellulaire ES XX de type HP3-10.....	107
II.5.2	Entretien des cellules.....	107
II.5.3	Préparation d'extrait nucléaire (adapté d'après Dignam, 1983).....	108
II.5.4	Analyse d'épissage <i>in cellulo</i> .....	108
a.	Transfection.....	108
b.	Analyse par RT-PCR.....	109
II.5.5	Mesure de l'activité luciférase.....	111
II.5.6	Analyse du temps de demi-vie des ARN.....	111
II.5.7	Quantification et significativité .....	111

<b>RESULTATS .....</b>	<b>113</b>
<b>I. Analyse de la structure 2D de la région des A-repeats par FRET .....</b>	<b>113</b>
I.1 Introduction .....	113
I.2 Présentation de la technique de FRET .....	115
I.3 Mise au point du protocole expérimental .....	117
I.3.1 Choix des fluorophores .....	117
I.3.2 Choix des oligonucléotides.....	118
I.3.3 Hybridation ARN-oligonucléotides.....	121
a. La concentration en ARN .....	121
b. Le tampon et le temps d'hybridation.....	121
c. La concentration en oligonucléotides et la purification des complexes .....	122
I.3.4 Paramètres de mesure .....	124
I.3.5 Protocole utilisé.....	124
I.4 Résultats obtenus.....	128
<b>II. « 2D structure of the A region of Xist RNA and its implication for PRC2 association » .....</b>	<b>131</b>
<b>III. Etude de l'interaction de la région des A-repeats avec des facteurs protéiques. 163</b>	
III.1 Stratégie de purification et d'identification des complexes ribonucléoprotéiques formés avec la région des A-repeats .....	163
III.2 Analyse biochimique de l'interaction avec la protéine KSRP.....	163
III.2.1 Clonage et production de la protéine KH1-4.....	165
III.2.2 Expériences de retard sur gel.....	168
III.2.3 Etude par empreinte aux RNases T1 et T2 des sites de fixation de la protéine KH1-4 sur la région des A-repeats .....	170
III.2.4 Etude de la compétition KSRP/PTB par pontage aux UV .....	174
III.3 Autres protéines partenaires de la région des A-repeats.....	177
III.3.1 ASF/SF2 .....	177
a. Etude de l'interaction par expérience de retard sur gel .....	177
b. Prédiction informatique des sites de fixation d'ASF/SF2 .....	178
c. Analyse des sites de fixation d'ASF/SF2 sur la région des A-repeats par expériences d'empreintes .....	179
III.3.2 Sin3A.....	181
III.3.3 Rif1 .....	182
III.3.4 Pds5a .....	184
<b>IV. Analyse du rôle de la région des A-repeats dans les processus de transcription et maturation de l'ARN Xist.....</b>	<b>186</b>
IV.1 Etude de l'épissage de l'ARN Xist <i>in cellulo</i> .....	186
IV.1.1 Clonage des exons 2 à 6 du gène Xist dans le plasmide pSpliceExpress .....	186
IV.1.2 Analyse des produits d'épissage entre les exons 2 et 6 en absence de la région des A-repeats .....	188
IV.2 Etude de l'épissage de l'ARN C2 sous le contrôle de la région des A-repeats .....	191
IV.2.1 Construction du plasmide pGL3-Arep-C2 .....	191

IV.2.2	Etude de l'épissage de C2 par RT-PCR.....	192
IV.3	Etude du rôle de la région des A-repeats dans l'épissage de l'ARN Xist .....	196
IV.3.1	Clonage du gène <i>Xist</i> murin dans le plasmide pGL3.....	196
IV.3.2	Analyse de l'épissage entre les exons 1 et 6 de l'ARN Xist .....	198
a.	Analyse du taux de transcrits Xist.....	198
b.	Analyse du temps de demi-vie de l'ARN Xist .....	201
c.	Analyse de l'épissage de l'ARN Xist.....	203
IV.4	Analyse du rôle de différentes protéines sur l'épissage de l'ARN Xist .....	205
	<b>DISCUSSION ET PERSPECTIVES.....</b>	<b>211</b>
	<b>ANNEXES .....</b>	<b>221</b>
	<b>BIBLIOGRAPHIE .....</b>	<b>225</b>



---

# INTRODUCTION

---



# Introduction

---

## I. La régulation épigénétique et les ARN non-codants

### I.1 La notion d'épigénétique

#### I.1.1 Généralités

« Si les caractères de l'individu sont déterminés par les gènes, pourquoi toutes les cellules d'un organisme ne sont-elles pas identiques ? ». Tout le mystère qui a longtemps entouré l'épigénétique se trouve finalement assez bien résumé dans cette question de Thomas Morgan datant de 1940. Aujourd'hui, on définit l'épigénétique comme le domaine qui étudie comment l'environnement et l'histoire individuelle influent sur l'expression des gènes, et plus précisément l'ensemble des modifications transmissibles d'une génération à l'autre et réversibles de l'expression génique sans altération des séquences nucléotidiques.

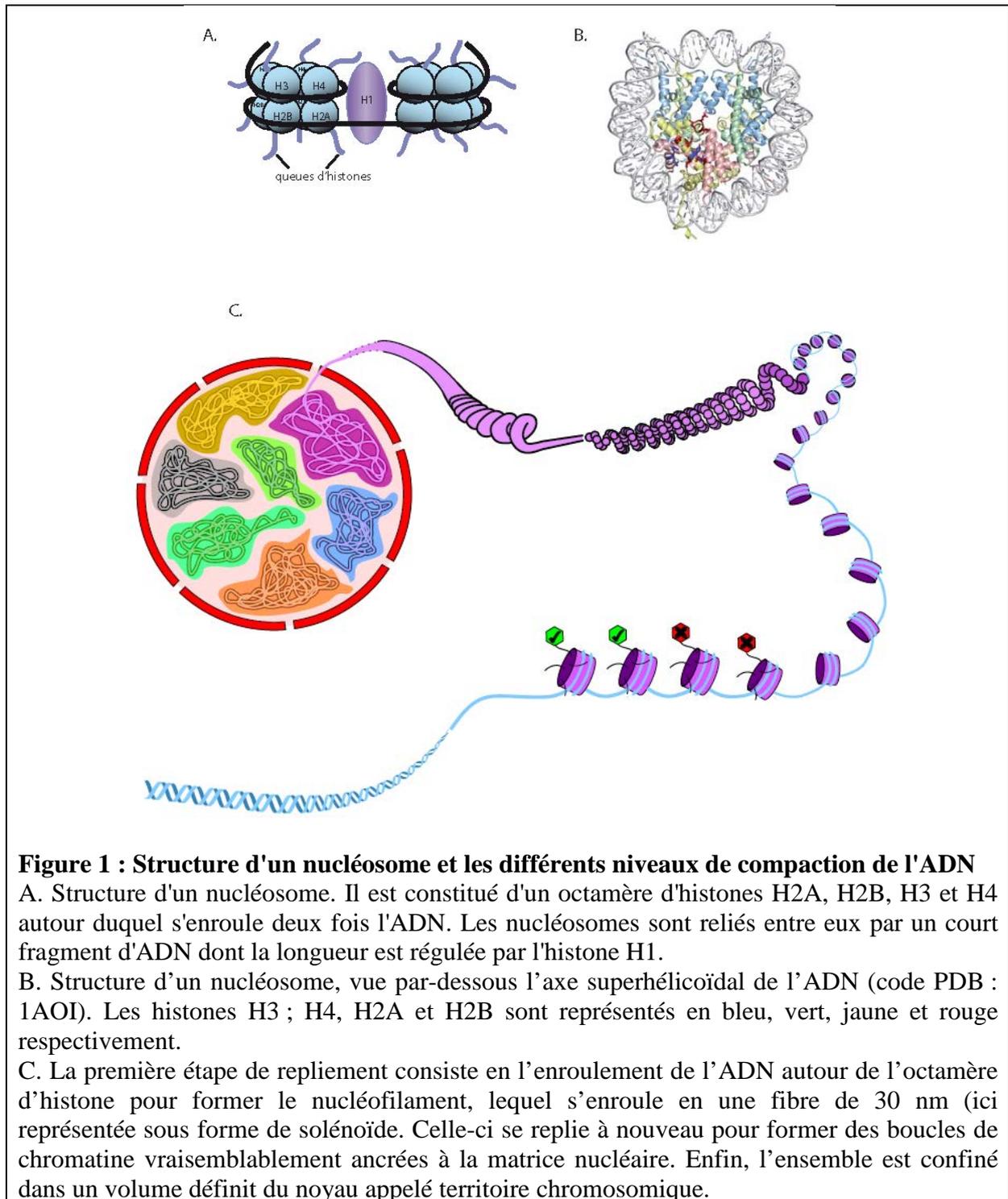
On attribue la paternité de l'épigénétique dans son sens moderne au biologiste Conrad H. Waddington qui la définit en 1942 comme « la branche de la biologie étudiant les implications entre les systèmes [gènes + environnement] et leurs produits donnant naissance au phénotype d'un individu ». Il apparut d'abord que le « modèle génétique » postulant une équivalence unique entre phénotype et génotype, ne pouvait expliquer tous les phénomènes liés à la différenciation cellulaire (cf. citation de T. Morgan plus haut). Par exemple, le noyau d'une cellule de peau d'amphibien transféré dans un œuf énucléé donne des animaux entiers. Un même génome peut donc avoir plusieurs destinées et sa détermination est modulable. Il fut alors élaboré une théorie selon laquelle chaque cellule indifférenciée passait par un état critique responsable de son développement futur qui n'est pas uniquement lié à ses gènes (et pour cette raison qualifié d'*épigénétique*). Avec la découverte de la double-hélice en 1953, cette théorie a été mise à l'écart jusque dans les années 90, durant lesquelles le séquençage complet de plusieurs génomes et l'incapacité de comprendre leur expression remirent sur le devant de la scène l'épigénétique.

Pour bien comprendre comment des modifications épigénétiques peuvent influencer l'expression d'un gène, un bref rappel sur la structure de l'ADN et sur les différents niveaux de compaction de la chromatine est nécessaire.

Dans les cellules eucaryotes, le matériel génétique est localisé dans le noyau et il est organisé en une structure complexe constituée d'ADN et de protéines, la chromatine. Environ deux

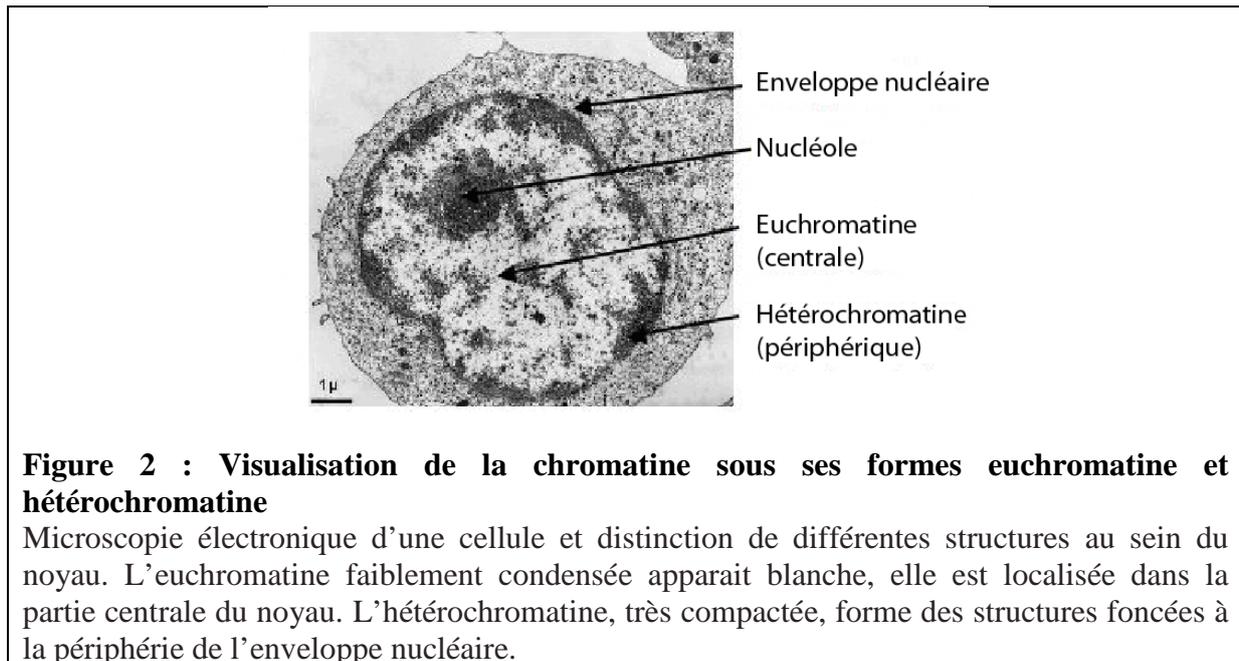
mètres d'ADN dans chaque cellule doivent être contenus dans un noyau de quelques microns de diamètre et d'un volume d'environ  $50 \mu\text{m}^3$ . En plus de cet énorme degré de compaction (400 000 fois), l'ADN doit être rapidement accessible afin de permettre son interaction avec les machineries régulant les fonctions de la chromatine (la réplication, la transcription, la réparation et la recombinaison). L'organisation dynamique de la structure chromatiniennne influence, potentiellement, toutes les fonctions du génome.

L'unité fondamentale de la chromatine est appelée le nucléosome qui est composé d'ADN et d'histones. Il est constitué d'un cœur de 8 protéines histones (deux exemplaires de H2A, H2B, H3 et H4) autour duquel s'enroulent environ 147 paires de bases d'ADN sur un tour trois quart (**Figure 1.A**). Il constitue le premier niveau de compaction de l'ADN puisque les 147 paires de bases d'ADN entourées autour du cœur de la structure sont six fois plus compactes qu'un fragment d'ADN nu de même longueur. L'enchaînement des nucléosomes forme le nucléofilament (ou fibre de 11 nm, ou encore « collier de perles »), puis les nucléosomes s'empilent pour former la fibre de 30 nm. Le taux de compaction augmente jusqu'à former le chromosome (**Figure 1.B**), le niveau le plus élevé étant celui du chromosome métaphasique (10 000 fois pour chaque chromatide).



Au sein du noyau interphasique, la chromatine est organisée en territoires fonctionnels. En fonction de caractères morphologiques (en microscopie électronique notamment) et de propriétés fonctionnelles, la chromatine a été divisée en (**Figure 2**) :

- euchromatine
- hétérochromatine



**Figure 2 : Visualisation de la chromatine sous ses formes euchromatine et hétérochromatine**

Microscopie électronique d'une cellule et distinction de différentes structures au sein du noyau. L'euchromatine faiblement condensée apparaît blanche, elle est localisée dans la partie centrale du noyau. L'hétérochromatine, très compactée, forme des structures foncées à la périphérie de l'enveloppe nucléaire.

L'hétérochromatine a été définie comme une structure qui ne change pas d'état de condensation au cours du cycle cellulaire, tandis que l'euchromatine apparaît décondensée pendant l'interphase. L'hétérochromatine est localisée principalement en périphérie du noyau et du nucléole, tandis que l'euchromatine est répartie à l'intérieur du nucléoplasme.

On distingue :

- L'hétérochromatine constitutive qui contient peu de gènes, formée principalement de séquences répétées et dont les plus grandes régions sont situées à proximité des centromères et des télomères des chromosomes.
- L'hétérochromatine conditionnelle qui contient des régions codantes pouvant adopter les caractéristiques structurales et fonctionnelles de l'hétérochromatine, comme le chromosome X inactif chez la femelle des mammifères.

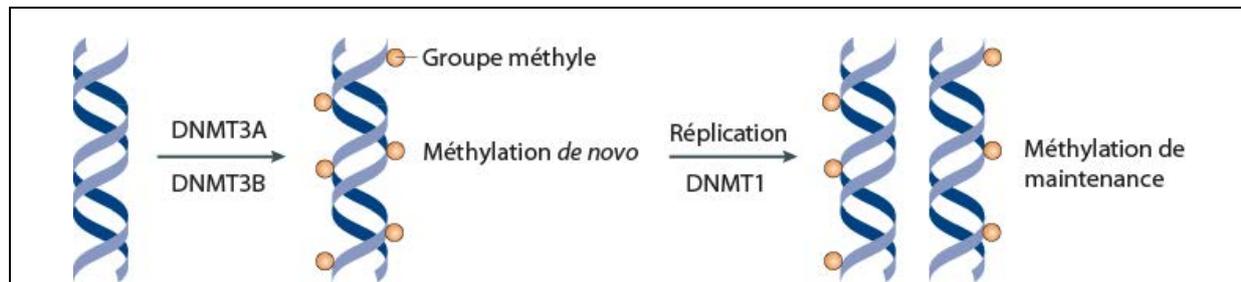
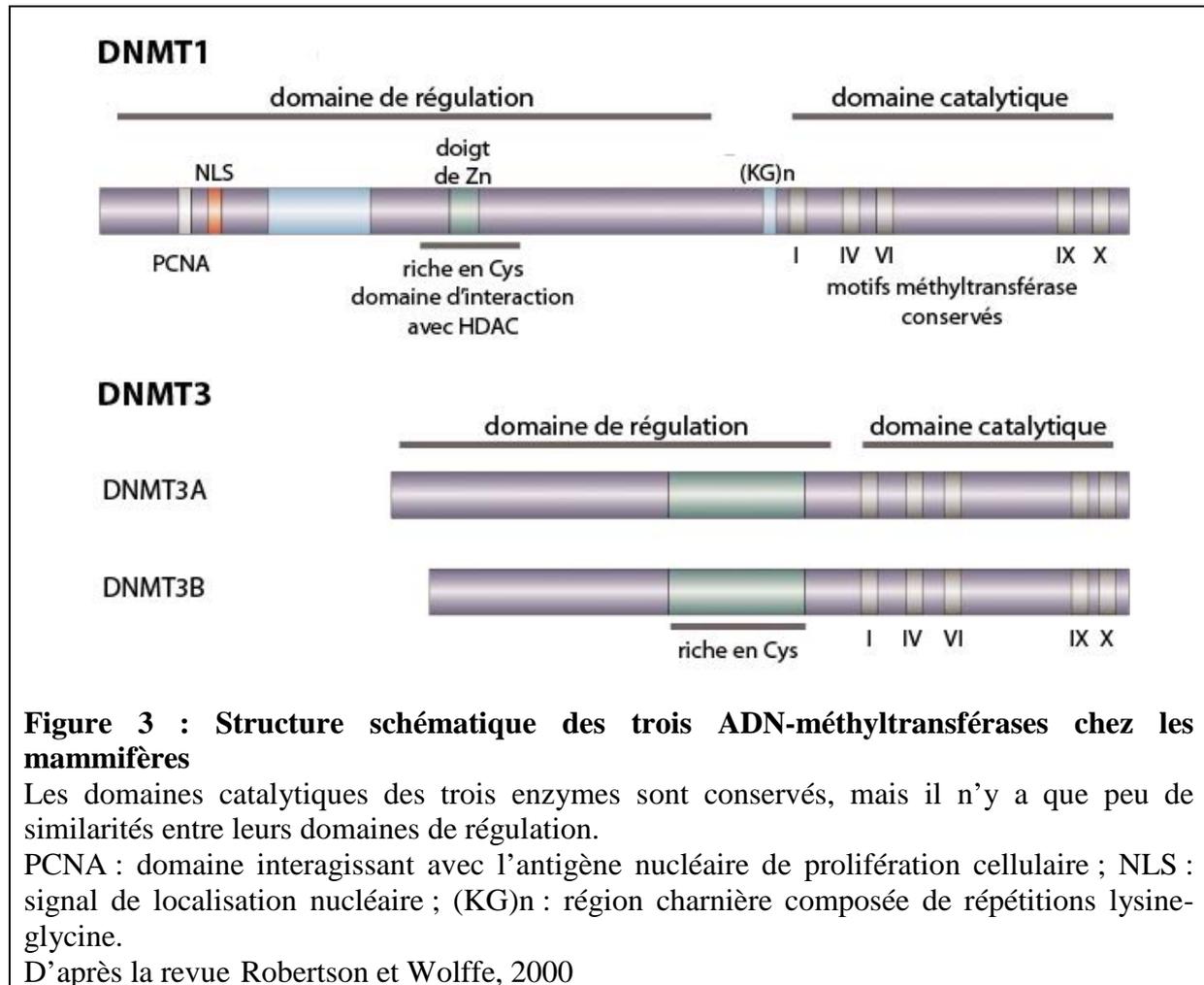
L'euchromatine est généralement associée à une expression active des gènes qu'elle regroupe. L'hétérochromatine, quant à elle, du fait de son état condensé, est associée au silence transcriptionnel. Néanmoins, l'état de compaction d'une région n'est pas irréversible, il est dynamique. En effet, une région condensée peut être décondensée dans le but de la rendre accessible à la machinerie de transcription. La modulation de la compaction chromatinienne permet ainsi de réguler l'expression génique. Cette régulation peut s'effectuer : (i) par l'apposition de marques épigénétiques qui influencent l'accessibilité de l'ADN aux diverses

machineries nucléaires, (ii) par des modifications post-traductionnelles des histones qui ont un impact épigénétique et (iii) par des complexes modifiant la conformation de la chromatine.

### I.1.2 Les régulations de l'expression génique par l'état de compaction de la chromatine

#### a. La méthylation de l'ADN

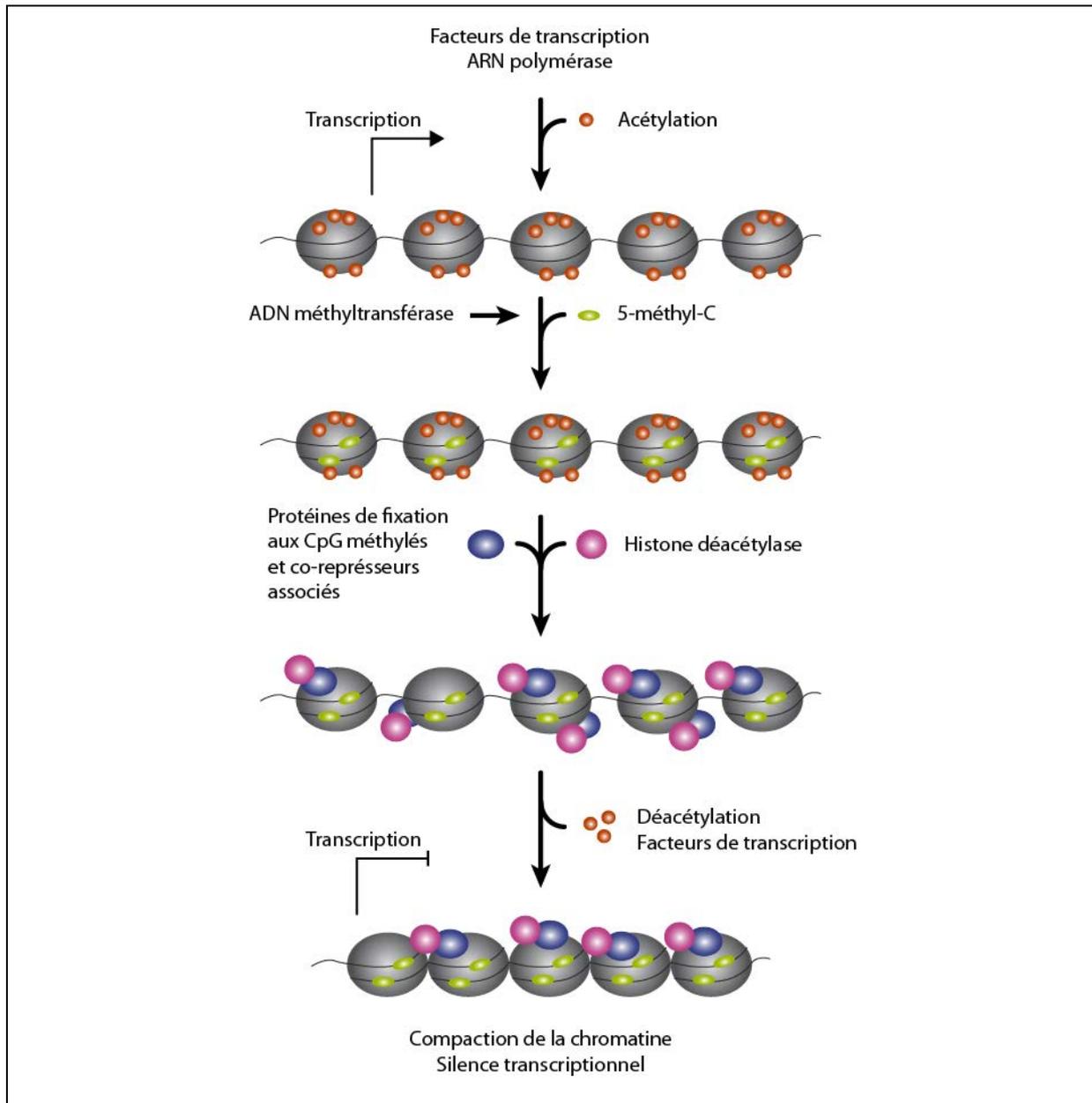
Dans les cellules mammifères, la méthylation de l'ADN s'opère principalement sur la position C5 des dinucléotides CpG. Les enzymes responsables de cette modification, les ADN méthyltransférases (DNMTs), sont bien caractérisées et sont conservées chez les mammifères et les plantes (Law et Jacobsen, 2010). On peut diviser ces enzymes en deux catégories d'activité : la maintenance de la méthylation et la méthylation *de novo* (Goll et Bestor, 2005). Les profils de méthylation d'ADN sont initialement établis par les méthyltransférases *de novo* DNMT3A et DNMT3B (**Figure 3**). Outre lors de la gamétogénèse où la méthylation de l'ADN joue un rôle capital dans l'établissement de l'empreinte génétique, ces enzymes interviennent au cours du stade blastocyste afin d'établir un nouveau profil de méthylation après l'effacement des marques héritées des gamètes (Okano *et al.*, 1998 ; Okano *et al.*, 1999). Ces marques sont ensuite fidèlement maintenues au cours des divisions cellulaires sous l'action de la méthyltransférase de maintenance, DNMT1 (**Figure 3**), qui a une préférence pour l'ADN hémiméthylé (Bestor et Ingram, 1983 ; Bestor *et al.*, 1988 ; Hermann *et al.*, 2004) (**Figure 4**).



La relation entre la méthylation des îlots CpG et la répression transcriptionnelle chez les vertébrés est reconnue depuis plus de 30 ans. En 1992, MECP2 (methyl-CpG binding protein 2), un répresseur transcriptionnel reconnaissant sélectivement l’ADN méthylé, a été

caractérisé (Lewis *et al.*, 1992). Cette protéine peut être divisée en deux domaines structuraux : un domaine de fixation aux CpG méthylés (MBD) qui reconnaît les dinucléotides CpG symétriquement méthylés par des contacts avec le grand sillon de la double hélice (Wakefield *et al.*, 1999), et un domaine de répression transcriptionnelle (TRD) qui interagit avec plusieurs autres protéines régulatrices (Nan *et al.*, 1997). MECP2 réprime sélectivement la transcription de matrices méthylées en absence de structure chromatinienne organisée (Nan *et al.*, 1997 ; Kaludov et Wolffe, 2000). La répression transcriptionnelle observée *in vitro* pourrait donc être le résultat de l'agrégat de MECP2 sur la matrice, empêchant la formation des complexes de transcription fonctionnels (Kaludov et Wolffe, 2000).

MECP2 peut également reconnaître des CpG méthylés présents au niveau des nucléosomes (Chandler *et al.*, 1999). Dans certaines circonstances, elle peut même déplacer l'histone H1 de la chromatine (Nan *et al.*, 1997). MECP2 est donc un facteur qui peut modifier la structure de la chromatine directement ou par le recrutement d'autres enzymes répressives. En effet, les protéines de fixation à l'ADN méthylé, dont MECP2, peuvent s'associer à des complexes co-répresseurs tels que les histone déacétylases (Jones *et al.*, 1998 ; Nan *et al.*, 1998 ; Ng *et al.*, 1999 ; Wade *et al.*, 1999). Le recrutement d'une histone déacétylase par MECP2 s'effectue par l'intermédiaire de la protéine adaptatrice Sin3A. Le résultat en est un silence transcriptionnel, en partie dû à la déacétylation des histones (Jones *et al.*, 1998 ; Nan *et al.*, 1998). Cette interaction conduit à la formation de structures de chromatine réprimée de manière stable (Tse *et al.*, 1998). La **Figure 5** présente un modèle de coopération entre la méthylation de l'ADN et la déacétylation des histones pour réprimer la transcription.



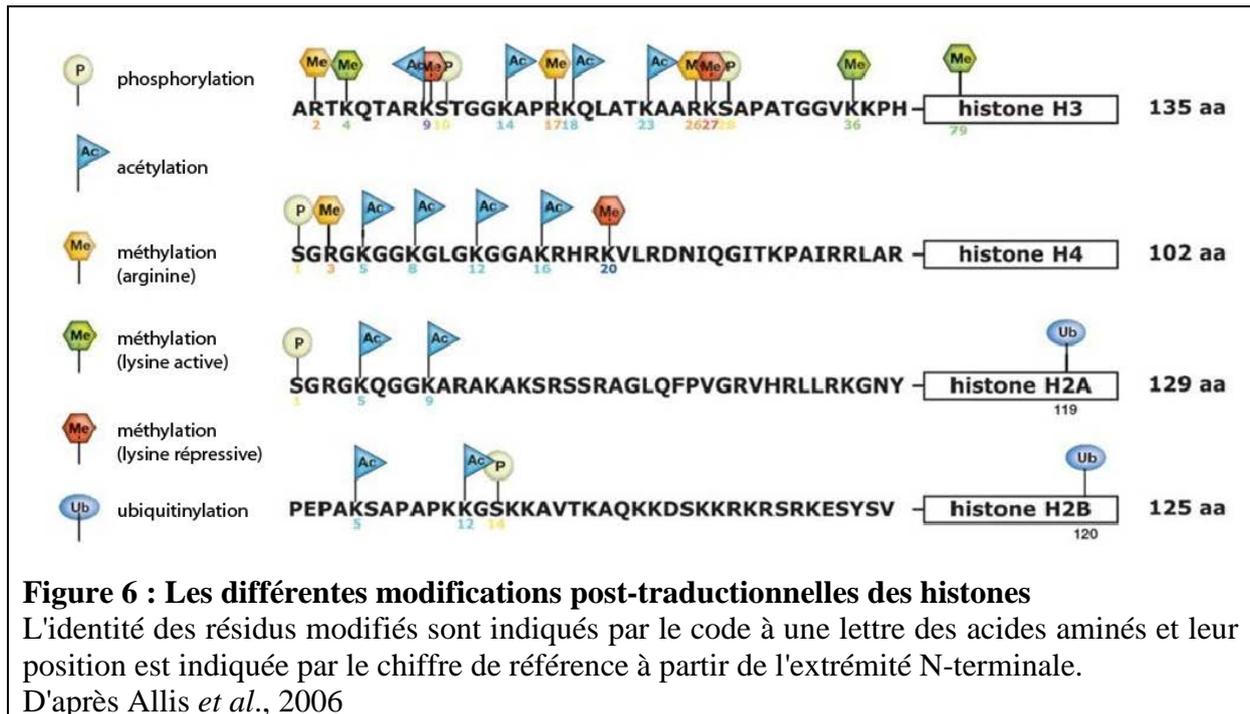
**Figure 5 : Mécanisme par lequel la méthylation de l'ADN et la déacétylation des histones coopèrent pour réprimer la transcription**

Il semblerait qu'une région transcriptionnellement active ciblée pour être réprimée est tout d'abord méthylée sur son ADN, ce qui permet le recrutement des protéines de fixation aux CpG méthylés ainsi que les co-répresseurs qui leur sont associés et des histones déacétylases (HDAC). Puisque l'ADN méthyltransférase 1 (DNMT1) peut interagir directement avec les histones déacétylases, il est également possible que la transcription soit d'abord réprimée par déacétylation, après quoi la machinerie de méthylation et les protéines de fixation aux CpG méthylés seraient recrutés pour « cimenter » le promoteur dans un état silencieux. Dans tous les cas, les nucléosomes déacétylés adoptent une structure plus condensée qui inhibe l'accès des facteurs de transcription à leurs sites de fixation.

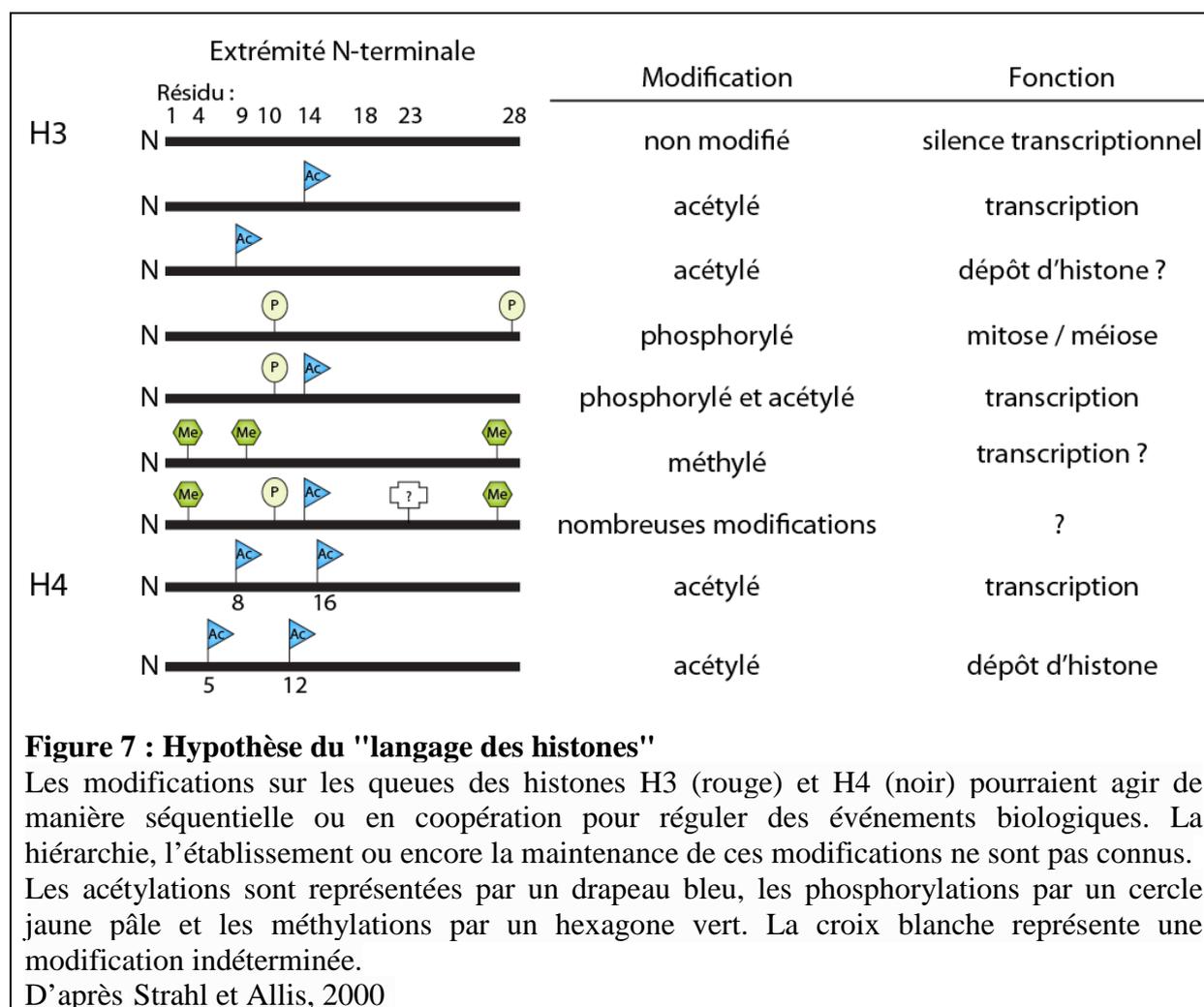
D'après la revue Robertson et Wolffe, 2000

### b. Les modifications post-traductionnelles des histones

L'état de condensation de la chromatine peut également être régulé par un autre type d'activité. Des modifications post-traductionnelles en N-terminal des queues d'histones telles que l'acétylation, la phosphorylation, la méthylation, la biotinylation, l'ubiquitination et l'ADP-ribosylation peuvent être ajoutées selon différentes combinaisons (**Figure 6**), laissant un signal pouvant être reconnu par différents complexes modifiant la chromatine et ce, via des domaines protéiques spécifiques (bromodomaines, chromodomaines).



La fonction de chacune de ces modifications n'est pas toujours identifiée. Quelques marques sont connues pour avoir un effet répresseur ou activateur, mais la compréhension de ce « langage » se complique avec la possibilité de former de nombreuses combinaisons différentes. La **Figure 7** présente quelques modifications, ainsi que leur fonction.



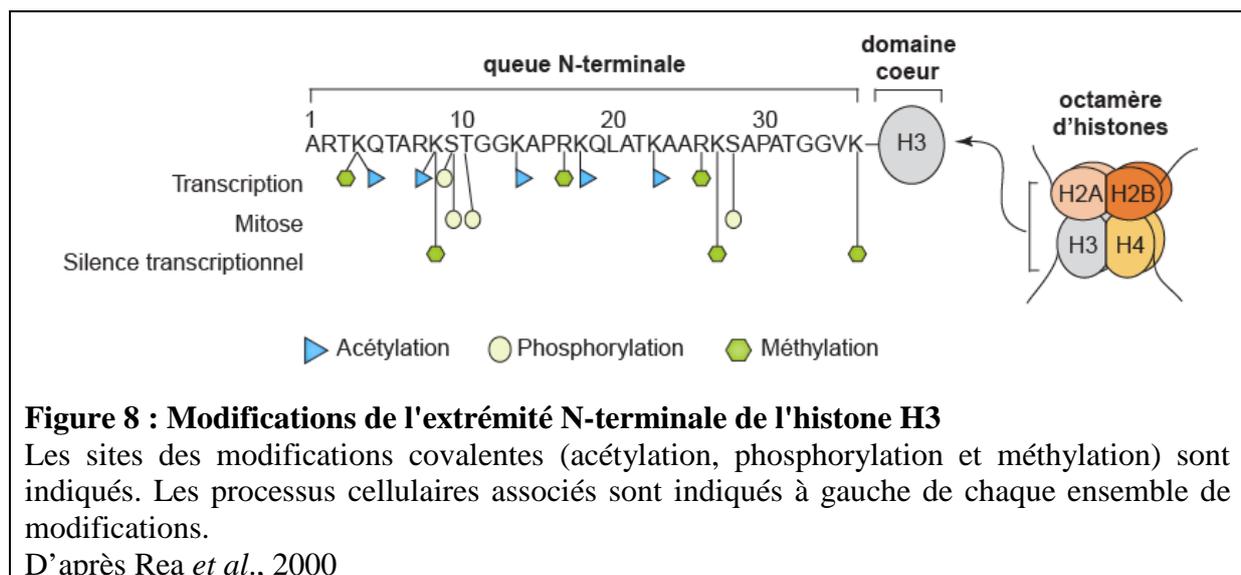
Le « code des histones » (Jenuwein et Allis, 2001) constitue donc la clé d'un mode de régulation génique extrêmement sophistiqué et dynamique, lui permettant de s'adapter rapidement aux différents changements cellulaires.

- La modification post-traductionnelle la mieux caractérisée jusqu'à aujourd'hui est sans aucun doute l'**acétylation**, une modification découlant du transfert d'un groupement acétyl à une lysine localisée au niveau de l'extrémité N-terminale d'une histone. L'ajout d'un tel groupement laisse non seulement une marque épigénétique sur les histones, mais permet également la neutralisation de la charge de ces lysines conduisant, entre autres, à la diminution des interactions histones/ADN et ainsi au relâchement du niveau de compaction de la chromatine. L'état d'acétylation des histones résulte d'un équilibre entre deux types d'activités opposées : l'activité histone acétyltransférase (HAT) et l'activité histone déacétylase (HDAC).

- La **méthylation**, quant à elle, peut s'effectuer soit sur des lysines soit sur des arginines et peut constituer en l'ajout d'un, de deux ou de trois groupements méthyles.

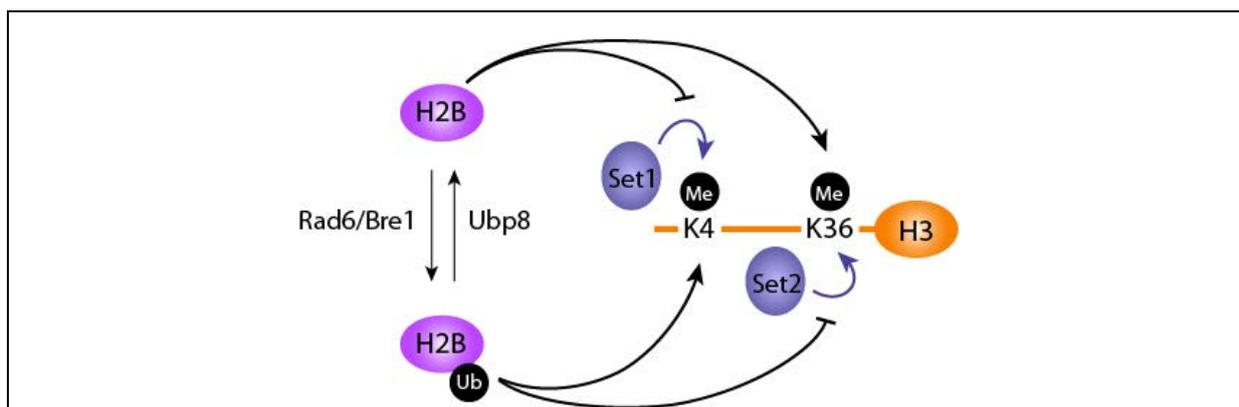
Selon les résidus méthylés et le nombre de groupements ajoutés, elle est associée à une activation ou une répression de la transcription. Longtemps considérée comme statique, la méthylation des histones s'avère être une modification réversible impliquée dans un processus dynamique, bien que plus stable que l'acétylation et la phosphorylation. Un nombre croissant d'histones déméthylases est ainsi identifié (Klose et Zhang, 2007).

- La **phosphorylation** de l'histone H3 peut se faire au niveau de plusieurs résidus. La phosphorylation des résidus Ser10, Ser28 et Thr11 a été associée aux phases de mitose et de méiose (Gurley *et al.*, 1978 ; Goto *et al.*, 1999 ; Preuss *et al.*, 2003). Cependant, depuis quelques années, il a été observé que la phosphorylation de la sérine 10 est également corrélée à la transcription (Sassone-Corsi *et al.*, 1999 ; Crosio *et al.*, 2000), permettant une rapide induction de celle-ci suite à l'apparition de stimuli. De plus, la sérine 10 est localisée dans une zone de la protéine où d'autres résidus peuvent être méthylés ou acétylés (**Figure 8**). Ces autres marques peuvent influencer sur la capacité de la Ser10 à être phosphorylée, et vice versa (Rea *et al.*, 2000).



- L'**ubiquitylation** de l'histone H2A est la première à avoir été découverte, un seul motif ubiquitine y est généralement attaché (Goldknopf *et al.*, 1975). Par la suite, des motifs d'ubiquitine ont également été observés sur les autres histones (West et Bonner, 1980 ; Chen *et al.*, 1998 ; Pham et Sauer, 2000). L'ubiquitylation est connue, lorsqu'elle est multiple, pour induire la dégradation des protéines par le protéasome. La présence d'un seul motif n'est pas suffisante pour cibler les protéines vers leur dégradation. Il a alors été proposé que cette modification contrôle la formation ou la dissociation de complexes multiprotéiques en

fournissant des sites de fixation à des protéines. Depuis cette hypothèse, certaines ubiquitinations ont été caractérisées. Par exemple, les E3 ubiquitine ligases responsables de l'ubiquitinylation de l'histone H2A sont toutes les deux associées au silence transcriptionnel (Wang *et al.*, 2004 ; Cao *et al.*, 2005 ; Zhou *et al.*, 2008). L'enzyme RING1B (l'une de ces E3 ubiquitine ligases) interagit notamment avec le complexe Polycomb répresseur PRC1 (de Napoles *et al.*, 2004). L'ubiquitinylation de l'histone H2B, quant à elle, est associée à une activation de la transcription (Osley, 2006). En effet, H2B est transitoirement ubiquitinylée au niveau de régions codantes de plusieurs gènes hautement exprimés. Cette modification est effectuée par les activités de plusieurs enzymes qui sont en interaction avec l'ARN polymérase II (Henry *et al.*, 2003 ; Kao *et al.*, 2004 ; Xiao *et al.*, 2005). Comme pour les autres types de modifications, l'ubiquitinylation d'un résidu a des répercussions sur les résidus alentours, voire sur des résidus d'autres histones. C'est le cas par exemple de la monoubiquitinylation de H2B qui a des effets sur la méthylation de H3 (Dover *et al.*, 2002 ; Sun et Allis, 2002) (**Figure 9**).



**Figure 9 : Diagramme représentant l'effet de l'ubiquitinylation de H2B sur la méthylation de H3**

Rad6/Bre1 catalyse l'ajout d'un motif ubiquitine sur H2B, ce qui favorise la méthylation de la lysine 4 de H3 par l'enzyme Set1 et inhibe celle de la lysine 36 par l'enzyme Set2. La déubiquitinylation par Ubp8 provoque les effets inverses. Ub : ubiquitinylation ; Me : méthylation

D'après Zhang, 2003

Les histones peuvent subir de nombreuses modifications et il est maintenant clair que chaque modification peut influencer les autres. En effet, il a été démontré que la phosphorylation de la sérine 10 de H3 menait à une plus grande acétylation de la lysine 14 par l'histone acétyltransférase Gcn5 (Berger, 2002). La double modification phospho-sérine 10/acétyl-lysine 14 sur la même histone H3 est maintenant vue comme une marque reflétant

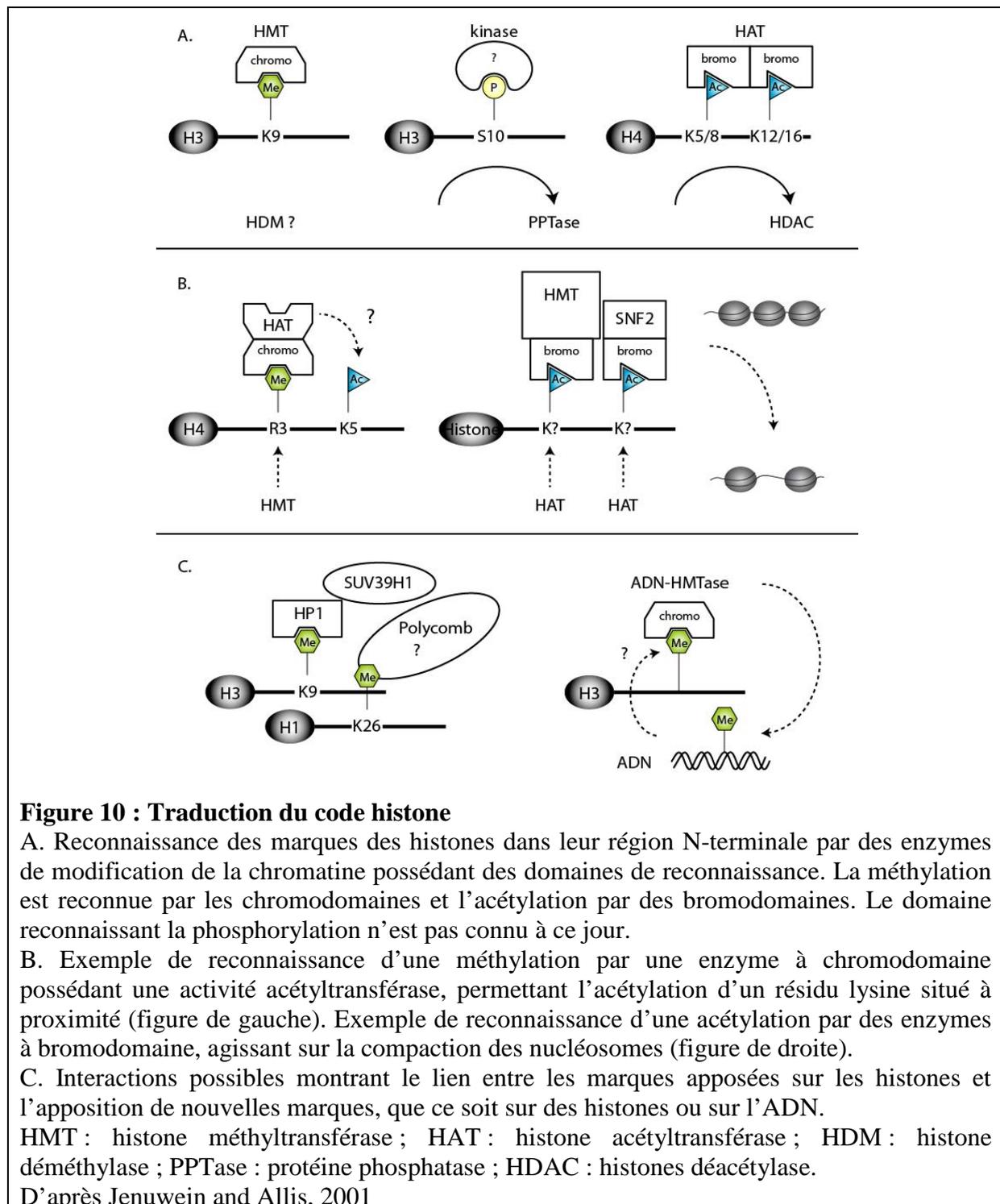
l'activation transcriptionnelle. De plus, d'autres expériences ont montré que la phosphorylation de la sérine 10 inhibe la méthylation de la lysine 9 de H3 et vice versa. Il en est de même pour l'acétylation de la lysine 14 qui inhibe la méthylation de la lysine 9.

De la même manière, les modifications du domaine N-terminal de l'histone H4 peuvent avoir des effets régulateurs les unes sur les autres. L'acétylation des lysines 8 et 12 par la CBP (CREB-binding protein) est augmentée lorsque l'arginine 3 est méthylée (Zhang et Reinberg, 2001). De plus, l'acétylation des lysines 5, 8, 12 et 16 défavoriserait la méthylation de l'arginine 3.

Il apparaît que ces modifications ne s'influencent non pas seulement lorsqu'elles sont présentes sur une même queue d'histone, mais aussi entre histones différentes. En effet, un lien existe entre l'ubiquitination de la lysine 123 de H2B et la méthylation des lysines 4 et 79 de H3 (Briggs *et al.*, 2002). L'ubiquitination de H2B *in vivo* par Rad6 est essentielle pour la méthylation des deux résidus de H3, à leur tour requis pour la formation de l'hétérochromatine.

Le signal épigénétique porté par le locus chromatinien doit par la suite être lu pour provoquer une réponse cellulaire. Certaines protéines responsables de cette lecture possèdent des domaines structuraux spécifiques servant à la reconnaissance des différentes modifications (Jenuwein et Allis, 2001). Les domaines les mieux caractérisés sont les bromodomaines et les chromodomaines. Il a été démontré que certains bromodomaines reconnaissent spécifiquement des lysines acétylées du domaine N-terminal des histones et permettent ainsi la rétention de complexes protéiques sur des régions de chromatine hyper-acétylées (Hassan *et al.*, 2002) (**Figure 10.A**). Cela explique partiellement le recrutement séquentiel de coactivateurs transcriptionnels à des séquences régulatrices de gènes *in vivo*, où l'acétylation par une HAT peut être prérequis au recrutement de complexes contenant une sous-unité à bromodomaine (**Figure 10.B**). Certains chromodomaines quant à eux peuvent se lier à des lysines méthylées (**Figure 10.A**). Cette dernière interaction est très largement étudiée car elle serait un mécanisme majeur pour la formation et la propagation de l'hétérochromatine. Le mécanisme proposé implique la méthylation de la lysine 9 de H3 par une HMT reliée à Su(var)3-9. Cette modification serait alors reconnue par le chromodomaine de HP1 et provoquerait sa polymérisation pour former l'hétérochromatine (Grewal et Elgin, 2002) (**Figure 10.C**). Il est intéressant de noter que cette méthylation de la lysine 9 de H3

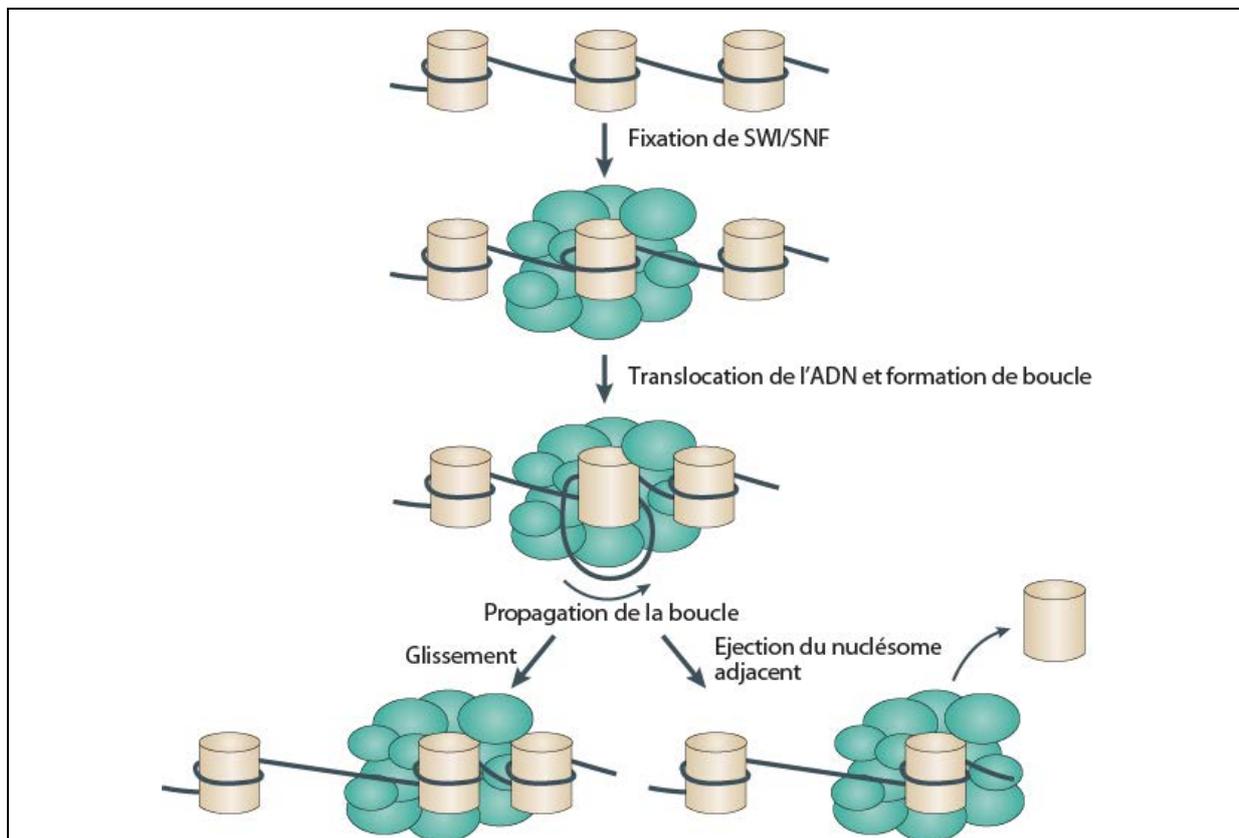
peut diriger la méthylation de l'ADN *in vivo*, liant ainsi fonctionnellement ces deux mécanismes de régulation épigénétique (Lachner, 2002) (**Figure 10.C**).



Le code des histones est donc interprété dans le contexte d'autres facteurs associés à la chromatine et c'est la combinaison d'interactions entre les histones modifiées et d'autres facteurs qui détermine si une protéine est recrutée sur la chromatine.

### c. Les complexes de remodelage modifiant la conformation de la chromatine

Il existe plusieurs familles de facteurs de remodelage de la chromatine. Tout d'abord, les complexes membres de la famille SWI/SNF (SWItch/Sucrose NonFermenting) possèdent tous un dénominateur commun : ils disposent à la fois d'une activité ATPase et d'un domaine similaire aux protéines possédant une activité hélicase. Chez *S. cerevisiae*, on retrouve deux représentants de cette famille : les complexes  $\gamma$ SWI/SNF (yeast SWI/SNF) et RSC (Remodel the Structure of Chromatin). Le complexe  $\gamma$ SWI/SNF a été d'abord caractérisé comme un co-activateur transcriptionnel, son activité de remodelage permettant le recrutement de la machinerie transcriptionnelle au site d'initiation de la transcription (Mohrmann et Verrijzer, 2005). En effet, ce complexe est recruté directement à ces sites par des activateurs et serait responsable de la disruption des interactions histone-ADN, changeant la structure de la chromatine (**Figure 11**).

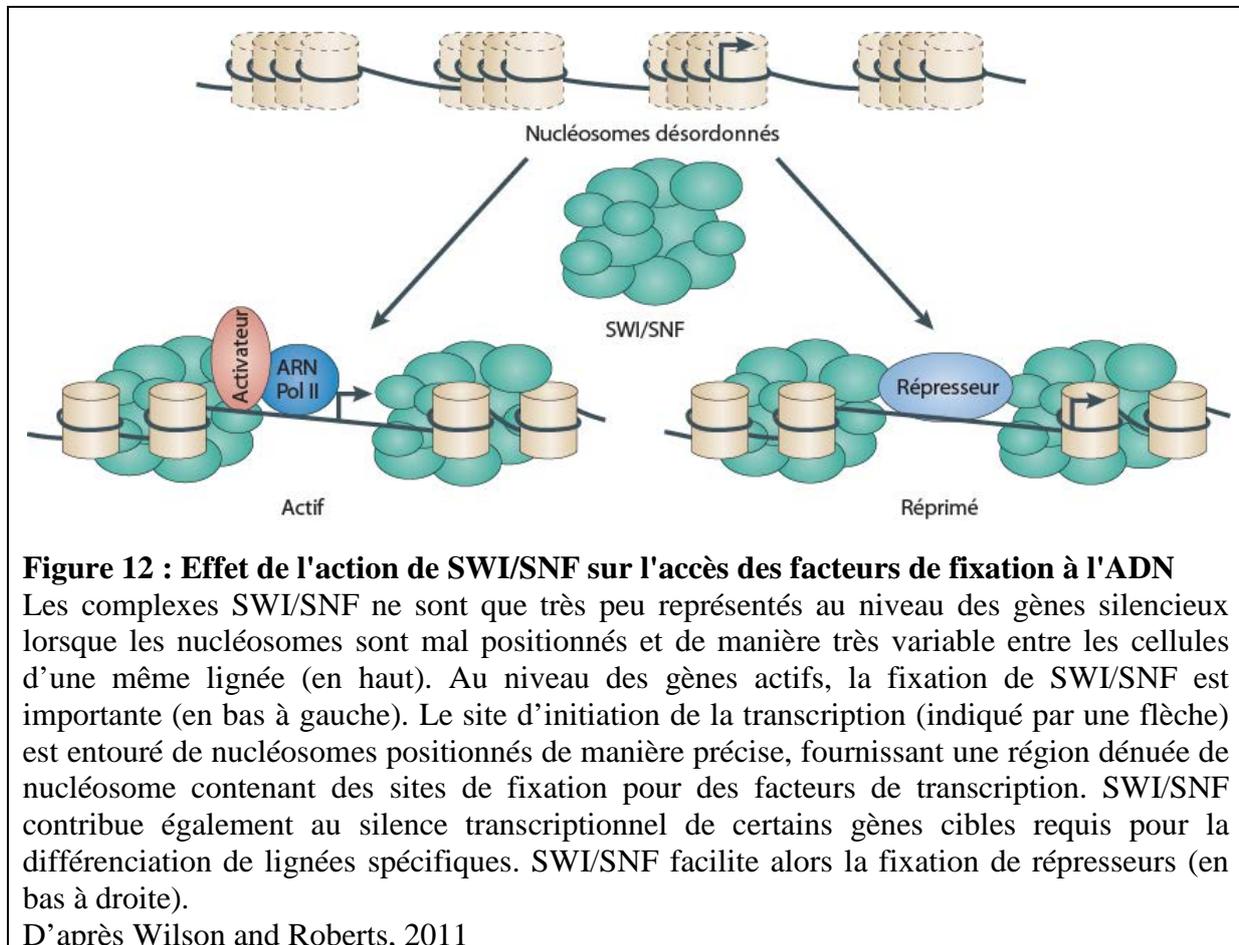


#### Figure 11 : Modèle d'action de SWI/SNF

Suite à la fixation du complexe SWI/SNF à l'ADN nucléosomal, les contacts histone-ADN se délient. La sous-unité ATPase permet ensuite la translocation de l'ADN et la formation d'une boucle qui peut se propager autour du nucléosome et générer ainsi des sites plus accessibles pour des facteurs de fixation à l'ADN. Le complexe SWI/SNF est également capable de catalyser l'éjection ou l'insertion d'octamères d'histones.

D'après Wilson et Roberts, 2011

Cette modification de la structure chromatinienne rend disponible l'ADN des régions promotrices aux facteurs initiant la transcription, mais aussi à des facteurs répresseurs (**Figure 12**). De plus, il a été établi que le complexe ySWI/SNF pouvait agir directement avec des séquences ADN, tel un activateur, via la sous-unité Swi2/Snf2.



Le complexe RSC est un complexe faisant aussi partie de la famille SWI/SNF. Il est essentiel à la cellule et y est beaucoup plus abondant que ySWI/SNF. Plusieurs évidences démontrent son importance au niveau de la régulation de l'expression des gènes. RSC serait impliqué à la fois dans la répression et l'activation de gènes. Le complexe RSC permettrait l'éjection totale d'un nucléosome ainsi que sa translocation sur une toute autre molécule d'ADN (cf. **Figure 11**) (Lorch *et al.*, 1999). De plus, ce complexe a été récemment impliqué dans la ségrégation chromosomique et la cohésion des chromatides sœurs (Hsu *et al.*, 2003 ; Baetz *et al.*, 2004 ; Huang et Laurent, 2004).

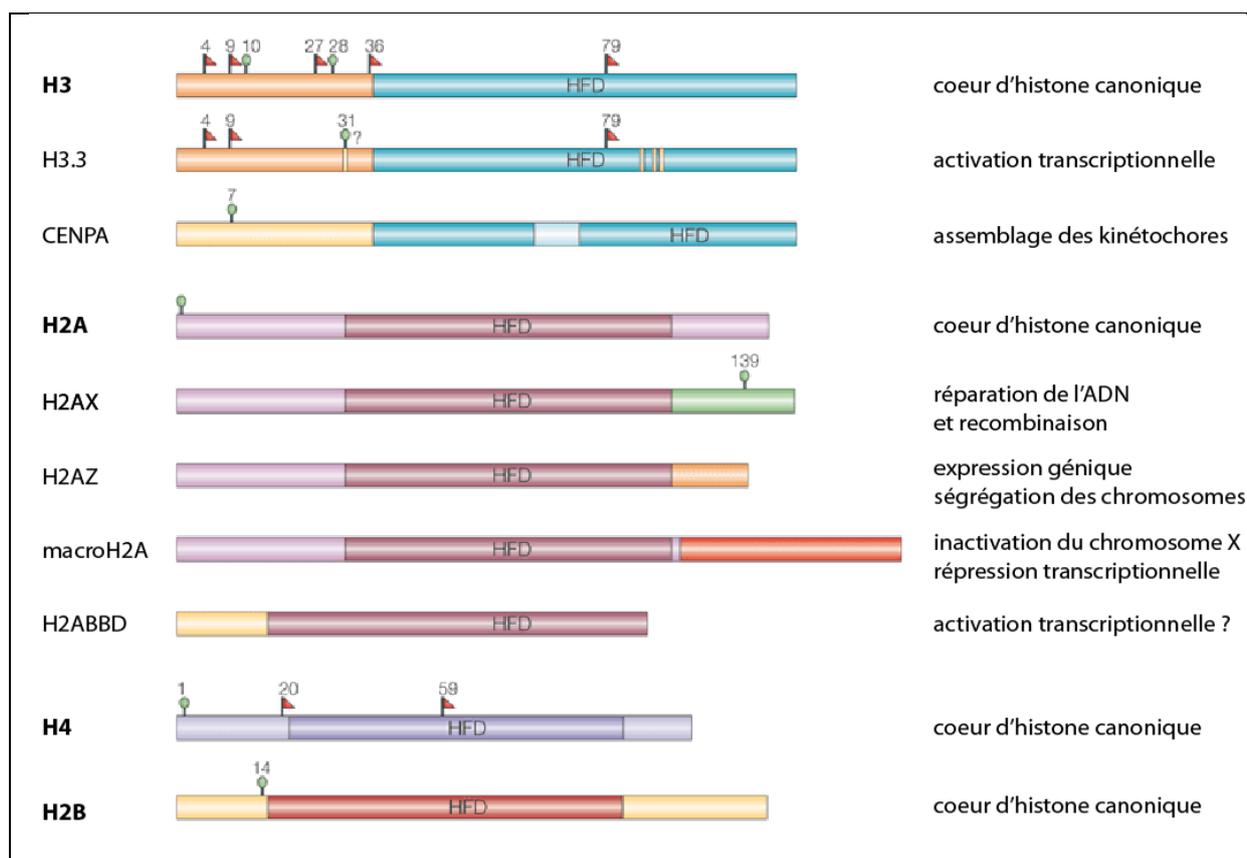
Au contraire des membres de la famille SWI/SNF, les membres de la famille ISW1 (Imitation of Switch 1) n'éjectent pas les nucléosomes. Ils les font glisser le long de l'ADN, changeant ainsi l'organisation nucléosomale. Ceci résulte non pas en la formation d'une région

entièrement dépourvue de nucléosomes, mais plutôt en une région où l'espacement entre les nucléosomes a été modifié. Chez la levure, il existe trois complexes de cette famille, Isw1a, Isw1b et Isw2. Isw1a est associé à la répression de la transcription, contrairement à Isw1b qui permet l'activation de la transcription (Morillon *et al.*, 2003). Ces deux complexes ont été également impliqués dans le contrôle des événements menant à l'élongation et à la terminaison transcriptionnelle (Morillon *et al.*, 2003).

#### d. Les variants d'histones

Deux classifications des variants d'histones ont été proposées. La première, basée sur la structure, distingue les variants homéomorphes des variants hétéromorphes. Les variants homéomorphes sont retrouvés dans toutes les classes d'histones, et ne diffèrent de leurs homologues conventionnels que par quelques substitutions d'acides aminés sans conséquence sur leur mobilité électrophorétique (à l'exception de H2A2). Au contraire, les variants hétéromorphes présentent une extrémité carboxy-terminale dont la taille et la séquence en acides aminés sont suffisamment variables pour modifier de façon significative leur mobilité électrophorétique. Ces variants sont retrouvés dans les classes d'histones H2A et H2B et sont intégrés dans la chromatine dans des proportions variables. La seconde classification repose quant à elle sur des critères plus fonctionnels : dépendance ou indépendance de la synthèse par rapport à la réplication, existence d'une spécificité tissulaire d'expression.

Quelle que soit la classification, la substitution au sein du nucléosome d'une histone conventionnelle par un variant peut créer un nouveau complexe nucléoprotéique de structure et de stabilité modifiées. Cette modification pourrait favoriser le recrutement de facteurs régulateurs pour assurer les fonctions nucléaires. La **Figure 13** présente des variants d'histones ainsi que leur fonction supposée.



**Figure 13 : Représentation schématique de variants d'histones**

Dans l'histone H3.3, les résidus qui diffèrent de l'histone canonique H3 sont surlignés en jaune. L'un de ces résidus, la sérine 31, est potentiellement un site de phosphorylation. L'histone CENPA est une histone centromérique qui possède une extrémité N-terminale unique qui ne ressemble à aucune autre. Deux sites de phosphorylation ont été identifiés dans cette région, dont l'un est essentiel pour l'achèvement de la cytokinèse. Les variants de l'histone H2A diffèrent fortement dans leur région C-terminale. L'extrémité C-terminale de H2AX présente une sérine conservée (Ser139) dont la phosphorylation constitue un élément de réponse aux coupures ADN double-brin. MacroH2A possède une région C-terminale plus longue dont la fonction n'est pas connue. H2ABBD est le plus court des variants de H2A qui possède un domaine N-terminal différent dans lequel sont absents tous les sites de modification présents dans H2A. Les fonctions potentielles de ces variants d'histones sont listées à leur droite.

La position des résidus modifiés est indiquée au-dessus de chaque modification. Drapeau rouge : méthylation de lysine ; cercle vert : phosphorylation de sérine ; HFD : histone-fold domain (domaine conserve entre les histones).

D'après Sarma et Reinberg, 2005

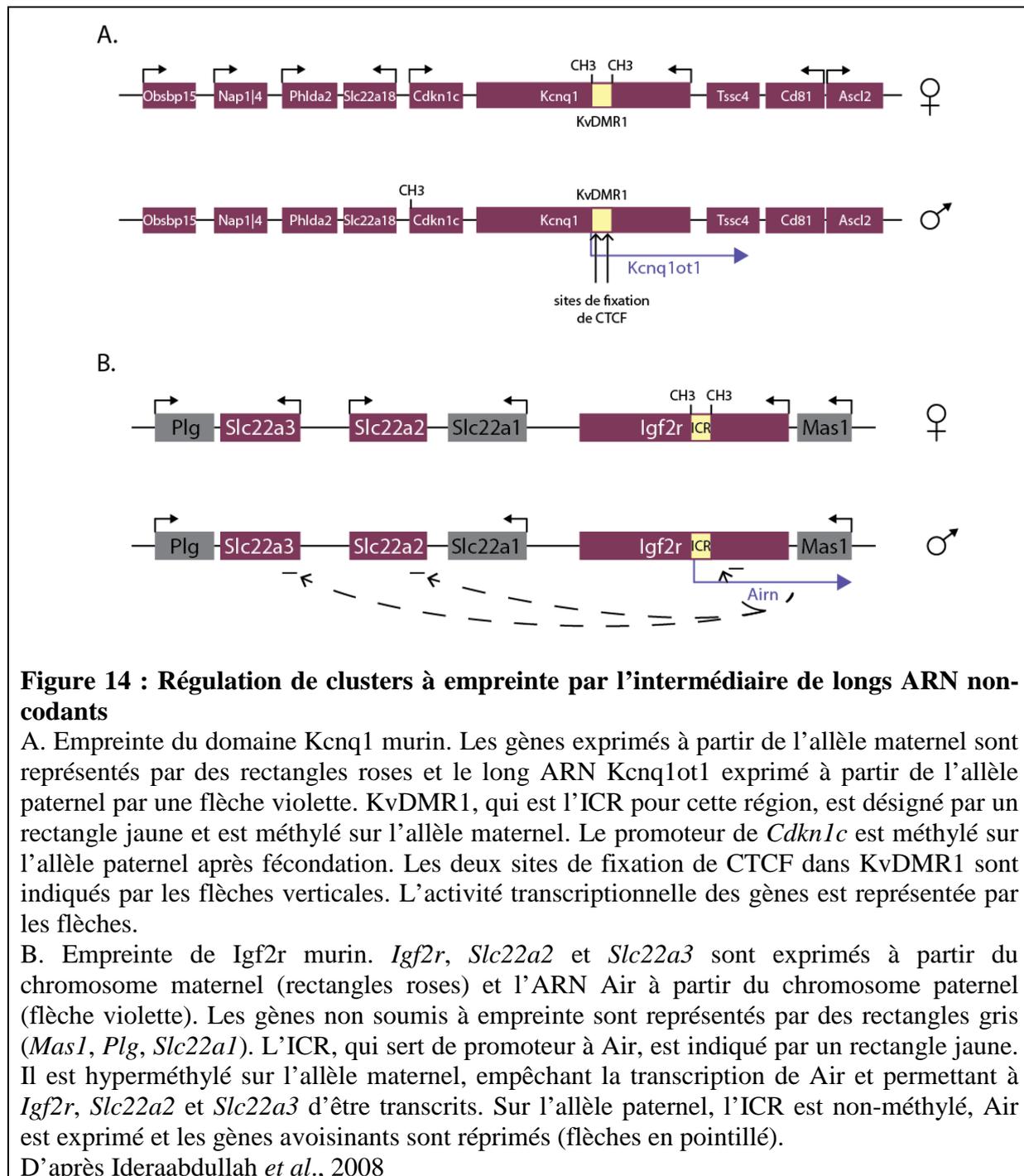
## I.2 L'implication des ARN dans la régulation épigénétique

De plus en plus d'études révèlent un rôle des ARN non-codants (ARNnc) dans la modulation de l'état épigénétique des gènes et dans la régulation de la transcription. La plupart des ARN mis en évidence sont de longs ARN non-codants mais de nouveaux rapports indiquent l'importance de nombreux petits ARN non-codants également.

### I.2.1 La régulation par de longs ARN non-codants

L’empreinte parentale est un bon exemple pour illustrer la régulation par les longs ARN non-codants. Les gènes soumis à empreinte sont régulés en *cis* par des régions de contrôle par empreinte (Imprinting Control Regions, ICR). Ces dernières peuvent réprimer les gènes adjacents par l’utilisation de longs ARN non-codants. Malgré le fait qu’elles régulent divers loci génomiques, elles partagent un mécanisme commun émergent : les longs ARN non-codants antisens interagissent avec des complexes de modification de la chromatine, ce qui résulte en l’extinction de l’allèle maternel ou paternel. Il y a quelques années, deux longs ARN non-codants impliqués dans la régulation par empreinte ont été montrés comme se liant à des complexes de modification de la chromatine et étant requis pour la localisation correcte des modifications d’histones.

- Le premier exemple est celui du gène *Kcnq1*. Le locus, et l’ARN non-codant en particulier, est gouverné par une ICR maternellement méthylée, KvDMR1, qui est située dans un intron du gène *Kcnq1*. Le promoteur de ce gène est lui-même situé dans KvDMR1. L’hypométhylation du promoteur sur l’allèle paternel est associée à l’expression de *Kcnq1* et à la répression des gènes adjacents soumis à empreinte, codant eux des protéines. A l’inverse, l’hyperméthylation de KvDMR1 sur l’allèle maternel est associée à la répression de l’ARNnc *Kcnq1ot1* et l’activation des gènes adjacents (**Figure 14.A**). En effet, la délétion de KvDMR1 sur l’allèle paternel conduit à un défaut d’expression de *Kcnq1* et à l’expression biallélique des gènes qui sont normalement uniquement exprimés à partir de l’allèle maternel, ce qui suggère que la transcription de l’ARNnc *Kcnq1ot1* est essentielle au silence transcriptionnel des 8 gènes codants, en *cis* (Mancini-Dinardo *et al.*, 2006).



Plus récemment, il a été montré que l'expression par empreinte du gène *Cdkn1c* peut être régulée par un mécanisme indépendant de l'ARNnc *Kcnq1ot1* (Shin *et al.*, 2008). Cependant, ce mécanisme est tout de même relié à des éléments de *KvDMR1*. Deux sites de fixation du facteur CTCF ont été identifiés. Le facteur CTCF (CCCTC-binding factor) est un régulateur transcriptionnel très conservé qui se lie à des séquences d'ADN spécifiques (Filippova *et al.*, 1996). Il est impliqué dans la régulation de la transcription par fixation à des insulateurs de la chromatine, empêchant ainsi l'interaction entre les promoteurs et les éléments activateurs ou

répresseurs alentours. Dans le cas de KvDMR1, les sites de fixation de CTCF ne sont occupés *in vivo* que sur l'allèle paternel non méthylé (Fitzpatrick *et al.*, 2007). Ainsi, il est possible que pour un gène aussi important dans le développement de l'embryon que *Cdkn1c*, des mécanismes redondants, dépendants ou indépendants de l'ARN *Kcnq1ot1*, soient mis en place afin d'assurer une expression par empreinte appropriée (**Figure 14.A**) (Ideraabdullah *et al.*, 2008).

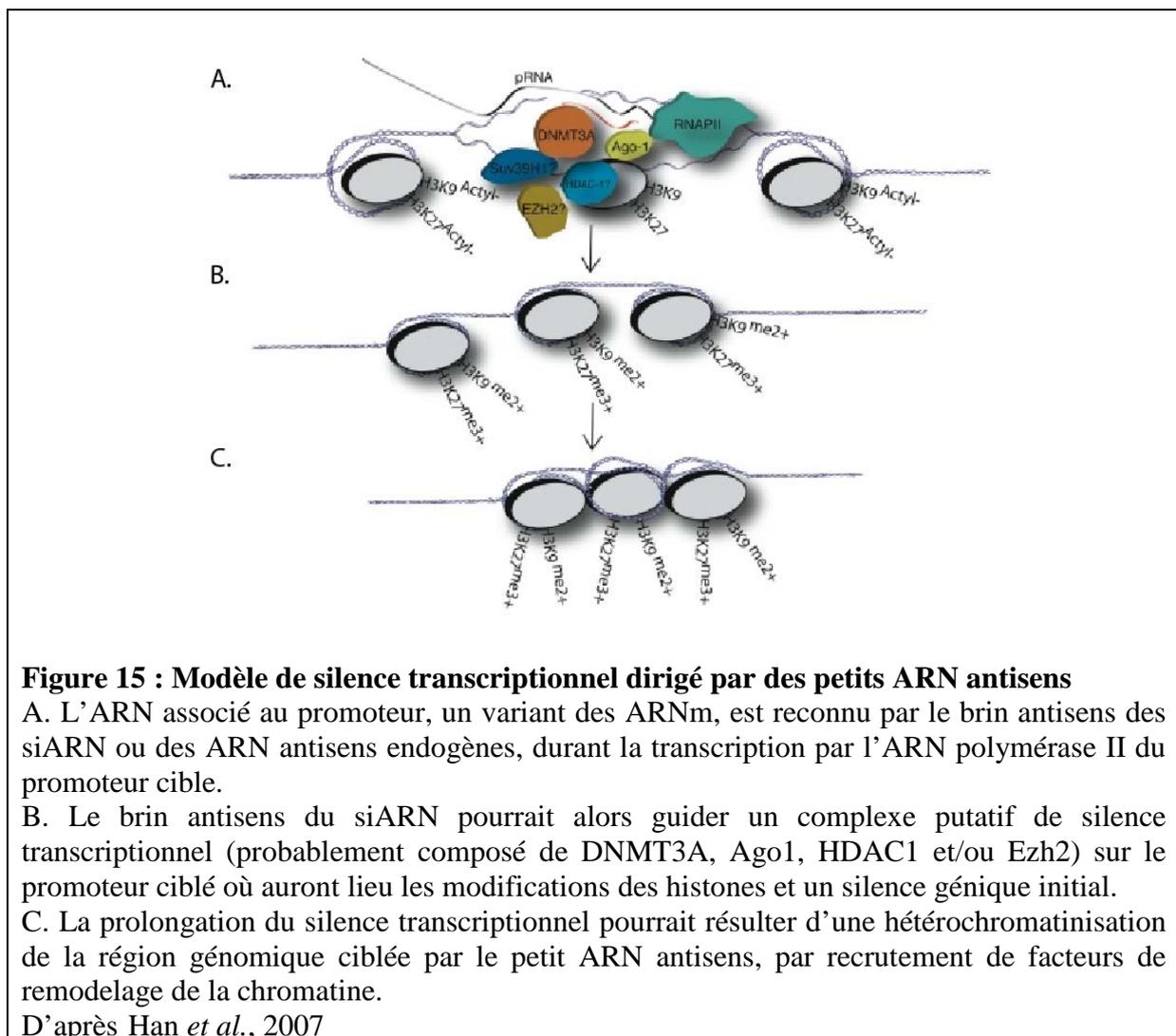
- Une observation similaire a été reportée concernant le cluster *Igf2r* (**Figure 14.B**). *Igf2r* et ses deux voisins, *Slc22a2* et *Slc22a3*, sont exprimés à partir de l'allèle maternel. Cette région contient également un transcrite exprimé paternellement, l'ARN Air (Antisens *Igf2r* RNA). Son expression est essentielle pour le silence transcriptionnel des gènes exprimés maternellement *in cis* (Sleutels *et al.*, 2002). L'expression de Air est également régulée par une ICR qui porte des modifications épigénétiques spécifiques selon l'allèle (Regha *et al.*, 2007).

### I.2.2 La régulation par de petits ARN non-codants

L'ARN interférence est un processus conservé par lequel un ARN double-brin spécifique d'une séquence est converti en un petit ARN double-brin interférant (siARN, small interfering ARN). Ce siARN peut induire le silence d'un gène par 2 voies : par silence post-transcriptionnel du gène, le plus connu jusqu'à présent, et par inhibition de la transcription (ou silence transcriptionnel du gène, TGS). Ce dernier permet une répression à long-terme des gènes mais il est beaucoup moins compris jusqu'à présent. Un modèle putatif du mécanisme de régulation dans les cellules humaines a pu être élaboré. Dans ce modèle, la génération et l'activité des petites molécules ARN requiert la protéine Argonaute A (Ago1), un des composants du complexe RISC (RNA-induced silencing complex) impliqué dans le processus d'ARN interférence. La reconnaissance des cibles s'effectue par complémentarité de séquence entre le brin antisens du petit ARN qui cible le promoteur et l'ARN associé au promoteur (pARN) (Han *et al.*, 2007 ; Schwartz *et al.*, 2008). Une fois exposé au petit ARN, le promoteur ciblé présente des taux plus élevés des marques épigénétiques associées au silence transcriptionnel, H3K9me2 et H3K27me3 (Janowski *et al.*, 2006 ; Kim *et al.*, 2006 ; Weinberg *et al.*, 2006 ; Han *et al.*, 2007). Ces marques d'une chromatine réprimée ont été mises en évidence non seulement au niveau du locus ciblé mais aussi en aval, suggérant un mécanisme par lequel les modifications des histones peuvent s'étendre sur les gènes voisins

ou sur des régions intergéniques (Morris, 2005 ; Weinberg *et al.*, 2006). Un effet sur la méthylation des îlots CpG reste incertaine puisque certains promoteurs présentent une méthylation accrue tandis que d'autres non.

Les protéines responsables de ces processus enzymatiques font partie d'un complexe de silence transcriptionnel (TSC) partiellement caractérisé. Des travaux récents montrent que ce TSC putatif contient les protéines DNMT3A (Weinberg *et al.*, 2006), Ago1 (Janowski *et al.*, 2006 ; Kim *et al.*, 2006), HDAC1 (Suzuki *et al.*, 2008) et une histone méthyltransférase telle que Ezh2 ou G9A (Viré *et al.*, 2006 ; Han *et al.*, 2007) (**Figure 15**).



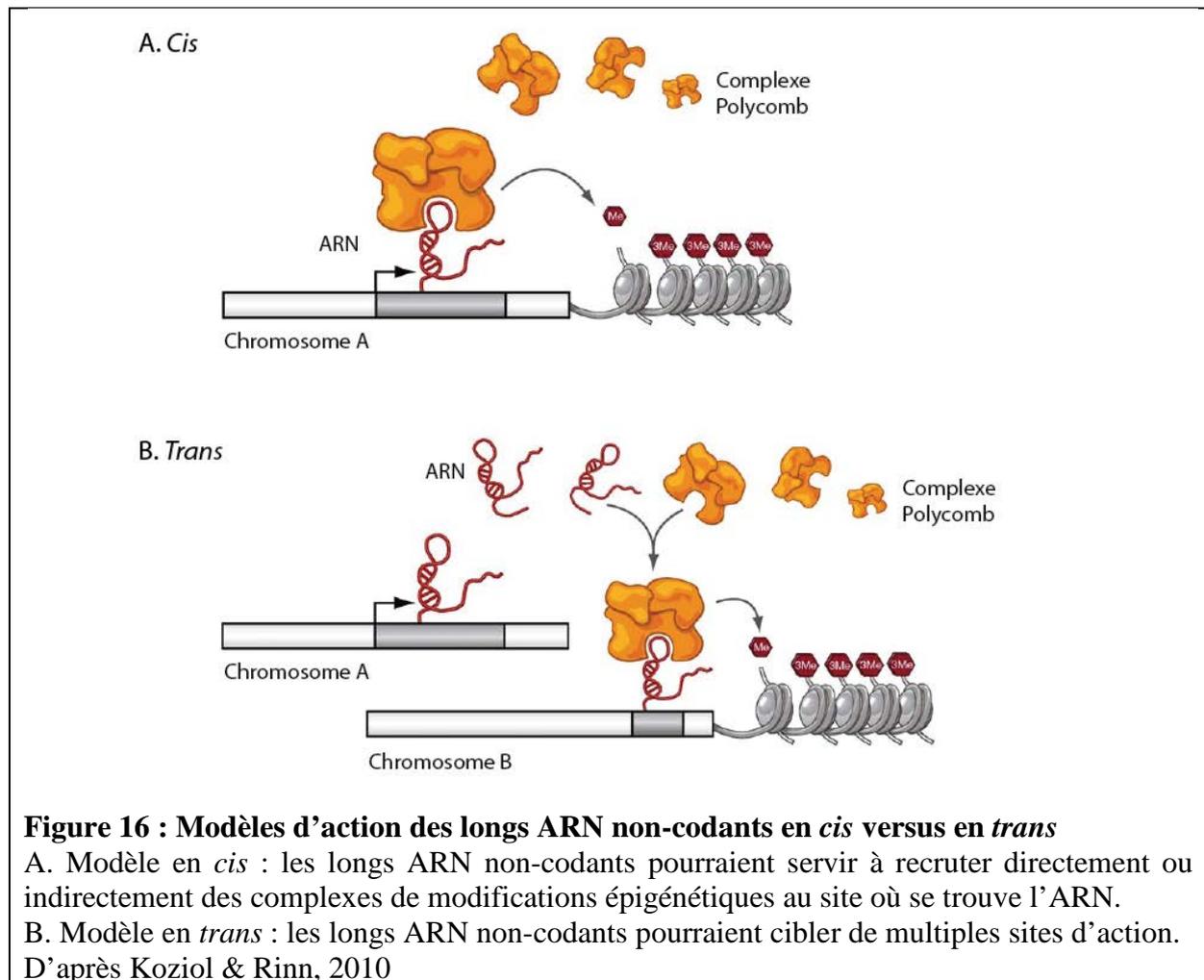
Dernièrement, plusieurs exemples de TGS médiés par des ARN endogènes ont été identifiés (Gonzalez *et al.*, 2008 ; Kim *et al.*, 2008). Il a également été proposé que certains transcrits antisens non-codants agissent par cette voie sur leur propre promoteur sens dans le but de modifier la signature chromatinienne locale (Yu *et al.*, 2008 ; Guttman *et al.*, 2009).

Un exemple de régulation par ces petits ARN non-codants est celui du silence transcriptionnel du gène de l'ubiquitine C (*Ubc*). L'induction du silence de ce gène est opérationnelle par méthylation des histones sur des sites spécifiques, et elle nécessite la présence de l'ARN non-codants associé au promoteur *Ubc* et des protéines HDAC1, DNMT3A et Ago1 (Hawkins *et al.*, 2009). La maintenance du silence de *Ubc* requiert quant à elle d'autres protéines, comme par exemple DNMT1 et peut-être DNMT3A aussi. De plus, une méthylation significative de l'ADN a été détectée après quelques jours de mise en présence du petit ARN non-codant, ce qui indique la mise en place d'une répression à long-terme.

### I.2.3 Les mécanismes possibles

Une étude récente a montré que de nombreux longs ARN non-codants « intervenants » (large intervening non-coding RNAs, lincRNAs) se lient à PRC2 et à de multiples autres complexes de modification de la chromatine (Khalil *et al.*, 2009). Cela suggère un rôle plus global des lincRNA, et il se pourrait qu'ils guident ces complexes vers leurs sites spécifiques. De plus, l'étude a également montré que les lincRNAs étaient requis pour la répression correcte des gènes, médiée par le PRC2. D'ailleurs, le thème général qui ressort des différentes études sur le sujet est que les longs ARN non-codants sont requis pour la mise en place correcte de domaines chromatiniens, probablement en dirigeant les complexes de modification vers leurs destinations spécifiques (Nagano *et al.*, 2008).

Dans un premier modèle, il est proposé que les longs ARN non-codants puissent servir de point d'ancrage pour le recrutement, direct ou indirect, de complexes de modification de la chromatine au locus spécifique de la transcription de ces ARN (**Figure 16.A**).



En parallèle de ce modèle d'action en *cis*, ces ARN pourraient également agir en *trans* et ainsi cibler de nombreux sites (**Figure 16.B**). Il existe de nombreux modèles mécanistiques selon lesquels les longs ARN non-codants peuvent agir en *trans*. L'une des possibilités est la formation de complexes ARN-ADN qui pourraient servir de balise pour le recrutement des complexes de modification de la chromatine (Koziol et Rinn, 2010).

Une autre possibilité de modulation des complexes de modification de la chromatine par l'ARN pourrait impliquer des longs ARN non-codants servant de plate-forme architecturale qui lierait directement ou indirectement les complexes de la chromatine aux protéines de fixation à l'ADN ou aux facteurs de transcription.

Enfin, des interactions de ces ARN avec des complexes de modification de la chromatine pourraient permettre des interactions entre des éléments ADN éloignés (intra et interchromosomiques) par formation de boucles d'ADN. En accord avec cette idée, l'interaction de *Kcnq1ot1* avec les complexes de modification de la chromatine ne réprime pas un domaine continu de gènes avoisinants. Puisque plusieurs gènes sont omis, il est

possible que *Kcnq1ot1* serve d'intermédiaire pour une interaction tri-dimensionnelle ADN-ARN-protéine qui définit précisément les gènes régulés au locus en question.

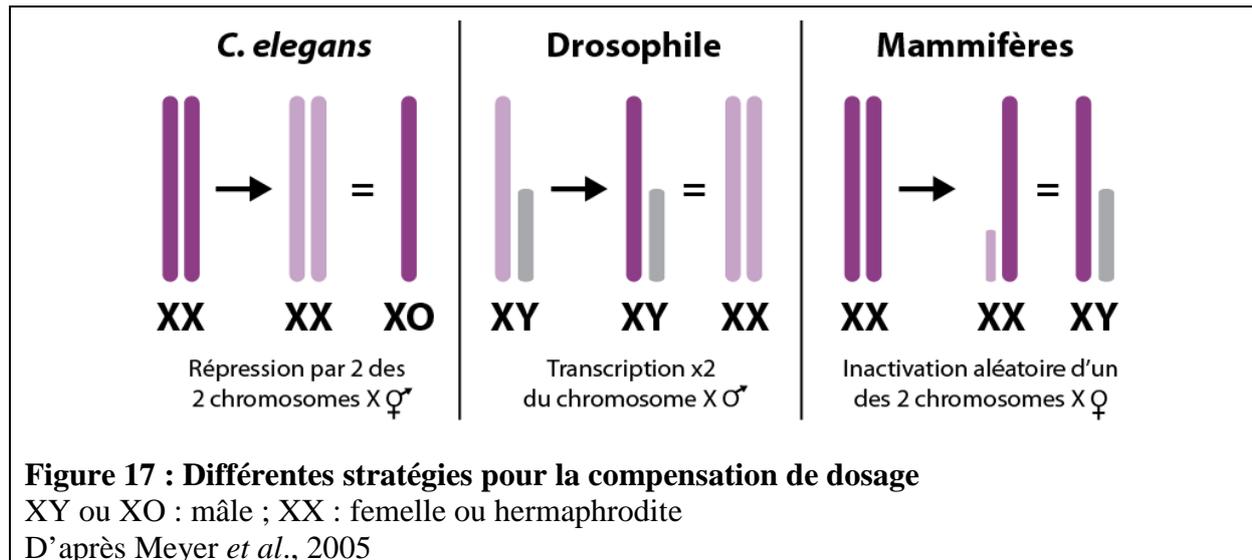
Il est important de noter que ce genre de régulation n'est pas dû à une transcription passive mais plutôt au fait que les longues molécules d'ARN non-codants elles-mêmes sont recrutées pour faciliter le recrutement et la répression à ces loci.

## II. La compensation du dosage des gènes

### II.1 Les principes généraux

Chez de nombreuses espèces, les chromosomes sexuels présentent des différences morphologiques. Les individus homogamétiques présentent deux chromosomes sexuels identiques (XX pour un organisme mammifère femelle par exemple), alors que les individus hétérogamétiques présentent deux chromosomes sexuels distincts (XY pour un organisme mammifère mâle). La caractéristique chromosomique majeure d'un individu hétérogamétique est un nombre de gènes liés aux chromosomes sexuels inférieur à celui présent dans un organisme homogamétique (pour revues, Marin *et al.*, 2000 ; Charlesworth *et al.*, 2005). Ainsi, le taux d'expression des gènes liés aux chromosomes sexuels peut provoquer un déséquilibre du taux de produits de ces gènes entre les autosomes et les gonosomes, mais aussi entre les deux sexes. Pour pallier à ce problème, un mécanisme de compensation, appelé compensation de dose des gènes, a été mis en place par la sélection naturelle au cours de l'évolution.

Les stratégies utilisées pour aboutir à la compensation de dosage sont diverses, mais dans tous les cas, des complexes spécialisés sont recrutés spécifiquement sur le ou les chromosomes X d'un seul des deux sexes afin de réguler la transcription. Chez le ver *C. elegans* par exemple, la compensation s'effectue par une répression des deux chromosomes X de l'individu hermaphrodite. L'expression de chaque chromosome étant réduite de moitié, l'expression totale correspond à celle des individus XO. Chez la Drosophile en revanche, le mécanisme est inverse puisque la compensation s'opère chez l'individu mâle dont le seul chromosome X est exprimé deux fois plus que chez l'individu femelle. Une autre stratégie est encore observée chez les mammifères où l'un des deux chromosomes X de l'individu femelle est inactivé, laissant un seul chromosome X actif, comme chez l'individu mâle (**Figure 17**).



## II.2 La compensation du dosage chez différentes espèces

### II.2.1 Chez *Drosophila melanogaster*

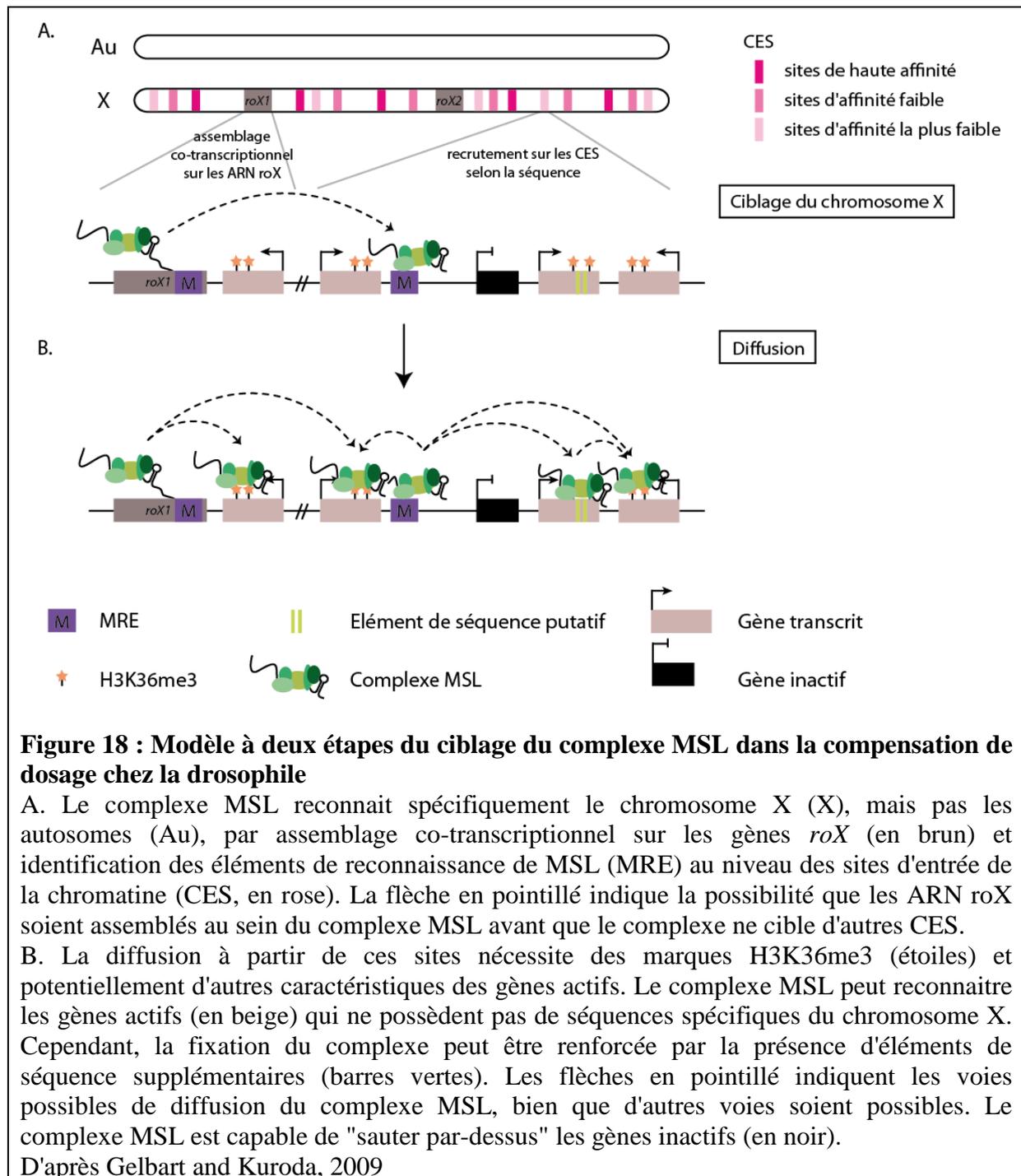
Chez la *Drosophile*, l'expression des gènes portés par le chromosome X est équilibrée par un taux de transcription doublé de l'unique chromosome X des mâles XY, comparativement aux femelles XX. Chez les mâles, le complexe MSL (Male-Specific Lethal) diffuse de manière bidirectionnelle sur la chromatine du X, en se fixant sur des centaines de sites, nommés sites d'entrée sur la chromatine (ou CES, Chromatin Entry Site), et double approximativement l'expression des gènes de ce chromosome. Le complexe MSL contient 6 protéines : MSL1 ; MSL2, une protéine avec un domaine RING-finger (RING : Really Interesting New Gene) ; MSL3, une protéine à chromodomaine ; MLE, une ARN hélicase ; MOF, une histone acétyltransférase à chromodomaine ; et JIL-1, une kinase. Deux ARN non-codants codés par le chromosome X, roX1 et roX2, font également partie du complexe (Gelbart et Kuroda, 2009 ; Hallacli et Akhtar, 2009). Le complexe MSL modifie covalamment la queue des histones grâce à l'activité de MOF qui acétyle la lysine 16 de l'histone H4 (H4Ac16), une modification liée à l'augmentation de la transcription, et de la protéine kinase JIL-1 qui phosphoryle l'histone H3 *in vitro*. La diffusion du complexe MSL requiert l'activité ATPase de l'hélicase MLE et l'activité acétyltransférase de MOF. Ainsi, ce complexe ribonucléoprotéique agit sur le recouvrement en *cis* du chromosome et sur le remodelage de la chromatine.

Par ailleurs, les ARN roX1 et roX2 fonctionnent de manière redondante dans la compensation, mais 5% des mâles survivent en leur absence. Les mâles déficients en ARN roX présentent une perturbation dans la localisation de la protéine MSL2 et une perte de

l'enrichissement du chromosome X en H4Ac16. De plus, chez ces mêmes mâles, MSL2 se relocalise sur des régions autosomales et semble les activer (Franke et Baker, 2000). Cela suggère que le complexe MSL partiel contient encore une activité de remodelage de la chromatine et que les ARN roX sont requis pour l'assemblage du complexe MSL aux sites appropriés de la chromatine sur le chromosome X et pour une compensation de dosage correcte. Plus récemment, il a été montré que l'ARN roX2, de par la régulation de son épissage, ait un rôle important dans l'assemblage ou la fonction optimale du complexe MSL (Park *et al.*, 2005).

Deux des sites d'entrée de la chromatine qui dirigent le complexe MSL sur le chromosome X se trouvent sur les gènes *roX*. Un rapport montre que les ARN roX sont en compétition pour limiter la quantité de protéines MSL pouvant former des complexes MSL fonctionnels, et que la concentration de complexes MSL sur le gène *roX* détermine le taux de diffusion en *cis* à partir de celui-ci (Park *et al.*, 2002). En effet, il a été suggéré que le taux de formation de complexes MSL est lent lorsque les protéines MSL sont limitantes, ce qui laisse le temps aux ARN roX de diffuser à partir de leur lieu de synthèse et de s'assembler au sein d'un complexe MSL au niveau d'un site d'entrée *trans* de la chromatine. Lorsque les protéines MSL ne sont pas limitantes, l'assemblage d'un complexe MSL fonctionnel est rapide, facilitant la diffusion de ce complexe en *cis* à partir du site de synthèse. Ainsi, chez les mâles, la concentration en protéines MSL pourrait déterminer si le complexe diffuse en *cis* à partir d'un site d'entrée de la chromatine localisé dans un gène *roX*, ou en *trans* à partir d'un des autres sites d'entrée du chromosome X.

A partir des sites d'assemblage, le complexe diffuse sur les autres sites de fixation à haute affinité. Cette diffusion aurait lieu, du moins en partie, par reconnaissance de motifs ADN spécifiques (**Figure 18**).



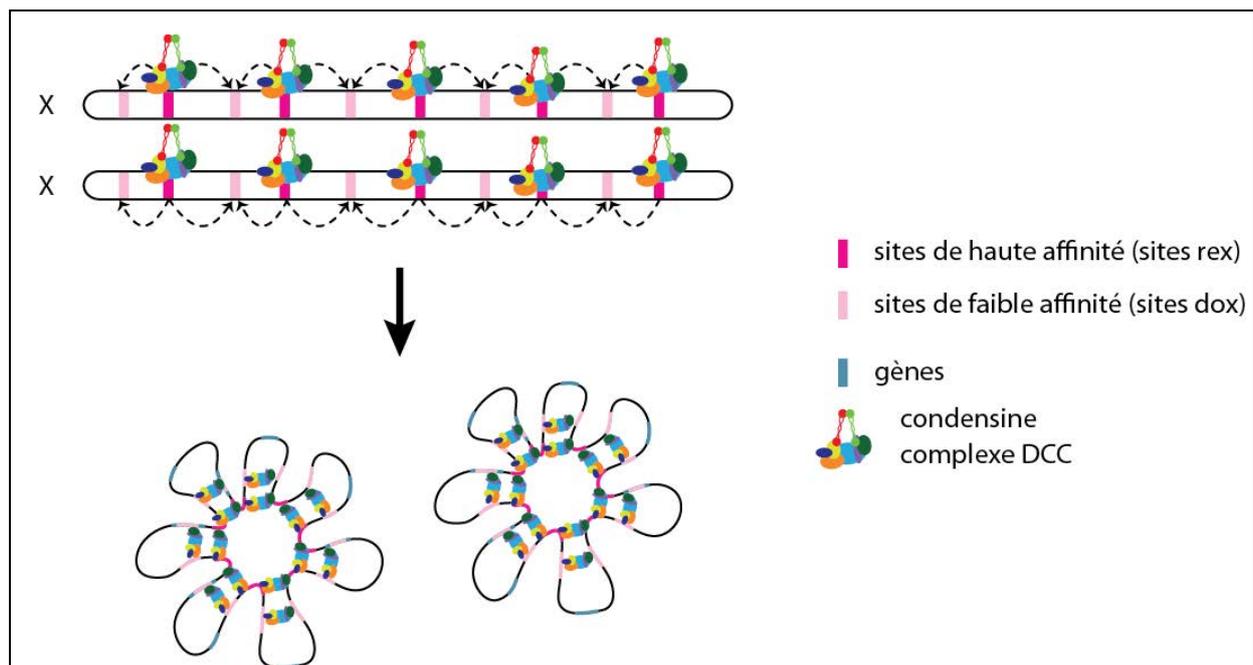
Néanmoins, les motifs ADN identifiés ne peuvent pas prédire seuls de manière univoque la fixation du complexe MSL puisque beaucoup de ces motifs ne sont pas associés avec ce complexe. D'autres facteurs doivent donc être impliqués. Il a été proposé que l'organisation nucléaire et le haut degré de structure de la chromatine puissent jouer un rôle, éventuellement en facilitant la diffusion des complexes vers les régions activement transcrites du chromosome. Il a été observé que le chromosome X mâle peut adopter une conformation

tri-dimensionnelle spécifique exposant des régions riches en sites à haute affinité pour le complexe MSL (Grimaud et Becker, 2009). De plus, les protéines MSL purifiées sont associées avec le complexe de pore nucléaire, Nup153 et Megator, et des expériences d'ARN interférence ont montré que ces facteurs sont essentiels pour la compensation du dosage médiée par MSL. Bien que la périphérie nucléaire soit souvent associée avec l'hétérochromatine et le silence transcriptionnel, il existe des sous-régions au niveau de l'enveloppe nucléaire, plus particulièrement au niveau des pores, qui sont associées avec une activité transcriptionnelle. Les protéines Nup153 et Megator se fixent à des régions associées aux nucleoporines (Nucleoporin-Associated Regions, NARs). Ces régions représentent des zones de chromatine ouverte car elles sont enrichies en marqueurs de transcription active comme l'ARN polymérase II et l'acétylation H4K16. Le chromosome X mâle hyperactif est particulièrement enrichi en ces régions et interagit avec des nucléoporines pour sa localisation périphérique. Le lien entre le complexe MSL, les composants des pores nucléaires et la périphérie nucléaire suggère que le chromosome X chez les mâles se localise spécifiquement au niveau de sous-domaines nucléaires via ses interactions avec les pores nucléaires et cela pourrait contribuer à son expression doublée. Ainsi, l'activation de l'expression du chromosome X chez les mâles pourrait être non seulement transcriptionnelle par des effets liés à la chromatine, mais également post-transcriptionnelle, probablement par une maturation et un export plus rapides des ARN messagers grâce à l'association avec les pores nucléaires.

### II.2.2 Chez *Caenorhabditis elegans*

Chez le ver *C. elegans*, la compensation du dosage implique une réduction par deux fois de l'activité des gènes du chromosome X chez les hermaphrodites (cf. **Figure 17**). Elle met en jeu un complexe appelé DCC (pour « Dosage Compensation Complex ») qui consiste en un cœur de protéines spécifiques des hermaphrodites, les protéines SDC (pour « Sex Determination and dosage Compensation ») (Hodgkin, 1987 ; Plenefisch *et al.*, 1989). Ces protéines sont requises pour la stabilisation et le ciblage du complexe sur le chromosome X. Le DCC interagit avec un autre complexe, le complexe condensine initialement connu pour sa fonction dans la condensation des chromosomes lors de la mitose et de la méiose. L'interaction avec la condensine induirait des changements de la chromatine nécessaires pour la compensation du dosage (Chan *et al.*, 2004) (**Figure 19**).

Cependant, les mécanismes par lesquels le complexe DCC permet une répression par deux fois de l'expression des gènes liés au chromosome X, et cela sur tout le chromosome, restent incompris. Il cible le chromosome X par interaction initiale avec un certain nombre d'éléments recruteurs, suivie par une dispersion ou diffusion vers des sites secondaires. Il a été proposé que le DCC ne réprime pas les gènes par une interaction directe mais qu'il agit plutôt à distance et pourrait impliquer la formation de boucles de fibres de chromatine, afin de rapprocher les gènes régulés et les sites de fixation du DCC (Jans *et al.*, 2009). L'implication de la condensine suggère que la répression du chromosome X pourrait inclure des mécanismes similaires à la condensation de la chromatine lors de la mitose (Chuang *et al.*, 1996 ; Lieb *et al.*, 1996 ; Lieb *et al.*, 1998 ; Tsai *et al.*, 2008 ; Csankovszki *et al.*, 2009). Le DCC pourrait altérer l'organisation du chromosome en induisant une condensation partielle du chromosome pendant l'interphase, ce qui pourrait à son tour aboutir à une réduction de la transcription (**Figure 19**).



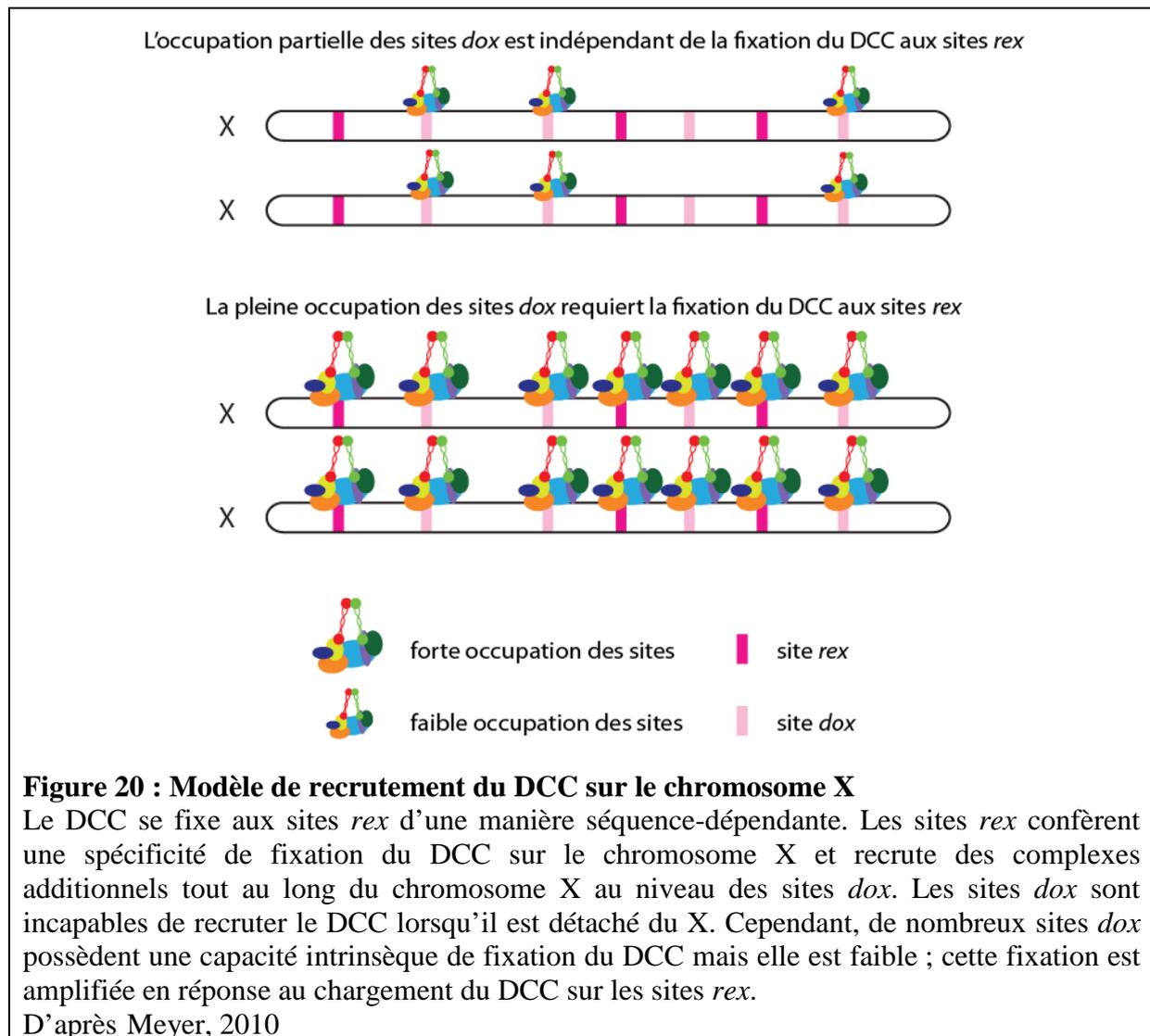
**Figure 19 : Stratégie de compensation de dosage chez les hermaphrodites *C. elegans***

Les hermaphrodites (XX) de *C. elegans* diminuent l'expression de chacun de leur chromosome X par deux. Le complexe DCC se fixerait sur des sites à haute affinité (sites *rex*, en rose foncé) puis diffuserait sur l'ensemble du chromosome (flèches en pointillés) par l'intermédiaire de sites à plus faible affinité (sites *dox*, en rose pâle). Cela conduirait à une réorganisation des chromosomes X par la formation de boucles de la chromatine, rapprochant ainsi les gènes à réguler (en bleu) à proximité des sites de fixation du DCC. Cette nouvelle organisation est supportée par l'implication de la condensine dans le processus de compensation de dosage.

D'après Chow & Heard, 2010

Cette idée est étayée par le fait que le complexe condensine du DCC diffère du complexe condensine de mitose de seulement une sous-unité. Bien que les mécanismes moléculaires exacts liant la condensine à la répression du chromosome X soient encore mal compris, l'idée partagée est que ce processus implique une réorganisation à haut degré du chromosome X dans le noyau.

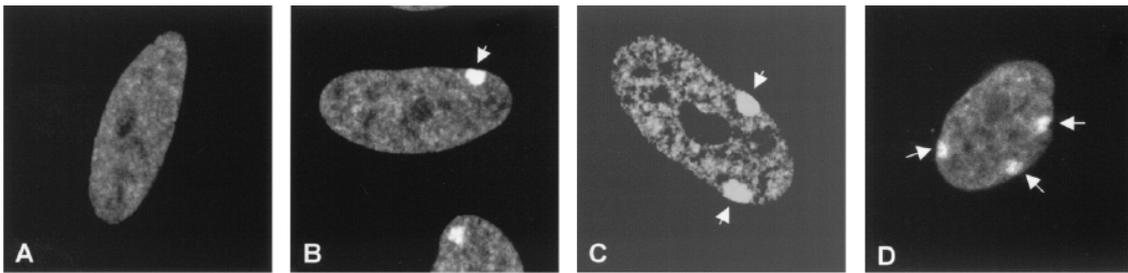
De plus, il semblerait que le recrutement et la dispersion du complexe DCC s'accomplissent par l'intermédiaire de deux sortes de sites de fixation. Les sites *rex* (recruitment elements on X) recruteraient le DCC de manière autonome et dépendante d'une séquence ADN enrichie sur le chromosome X. Il s'agit de la séquence consensus MEX (Motif Enriched on X) : (T/A)CGCG(C/A)AGGGA(G/C). La dispersion du complexe sur le reste du chromosome s'effectuerait ensuite par la fixation aux sites *dox* (dependent on X) qui ne possèdent pas la même capacité de recrutement que les sites *rex* (Jans *et al.*, 2009) (**Figure 20**).



On retrouve chez la drosophile et chez le ver des similarités quant au ciblage du chromosome X et à la fixation du complexe de compensation de dosage. En effet, comme nous l'avons évoqué précédemment, le complexe MSL de la drosophile est recruté sur le chromosome X en se fixant sur des sites d'entrée de la chromatine de manière ADN-dépendante. La similarité se poursuit avec la dispersion du complexe qui apparaît indépendante de la séquence grâce à des sites secondaires.

### II.2.3 Chez les mammifères

Chez les mammifères, la compensation de dosage se traduit par l'inactivation de l'un des deux chromosomes X chez la femelle (cf. **Figure 17**). La présence d'un chromosome X inactif (Xi) a été observée pour la première fois par Barr en 1949 qui identifia un corps plus dense à la coloration de Nissl, souvent localisé dans la région périnucléolaire ou à la périphérie nucléaire (**Figure 21**) (Barr et Bertram, 1949).



**Figure 21 : Corps de Barr visualisés par immunofluorescence**

Les cellules utilisées sont des fibroblastes humains diploïdes qui se différencient par le nombre de leurs chromosomes X (respectivement 1, 2, 3 et 4 pour les photos A, B, C et D). Les flèches pointent sur les corpuscules de Barr.

D'après Hong *et al.*, 2001

Cependant, l'hypothèse de l'inactivation de l'un des deux chromosomes X chez les cellules de mammifères femelles a été pour la première fois proposée par Mary Lyon en 1961 (Lyon, 1961). Elle est fondée sur 3 observations : (i) dans chaque cellule de mammifère femelle, seul un chromosome X est actif et le chromosome inactif est celui qui apparaît hétéropycnotique ; (ii) l'inactivation se produit au début du développement embryonnaire ; (iii) le chromosome inactivé est aléatoirement le maternel ou le paternel. Elle a observé l'effet de mutations dans des gènes liés au chromosome X qui régissent la couleur du pelage chez la souris, entre mâles et femelles. Tandis que les mâles porteurs d'une de ces mutations ont un pelage uniforme, les femelles présentent un pelage mixte, de phénotype sauvage par endroits, de phénotype mutant

à d'autres. Cela a conduit à l'hypothèse que chez les mammifères, seul un chromosome X était actif par lot d'autosomes, le(s) autre(s) étant inactivé(s) suite à une étape de comptage (Lyon, 1961 ; Ohno *et al.*, 1964).

En parallèle, il a été observé qu'un nombre de chromosomes X supérieurs à 2 entraînait de multiples zones de chromatine sexuelle condensée (Moore, 1966) et que, chez les femelles de rat et de souris, un chromosome X était condensé et hétéropycnotique tandis que l'autre était non condensé et d'apparence similaire aux autosomes.

Les mammifères peuvent être divisés en trois catégories en ce qui concerne l'inactivation : (i) la souris et la vache ; (ii) l'homme - ces espèces faisant partie des euthériens - et (iii) les metathériens, dont les marsupiaux. Dans tous les cas, il s'agit de l'inactivation de l'un des deux chromosomes X chez la femelle et l'un des acteurs principaux est un long ARN non-codant.

Chez la souris, cet ARN est l'ARN Xist (X-inactive specific transcript), transcrit à partir du chromosome qui sera inactivé. Nous reviendrons en détails dans le paragraphe III.2.1 sur le processus de l'inactivation tel qu'il est connu à ce jour. Notons simplement ici que l'inactivation chez la souris se déroule en 3 étapes : tout d'abord une inactivation soumise à empreinte dans toutes les cellules où le chromosome paternel est inactivé, puis une réactivation temporaire de ce chromosome avant une nouvelle inactivation distincte selon les tissus. Les tissus extra-embryonnaires conservent une inactivation soumise à empreinte tandis que les autres tissus présentent une inactivation aléatoirement maternelle ou paternelle.

Chez l'homme, l'ARN est également l'ARN Xist. Cependant, les dernières études ont montré que l'inactivation, même dans les tissus extra-embryonnaires tels que le placenta, se ferait de manière aléatoire, sans étape d'inactivation du chromosome X paternel (Moreira de Mello *et al.*, 2010).

Chez les marsupiaux, l'inactivation est partielle, variable selon les tissus et ne concerne que le chromosome X d'origine paternelle (Cooper *et al.*, 1993 ; Wakefield *et al.*, 1997). L'ARN Xist n'est pas retrouvé et l'initiation de l'inactivation est restée une grande question pendant des années. Nous ne détaillerons pas l'inactivation chez ces mammifères, mais il vient d'être découvert très récemment un ARN non-codant possédant des propriétés en adéquation avec un rôle dans l'inactivation du chromosome X. Il s'agit de l'ARN R<sub>sx</sub> (RNA-on-the-silent X) et il dispose des mêmes caractéristiques que Xist, à savoir qu'il est long (27 kb environ), riche en régions répétées, et surtout il n'est exprimé que chez les femelles, à

partir du futur Xi qu'il recouvre (Grant *et al.*, 2012). Cependant Rsx et Xist ne sont pas homologues, même si une région située à la position 3307 de Rsx est composée de 9 répétitions d'une séquence de 34 nucléotides qui ressemble globalement à une région située en 5' de Xist (la région des A-repeats, cf. paragraphe IV.2.2) et dont la structure pourrait être similaire également. Cette découverte apporte de nouvelles connaissances quant au mécanisme d'inactivation chez les métathériens, mais également de nouvelles questions, notamment sur l'évolution convergente de l'inactivation entre métathériens et euthériens.

### III. L'inactivation du chromosome X chez les mammifères

Après ces premières observations de l'inactivation du chromosome X, il restait à en comprendre les mécanismes moléculaires (hétérochromatinisation, extinction transcriptionnelle d'un chromosome entier, choix du chromosome à inactiver) ainsi que les acteurs. La caractérisation du centre d'inactivation du chromosome X puis de l'ARN Xist ont permis de grandes avancées dans cette compréhension. D'autre part, l'essor de l'épigénétique dans les années 2000 a quelque peu refaçoné la vision du phénomène d'inactivation.

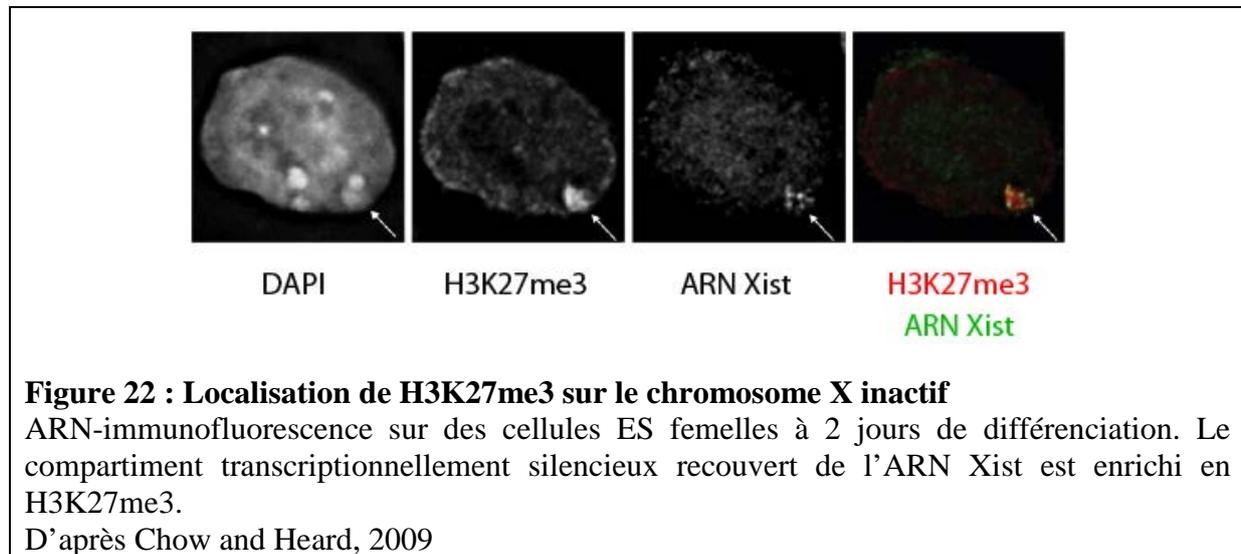
Le principal modèle utilisé pour l'étude de l'inactivation chez les mammifères est celui des cellules souches embryonnaires (cellules ES). Isolées à partir de la masse interne du blastocyste, elles peuvent être cultivées en présence du facteur LIF (Leukemia Inhibitory Factor). Elles sont alors pluripotentes et immortelles puisque le facteur LIF active une voie de signalisation inhibant la différenciation (pour revue, Burdon *et al.*, 2002). L'élimination du LIF du milieu de culture déclenche la différenciation des cellules ES et permet ainsi l'inactivation du chromosome X (pour revue, Fluckiger *et al.*, 2003).

#### III.1 Les caractéristiques du chromosome X inactif

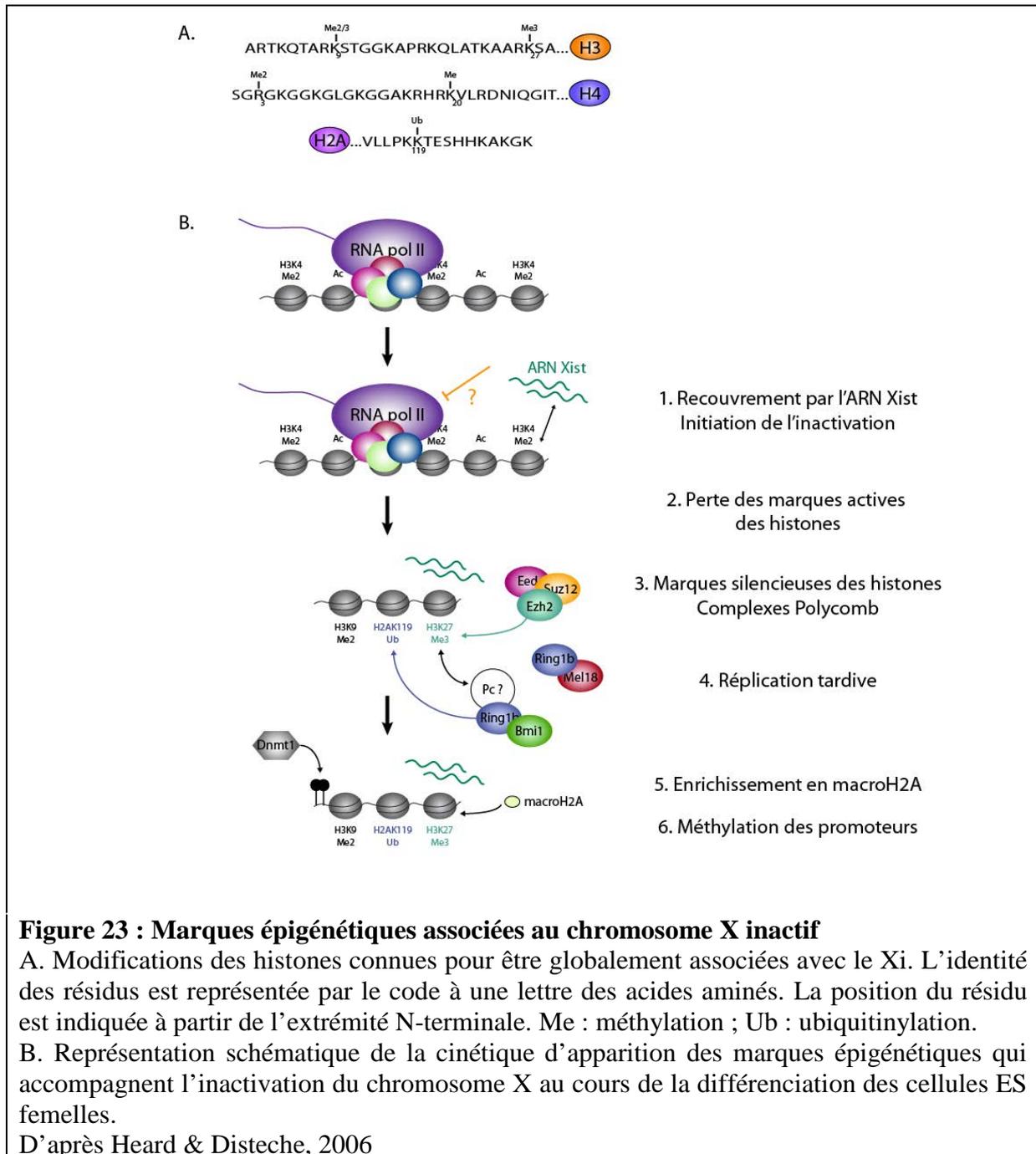
##### III.1.1 Les marques épigénétiques du Xi

L'accumulation de l'ARN Xist est le premier événement de l'initiation de l'inactivation. Elle constitue un élément déclencheur pour une cascade d'événements aboutissant à la mise en place de la répression transcriptionnelle stable du chromosome X, dont la méthylation des histones, modification importante dans la régulation de l'expression des gènes, comme nous l'avons vu précédemment. En effet, le chromosome X inactif (Xi), par rapport au X actif (Xa), est spécifiquement enrichi en modifications telles que H3K27me3 (Rougeulle *et al.*, 2004), H3K9me3, H4K20m1 (Kohlmaier *et al.*, 2004), H2AK119ub1

(Hernandez-Munoz *et al.*, 2005). Notons que la marque H3K27me3 est tellement enrichie sur le Xi qu'elle permet de localiser ce chromosome (**Figure 22**).



De même, il a été observé une diminution de H3K9ac, H3K4me3 et me4 ainsi qu'une déacétylation globale de H4 (Keohane *et al.*, 1996 ; Heard *et al.*, 2001) (**Figure 23**).



### III.1.2 La réplication tardive

Contrairement au chromosome X actif qui est répliqué en début de phase S, le Xi est répliqué à la fin. Dans les cellules ES femelles, la réplication tardive du Xi est détectable dès 2 jours après le début de la différenciation, ce qui correspond à la période d'induction de l'expression de l'ARN Xist ainsi qu'au silence transcriptionnel des gènes liés à l'X (Keohane *et al.*, 1996). De plus, il a été montré qu'une fois l'inactivation mise en place, la délétion conditionnelle du gène *Xist* sur le Xi n'a pas de conséquence sur la réplication tardive de ce

dernier (Csankovszki *et al.*, 1999). Ceci illustre le fait qu'après l'initiation de l'inactivation, la réplication asynchrone ainsi que les autres caractéristiques du Xi sont maintenues indépendamment de l'ARN Xist. De manière étonnante, la réplication du Xi se déroule avant celle du Xa dans les tissus extra-embryonnaires (Sugawara *et al.*, 1983).

### III.1.3 La méthylation de l'ADN

La méthylation du Xi est l'un des événements les plus tardifs de la mise en place de l'inactivation. Les îlots CpG sont fortement méthylés sur le Xi des cellules somatiques tandis qu'ils sont non méthylés sur le Xa (et ils sont non méthylés sur les 2 chromosomes dans le cas des gènes qui échappent à l'inactivation, qui sont le sujet du paragraphe suivant).

De plus, bien que la méthylation de l'ADN soit sans conteste un élément régulateur de l'expression des gènes et qu'elle soit observée sur le Xi, le chromosome X peut tout de même adopter une conformation d'état inactif en absence de méthylation des îlots CpG. La méthylation pourrait alors intervenir dans la diminution de l'accessibilité de la chromatine aux facteurs de transcription. La méthylation est plus ambiguë chez l'Homme car le Xi pourrait être hypométhylé par rapport au Xa, ce qui pose un problème dans la conception actuelle des mécanismes engagés.

### III.1.4 Les gènes échappant à l'inactivation

Même si l'on parle globalement de l'inactivation du chromosome X, tous les gènes liés à ce chromosome ne sont pas inactivés et sont donc exprimés à partir du Xa et du Xi. Environ 15% des gènes liés à l'X échappent à l'inactivation et la proportion de ces gènes diffère de beaucoup selon les régions du chromosome, ce qui reflète un lien avec l'évolution des chromosomes sexuels (Carrel et Willard, 2005). Cette étude montre également la présence de gènes présentant un profil variable d'expression (10%). A part quelques exceptions, la majorité des gènes liés à l'X homologues du chromosome Y échappent à l'inactivation, ce qui assure dans ce cas une expression à taux équivalents entre mâles et femelles. Chez l'homme, les gènes non soumis à l'inactivation sont regroupés au sein de régions échappant à l'hétérochromatinisation (Carrel et Willard, 2005), ce qui étaye l'idée que le contrôle de l'expression se situe au niveau de l'établissement de régions chromosomiques régulées de manière coordonnée. Chez la souris, les gènes échappant à l'inactivation semblent moins nombreux (environ 3 %) et l'expression à partir du Xi suggère qu'ils y échappent seulement de manière partielle (Berletch *et al.*, 2010). Les gènes échappant à l'inactivation sur le Xi

arborent les mêmes caractéristiques que ceux présents sur le Xa, telles qu'une réplication synchrone ou encore une hypométhylation des îlots CpG.

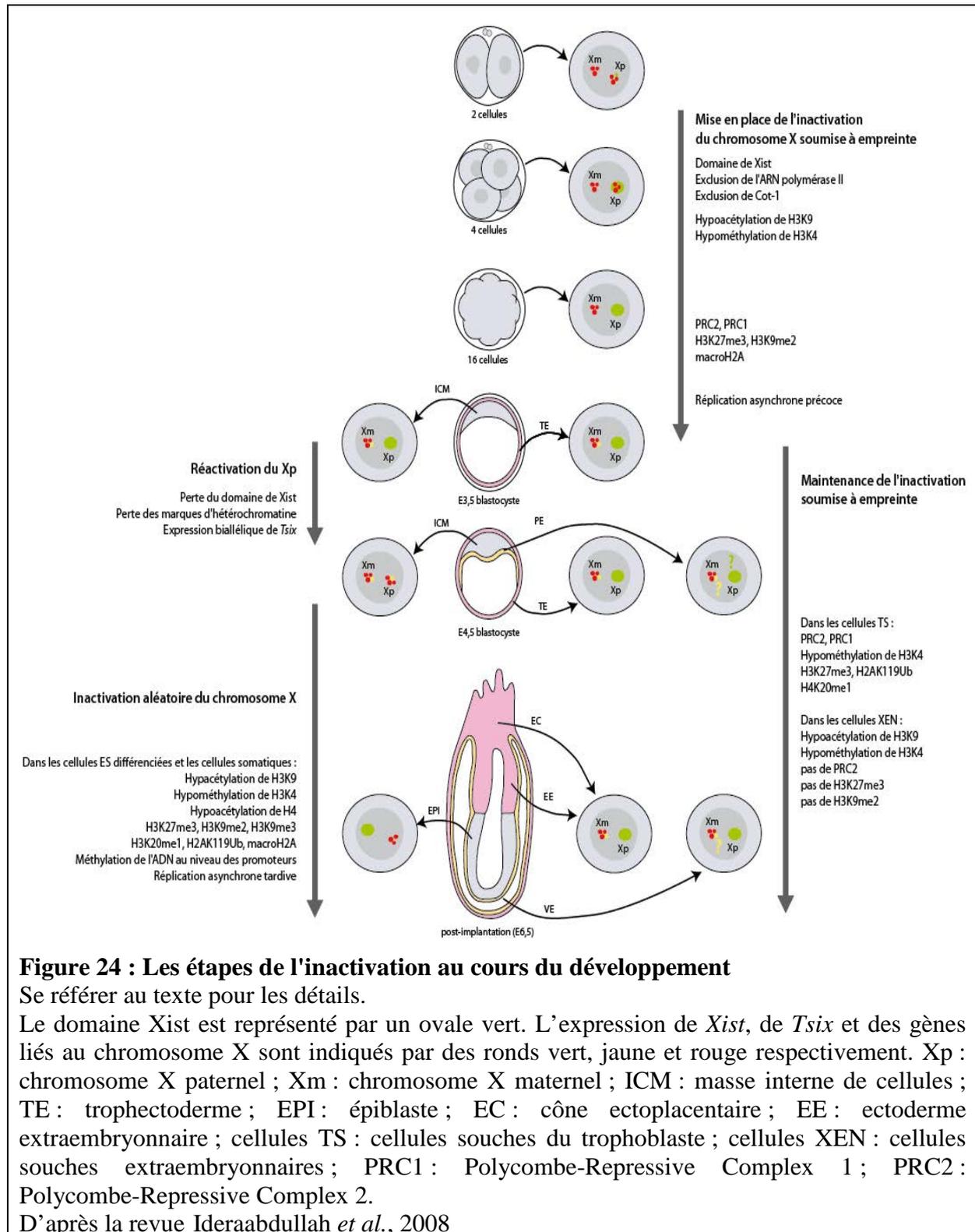
### III.2 Les mécanismes de l'inactivation

Nous avons vu dans les chapitres précédents que l'inactivation du chromosome X est un phénomène complexe nécessitant la mise en place de multiples événements. Nous allons désormais présenter les différentes étapes de l'initiation de l'inactivation, notamment l'expression de l'ARN *Xist*.

#### III.2.1 Les étapes de l'inactivation au cours de l'embryogénèse chez la souris

Chez la souris, l'inactivation du chromosome X s'effectue en deux phases au cours du développement. La première a lieu durant la phase de préimplantation, c'est-à-dire dans les étapes les plus précoces du développement. Elle est soumise à l'empreinte parentale et résulte en l'inactivation du chromosome X paternel (Xp). Bien qu'il ait été reporté que le Xp soit hérité dans un état déjà partiellement inactif (Huynh et Lee, 2003), d'autres preuves suggèrent que les deux chromosomes X sont actifs dans les embryons femelles après l'activation du génome zygotique suivi de l'initiation de l'inactivation par empreinte au stade 4-8 cellules (Okamoto *et al.*, 2004 ; Okamoto *et al.*, 2005). Il semblerait que les modifications épigénétiques spécifiques des parents, qui sont établies lors de la gamétogénèse, régulent l'inactivation par empreinte. Ces modifications impliquent probablement une marque qui réprime *Xist* sur le chromosome X maternel (Xm). Dans le cas de l'inactivation par empreinte, contrairement à l'inactivation aléatoire comme nous le verrons par la suite, la répression de *Xist* sur le Xm est indépendante de *Tsix*.

La répression du Xp est maintenue dans les tissus extra-embryonnaires qui proviennent du trophoctoderme et de l'endoderme primitif du stade blastocyste (Takagi et Sasaki, 1975). En revanche, dans les cellules pluripotentes de la masse interne de cellules (Inner Mass Cell, ICM), le chromosome X paternel précédemment inactivé retrouve provisoirement une activité transcriptionnelle. Ensuite, une inactivation aléatoire du chromosome X maternel ou paternel se met en place lors de la différenciation (**Figure 24**). Cette inactivation est stable puisqu'elle est transmissible aux générations suivantes, toutes les cellules filles provenant d'une même cellule mère ayant le même chromosome X inactivé.



**Figure 24 : Les étapes de l'inactivation au cours du développement**

Se référer au texte pour les détails.

Le domaine Xist est représenté par un ovale vert. L'expression de *Xist*, de *Tsix* et des gènes liés au chromosome X sont indiqués par des ronds vert, jaune et rouge respectivement. Xp : chromosome X paternel ; Xm : chromosome X maternel ; ICM : masse interne de cellules ; TE : trophoctoderme ; EPI : épiblaste ; EC : cône ectoplacentaire ; EE : ectoderme extraembryonnaire ; cellules TS : cellules souches du trophoblaste ; cellules XEN : cellules souches extraembryonnaires ; PRC1 : Polycombe-Repressive Complex 1 ; PRC2 : Polycombe-Repressive Complex 2.

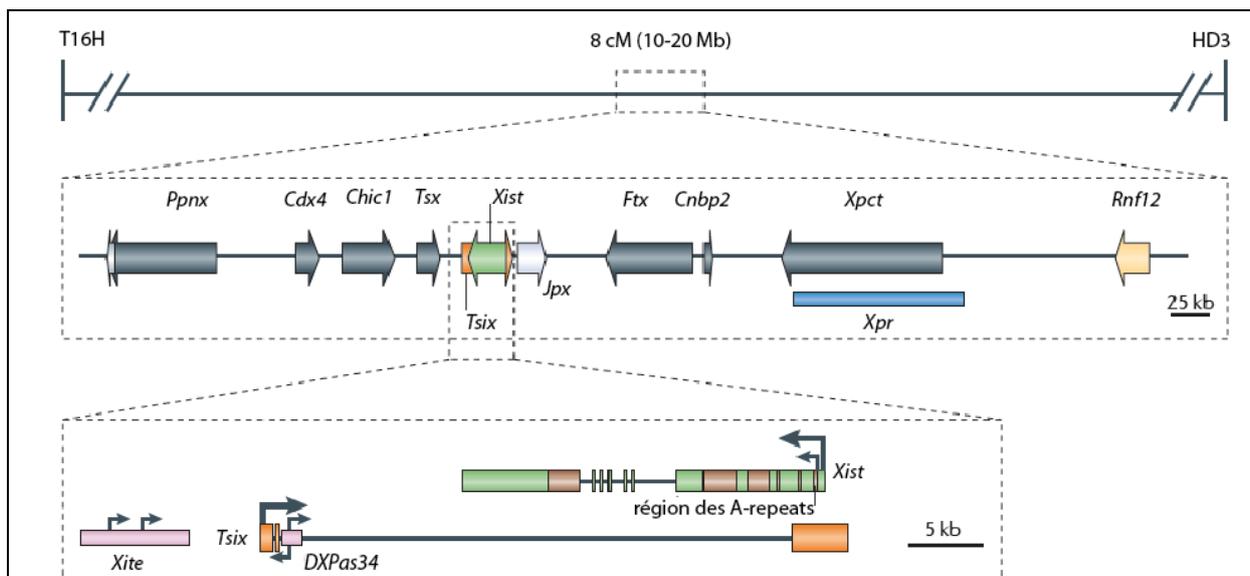
D'après la revue Ideraabdullah *et al.*, 2008

### III.2.2 Le centre d'inactivation du chromosome X (XIC)

L'initiation de l'inactivation se déclenche après une évaluation du nombre de chromosomes X par rapport au nombre d'autosomes puis un choix du chromosome X à inactiver. La répression transcriptionnelle se met alors en place sur le chromosome

sélectionné. Ces étapes dépendent d'un locus appelé XIC (X-inactivation center). Il a été montré par insertion sur un chromosome autosome que cette région porte les éléments nécessaires et suffisants pour initier l'inactivation de ce chromosome entier (Lee *et al.*, 1996 ; Lee et Jaenisch, 1997).

La région XIC est longue d'environ 1 Mb et elle est localisée entre les points T16H et HD3, sur le bras long proximal du chromosome X murin. Elle contient plusieurs gènes dont le rôle dans l'inactivation n'est pas déterminé pour tous (**Figure 25**). Elle porte ainsi le gène *Xist* (X-inactivation specific transcript), identifié dès le début des années 1990 (Borsani *et al.*, 1991 ; Brockdorff *et al.*, 1991 ; Brown *et al.*, 1991), produisant un long ARN non-codant qui est un élément clé dans l'initiation de l'inactivation. Au niveau de cette région se trouvent également les gènes *Tsix* (gène antisens de *Xist*) et *Rnf12* (RING finger protein 12), connus pour être impliqués dans la régulation de *Xist* (Jonkers *et al.*, 2009). Nous reviendrons sur l'ARN *Xist* et ses régulations dans les paragraphes suivants. De même, les locus *Xite* (X-inactivation intergenic transcription element) et *DXPas34* ont été identifiés comme des régulateurs de *Tsix* (Ogawa et Lee, 2003 ; Stavropoulos *et al.*, 2005).



**Figure 25 : Le centre d'inactivation du chromosome X (XIC) murin**

La région XIC est située entre les points T16H et HD3, une région couvrant 8 cM (10-20 Mb). Seuls les éléments avoisinants le gène *Xist* sont représentés ici. Certains de ces éléments, comme le gène antisens de *Xist* (*Tsix*) ou le gène de la protéine *Rnf12* sont connus pour être impliqués dans la régulation de *Xist*. *Xist* et son antisens *Tsix*, ainsi que *Xite* (X-inactivation intergenic transcription element), le régulateur de *Tsix*, et *DXPas34*, sont représentés plus en détails en-dessous.

*Cdx4*, caudal X-linked gene 4 ; *Chic1*, cysteine rich hydrophobic 1 ; *Cnbp2*, cellular nucleic acid binding protein 2 ; *Ftx*, five prime to *Xist* ; *Jpx*, ou *Enox* (expressed neighbour of *Xist*) ; *Tsx*, testis specific X-linked ; *Xpct*, X-linked PEST-containing transporter.

D'après Augui *et al.*, 2011

L'organisation structurale du XIC, bien que ce soit une région longue de plusieurs Mb, n'est pas aléatoire. Très récemment, des expériences de conformation de chromosome sur une région de 4,5 Mb, incluant *Xist* et la plupart des loci connus pour être impliqués dans l'inactivation du chromosome X (correspondant à la première région encadrée de la **Figure 25**), ont été effectuées (Nora *et al.*, 2012). Il en ressort que la conformation de la chromatine partitionne cette région en une succession de domaines topologiquement associés (TADs, topologically associated domains). Ces domaines permettent, par la formation de boucles, le rapprochement d'éléments régulateurs avec les éléments qu'ils régulent, tels que *Tsix*, son promoteur et son activateur *Xite*. Ils ont ainsi découvert l'existence d'un TAD d'environ 550 kb, contenant *Xist* et tous ses activateurs connus (*Jpx*, *Ftx*, *Xpr/Xpct* et *Rnf12*) (Nora *et al.*, 2012). Ce domaine TAD a la particularité d'être en corrélation inverse avec la plupart des autres gènes de la région étudiée puisque les transcrits issus de ce domaine sont augmentés au cours de la différenciation. Ceci suggère que les éléments présents sont intégrés dans un même réseau de régulation en *cis*. Ils ont alors proposé que ce TAD représente la région régulatrice 5' minimale pour une expression de *Xist* optimale.

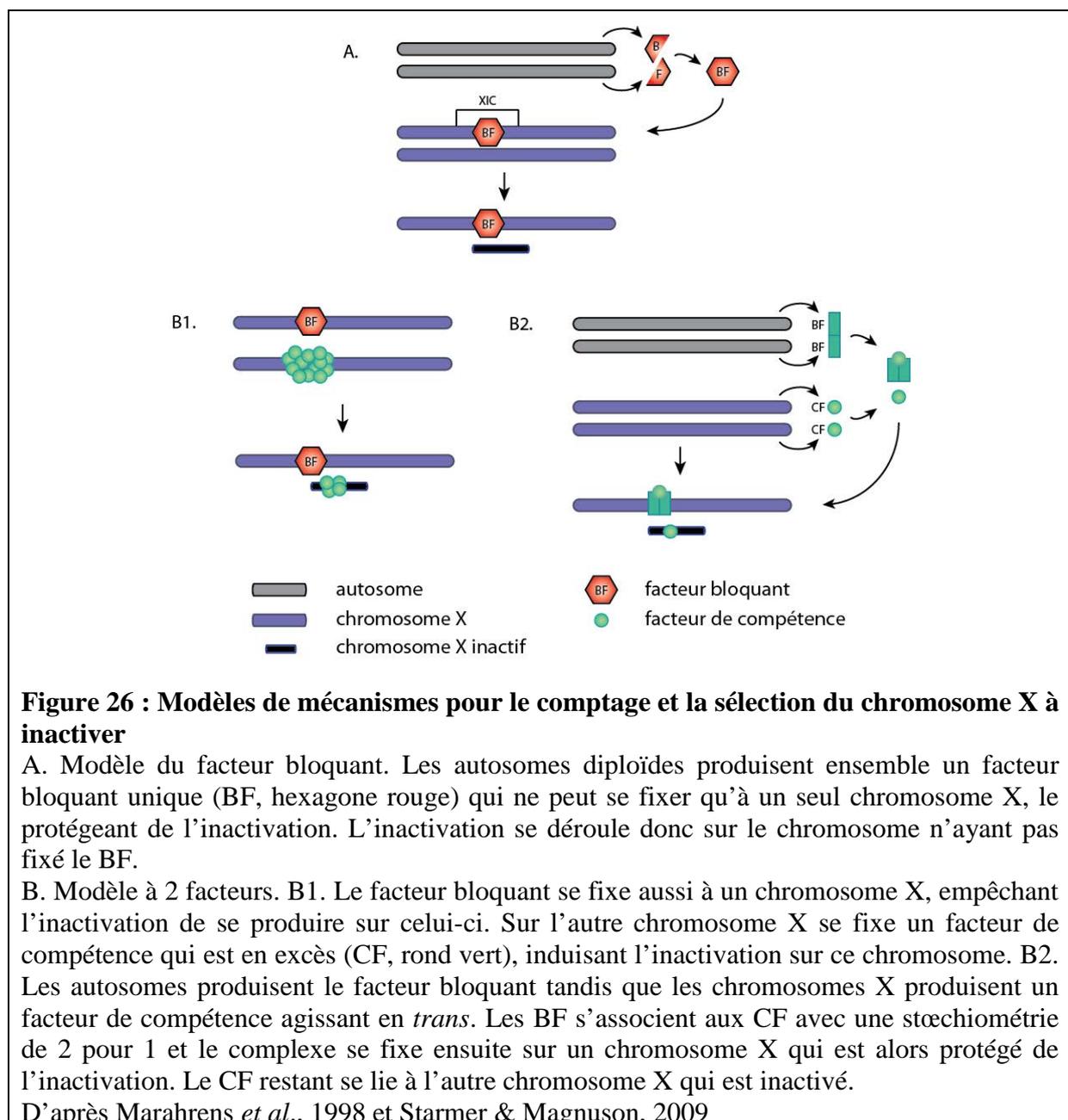
### III.2.3 Les étapes de l'inactivation aléatoire

Chez les mammifères, la mise en place de l'inactivation peut être divisée en cinq étapes : le comptage, le choix d'un des deux X, l'initiation, l'établissement et enfin la maintenance de l'état transcriptionnellement silencieux. Ces différentes étapes sont sous la régie du locus XIC.

#### a. Le comptage et le choix du chromosome X à inactiver

Les individus possédant un ou plusieurs chromosomes X surnuméraires par rapport au nombre d'autosomes (par exemple AA:XXXX) présentent malgré tout un seul X actif (Giannelli, 1963 ; Grumbach *et al.*, 1963). De plus, pour un génotype AAAA:XXXX, seuls 2 chromosomes X sont inactivés, ce qui revient également à un chromosome X pour une paire d'autosomes. Pour cela, la cellule doit procéder à une évaluation du nombre relatif de chromosomes X par rapport au nombre d'autosomes. L'inactivation qui s'en suit repose sur la règle selon laquelle le ratio du nombre de X actifs par rapport au nombre d'autosomes doit être égal à  $\frac{1}{2}$  (Rastan, 1983). Les étapes de comptage et de choix sont étroitement liées puisqu'elles ont lieu de manière quasi concomitante, les mécanismes mis en jeu pour le comptage permettant simultanément le choix du futur Xi. Plusieurs modèles concernant ces

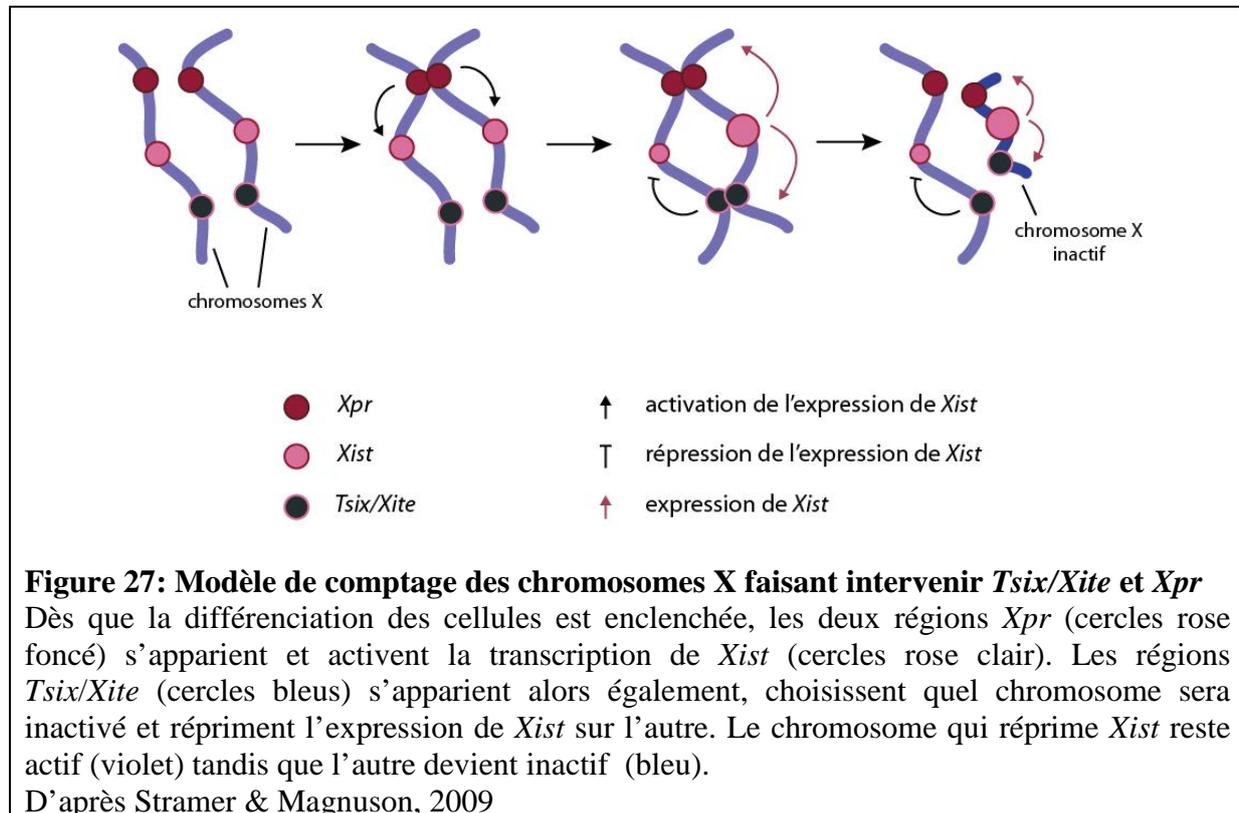
mécanismes ont été proposés. Le modèle du facteur bloquant (BF, blocking factor) présume qu'un facteur limitant bloque l'inactivation de l'un des chromosomes X par cellule diploïde, que ce soit l'un des 2 chromosomes X chez la femelle ou l'unique chromosome X chez le mâle (Rastan, 1983). Le phénomène de comptage vient du fait que ce facteur serait encodé par un gène autosomal. Ainsi l'excès d'autosomes par rapport aux chromosomes X détermine s'il y a suffisamment de facteur bloquant pour inhiber l'inactivation sur l'un des X dans un contexte diploïde. Ce facteur bloquant se fixerait sur un élément du XIC qui empêcherait l'ARN Xist de s'accumuler sur ce chromosome (**Figure 26.A**). La région du XIC qui fixerait le BF engloberait les locus *DXPas34* et *Xite*, connus pour réguler l'expression de *Tsix* (Ogawa et Lee, 2003 ; Stavropoulos *et al.*, 2005).



Cependant, les cellules avec une délétion du gène *Xist* sur un seul des 2 chromosomes X font toujours le « bon choix » et inactivent le chromosome non délété (Marahrens *et al.*, 1998 ; Gribnau *et al.*, 2005). Ce cas de figure est impossible dans l'hypothèse du modèle précédent : supposons que le facteur bloquant se fixe sur le XIC du chromosome non délété, il empêcherait donc son inactivation. Or l'inactivation ne peut pas non plus se mettre en place sur l'autre chromosome, car il est délété du gène *Xist*. Il en résulterait la mort de la moitié de la population cellulaire, ce qui n'est pas le cas. C'est pourquoi la présence d'un autre facteur a été proposée, un facteur favorisant l'inactivation cette fois-ci, appelé facteur de compétence (CF, competence factor). Un modèle a ainsi émergé, combinant l'action du BF et du CF, le CF en excès inactivant tous les chromosomes X présents dans la cellule à l'exception de celui sur lequel le BF serait fixé (Marahrens *et al.*, 1998) (**Figure 26.B1**). Un second modèle faisant lui aussi intervenir les deux types de facteurs propose que le CF, codé par le chromosome X, soit limitant et titré par le BF avec un ratio 1:2. Si plusieurs chromosomes X sont présents par cellule diploïde, le ratio 1:2 n'est plus à l'équilibre et les CF ne peuvent plus être tous titrés. Il en résulterait que les CF en excès initieraient l'inactivation des chromosomes X supplémentaires (Lee, 2005) (**Figure 26.B2**).

Par ailleurs, des expériences de délétion chez la souris d'une région de 65 kb se trouvant en aval de l'exon 6 de *Xist* ont eu pour conséquence une forte réduction de l'expression de *Xist* dans des cellules ES XX non différenciées et une inactivation du chromosome X dans des cellules ES XO différenciées, malgré l'absence d'un second XIC, en principe nécessaire pour l'étape de comptage (Clerc et Avner, 1998). C'est à partir de ce moment qu'un élément contenu dans cette région a été recherché pour son rôle dans le processus de comptage. Il semblait notamment empêcher l'inactivation lorsqu'un seul chromosome X est présent dans des cellules diploïdes, et ceci par un mécanisme de répression de l'expression de *Xist*. Cet élément est l'ARN *Tsix*, issu d'un gène antisens à *Xist* (Lee *et al.*, 1999). Son promoteur se trouve 15 kb en aval de *Xist*, et sa transcription résulte en un produit de 40 kb qui recouvre donc le locus *Xist*. De plus, il a été observé une colocalisation des régions XIC des deux chromosomes X préalablement à l'inactivation. Certaines des régions impliquées dans cet appariement transitoire ont été identifiées. Tout d'abord la région comprenant justement les gènes *Tsix* et *Xite* (Bacher *et al.*, 2006 ; Xu *et al.*, 2006 ; Xu *et al.*, 2007). Une autre région a été découverte comme s'appariant avec des copies d'elle-même et induisant la transcription de *Xist* dans des cellules ES mâles (Augui *et al.*, 2007). Cette région

se situe environ 400 kb en aval de *Xist* et elle a été nommée X-pairing, *Xpr*. Un autre modèle de comptage et de choix a donc été proposé. Il suggère que *Xpr* permettrait de repérer la présence d'une autre (ou plusieurs autres) *Xpr* et donc l'initiation de l'inactivation. En parallèle, l'appariement des régions de *Tsix* et *Xite* de chaque chromosome X induirait un mécanisme empêchant l'inactivation du second chromosome (Augui *et al.*, 2007) (**Figure 27**).



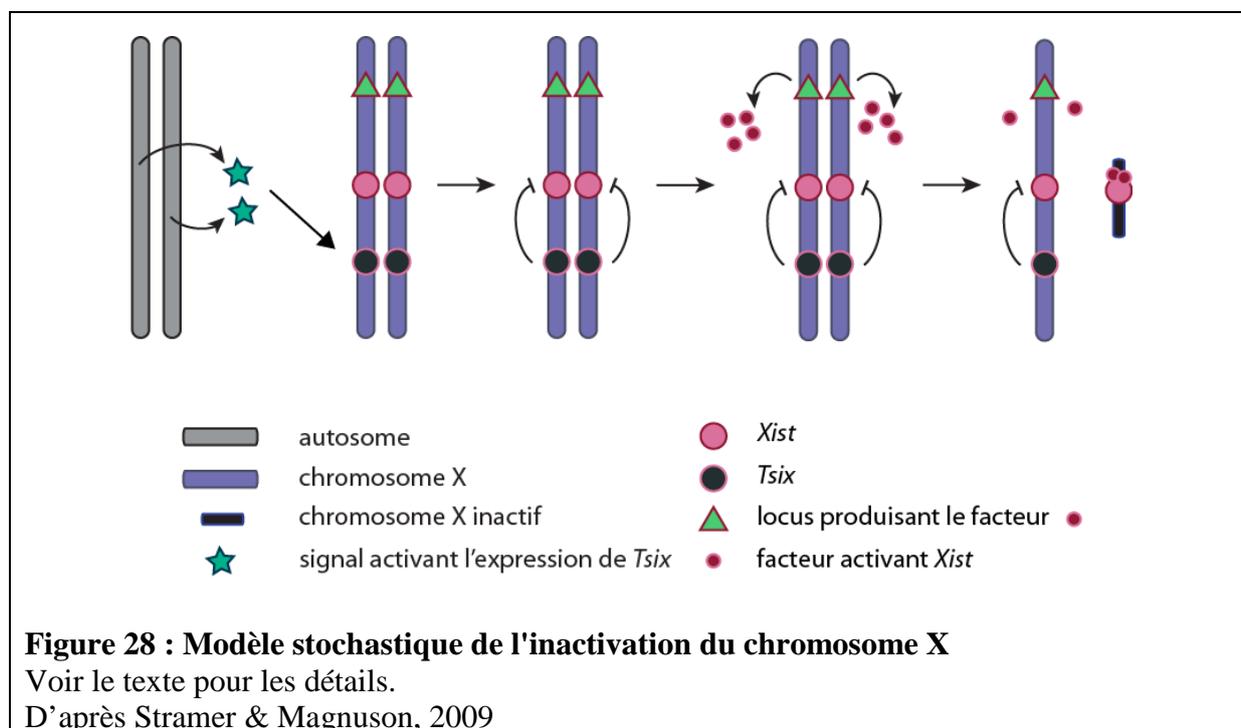
**Figure 27: Modèle de comptage des chromosomes X faisant intervenir *Tsix/Xite* et *Xpr***

Dès que la différenciation des cellules est enclenchée, les deux régions *Xpr* (cercles rose foncé) s'apparient et activent la transcription de *Xist* (cercles rose clair). Les régions *Tsix/Xite* (cercles bleus) s'apparient alors également, choisissent quel chromosome sera inactivé et répriment l'expression de *Xist* sur l'autre. Le chromosome qui réprime *Xist* reste actif (violet) tandis que l'autre devient inactif (bleu).

D'après Stramer & Magnuson, 2009

Les régions du XIC ne sont pas suffisantes pour induire de tels phénomènes à elles seules, la présence de facteurs protéiques est également indispensable. On peut citer le facteur CTCF (CCCTC-binding factor) qui aurait de multiples fonctions dans l'inactivation du chromosome X. En effet, il joue de nombreux rôles dans la régulation de la transcription mais aussi dans l'architecture nucléaire et il permettrait de séparer les gènes inactivés de ceux échappant à l'inactivation (Filippova *et al.*, 2005). Il pourrait également participer au choix du futur Xi en facilitant la formation d'une structure particulière de la chromatine sur l'un des chromosomes X, ce qui la rendrait conforme pour la transcription de *Xist* (Pugacheva *et al.*, 2005). De plus, une concentration élevée de ce facteur a été observée sur le Xi, où il est associé au promoteur du gène *Xist*, mais aussi à celui de *Tsix*. Il a été proposé que CTCF permet d'activer la transcription de *Tsix* sur le Xa (Chao *et al.*, 2002). De plus, des sites de fixation de cette protéine ont été identifiés dans le gène *Tsix* (Xu *et al.*, 2007). Il a ainsi été

proposé un modèle stochastique pour le comptage et le choix du Xi. La première étape de ce modèle est la production par les autosomes d'un facteur agissant en *trans* pour activer l'expression de *Tsix*. Au moment de la différenciation, le chromosome X, lui, exprime un facteur agissant en *trans* pour l'activation de *Xist*. La probabilité que le XIC initie l'inactivation dépend de la concentration de chaque facteur sur chaque chromosome X. Si la concentration du facteur activant *Xist* est suffisante, les cellules créent une opportunité pour l'ARN *Xist* de s'accumuler et d'initier l'inactivation, réprimant ainsi *Tsix* et les gènes liés au chromosome X en *cis*, incluant le gène codant le facteur activateur de *Xist* (Monkhorst *et al.*, 2008) (**Figure 28**).



Ce modèle ne correspondant pas encore exactement aux observations faites *in vivo*, un nouveau modèle a été proposé par Stramer & Magnuson (2009). Dans ce modèle appelé modèle par « retour d'information », le chromosome X produit un facteur agissant en *trans* qui se fixe sur des sites spécifiques se trouvant sur des autosomes. Une fois tous les sites saturés, les autosomes produisent alors un facteur inhibiteur en grande quantité. Ces facteurs se lient entre eux et aux XIC, et lorsqu'un chromosome X a accumulé suffisamment de facteurs inhibiteurs, le XIC initie l'inactivation. Seul un chromosome X reste actif et délivre encore le facteur mais en quantité insuffisante pour déclencher la production de facteur inhibiteur par les autosomes.

### b. L'initiation de l'inactivation

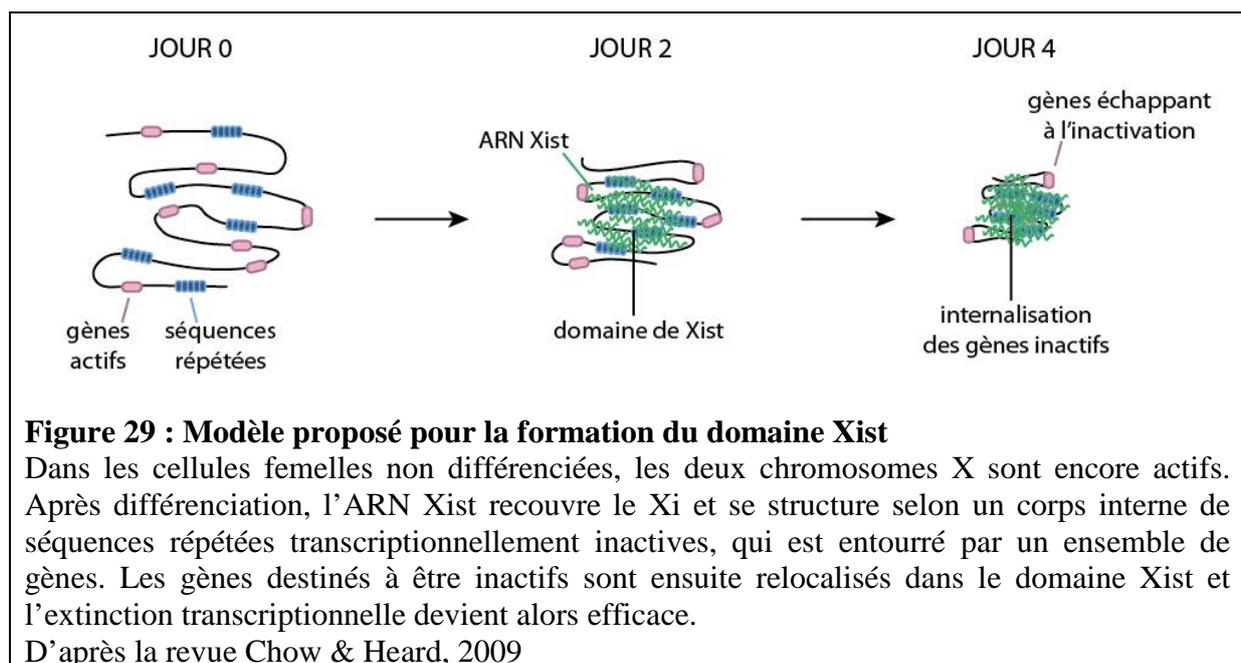
L'inactivation du chromosome X est initiée par la surexpression de l'ARN *Xist*. Elle démarre non seulement très tôt au cours du développement embryonnaire, mais également dans une fenêtre de temps courte et précise après la différenciation cellulaire. Il a été montré qu'une surexpression de *Xist* grâce à un transgène sous contrôle d'un promoteur inductible dans des cellules ES déjà différenciées n'induit aucune inactivation du chromosome porteur du transgène, contrairement à des cellules encore indifférenciées (Wutz et Jaenisch, 2000). L'inactivation démarre donc très tôt au cours du développement et elle ne peut plus être initiée après un certain stade. Par induction de *Xist* à différents stades de la différenciation, il a ensuite été possible de réduire la fenêtre d'opportunité à une période allant de l'état indifférencié jusqu'aux 24 premières heures après initiation de la différenciation. Cette répression initiale dépend de l'expression de *Xist* et est réversible. Nous détaillerons les facteurs impliqués dans la régulation de l'expression de *Xist* dans le paragraphe IV.3.1.

### c. La propagation de l'ARN *Xist*

La surexpression de *Xist* n'est pas le seul phénomène observé lors de l'initiation de l'inactivation. L'ARN *Xist* n'est pas exporté dans le cytoplasme mais reste associé au chromosome X à partir duquel il est exprimé et le recouvre en quasi-totalité (rappelons que certains gènes liés au chromosome X restent actifs). La question de savoir comment cet ARN s'associe au futur Xi et comment il se propage le long de celui-ci reste non résolue, même si plusieurs hypothèses ont été émises. Il a par exemple été observé que lorsque des translocations de portions du chromosome X contenant le XIC ont lieu sur un autosome, l'ARN *Xist* est incapable de se propager au-delà du point de cassure de la translocation (Popova *et al.*, 2006). L'absence de propagation à longue distance pourrait s'expliquer par l'existence sur le chromosome X de « relais » ou « tremplins » (*way stations* ou *booster elements*, dans les termes des auteurs), favorisant la propagation sur ce chromosome (Gartler et Riggs, 1983). M. Lyon a de son côté proposé que les éléments répétés LINE-1 (long interspersed element-1) soient des cibles de haute affinité pour la fixation de l'ARN *Xist* (Lyon, 1998). Ces éléments se trouvent d'ailleurs en forte concentration autour du gène *Xist* (Jonkers *et al.*, 2008). Toutefois, suite à la découverte de l'ARN *Rsx* chez l'oppossum, la composition en LINEs du chromosome X de cette espèce a été étudiée et il n'a pas été observé d'enrichissement particulier, alors que *Rsx* peut recouvrir le Xi (Grant *et al.*, 2012). Il

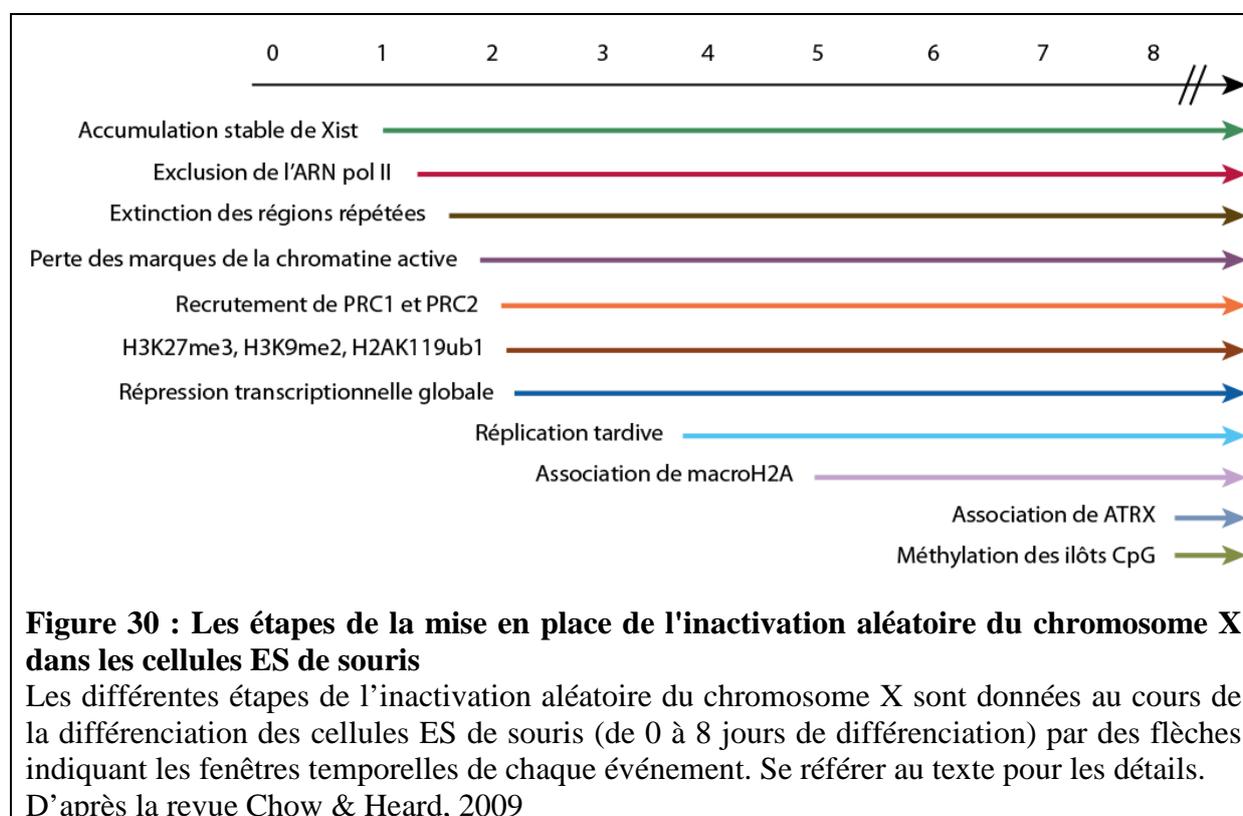
n'est donc pas à exclure que d'autres régions d'ADN ou marques épigénétiques spécifiques pourraient également être impliquées dans la propagation de l'ARN Xist.

En parallèle du recouvrement du futur chromosome X inactif par l'ARN Xist, ce chromosome adopte une conformation condensée jusqu'à la formation d'un compartiment nucléaire répressif pour la transcription. Une hypothèse controversée concernant la structuration de ce compartiment fait intervenir les éléments de la matrice nucléaire (Clemson *et al.*, 1996). Cette proposition est confortée par l'observation d'un enrichissement de la protéine SAF-A (Scaffold Attachment Factor A) sur le Xi. Cet enrichissement dépend du domaine d'interaction de SAF-A avec les ARN, ce qui suggère une interaction fonctionnelle avec Xist (Helbig et Fackelmayer, 2003 ; Fackelmayer, 2005). Ce compartiment a été nommé domaine Xist. Il est structuré d'un corps interne de séquences répétées entouré par une bordure de gènes, le tout recouvert par l'ARN Xist et enrichi en H3K27me3 (Chaumeil *et al.*, 2006 ; Clemson *et al.*, 2006). La mise en place de ce domaine est séquentielle et l'organisation finale influe sur l'expression ou non des gènes puisque ceux situés à l'extérieur de ce domaine échappent à l'inactivation. Il a été proposé que suite au recouvrement du futur Xi par l'ARN Xist, l'ARN polymérase II est exclue de la fraction riche en séquences répétées (celle qui constitue le corps interne du domaine), ce qui conduit à son extinction transcriptionnelle (Chaumeil *et al.*, 2006). Une relocalisation des gènes à l'intérieur de ce domaine Xist permet ensuite un silence transcriptionnel efficace (**Figure 29**). Nous reviendrons sur la mise en place de ce domaine lorsque nous aborderons le rôle de la région des A-repeats dans l'inactivation.



## d. La mise en place des marques épigénétiques et la maintenance de l'état réprimé

L'état inactif du chromosome X devient irréversible 48 heures après le début de la différenciation, même après arrêt de l'expression de *Xist* (Wutz *et al.*, 2002). Après accumulation de l'ARN *Xist* sur le Xi, plusieurs événements surviennent rapidement pour aboutir à un état silencieux stable. Nous avons déjà parlé de l'exclusion de l'ARN polymérase II. De même, des marques épigénétiques caractéristiques de l'euchromatine disparaissent au bénéfice de marques distinctives de l'hétérochromatine. En effet, les H3K9Ac, H3K4me2, H3K4me3 et l'acétylation globale des histones H4 sont perdues (Keohane *et al.*, 1996 ; Heard *et al.*, 2001 ; Chaumeil *et al.*, 2002). Ces marques sont remplacées par des H3K27me3 (Rougeulle *et al.*, 2004), H3K9me2 (Heard *et al.*, 2001 ; Okamoto *et al.*, 2004), H3K9me3, H4K20me1 (Kohlmaier *et al.*, 2004) et H2AK119ub1 (de Napoles *et al.*, 2004 ; Hernandez-Munoz *et al.*, 2005). Parmi les protéines apposant ces modifications, les complexes PRC1 et PRC2 ont été retrouvés sur le Xi dès 48 heures après le début de la différenciation. Les événements survenant ensuite sont entre autres l'association à un variant d'histone, la macroH2A, et à la protéine ATRX. Le dernier élément notable conduisant à un état inactivé stable et irréversible est la méthylation des îlots CpG (Chow et Heard, 2009) (**Figure 30**).



Ces marques apparaissant de manière séquentielle au cours de l'inactivation, il est maintenant admis qu'elles jouent des rôles distincts dans le processus, les marques les plus précoces participant à la mise en place de l'inactivation tandis que les plus tardives permettant le verrouillage de la chromatine dans un état inactif (Kohlmaier *et al.*, 2004).

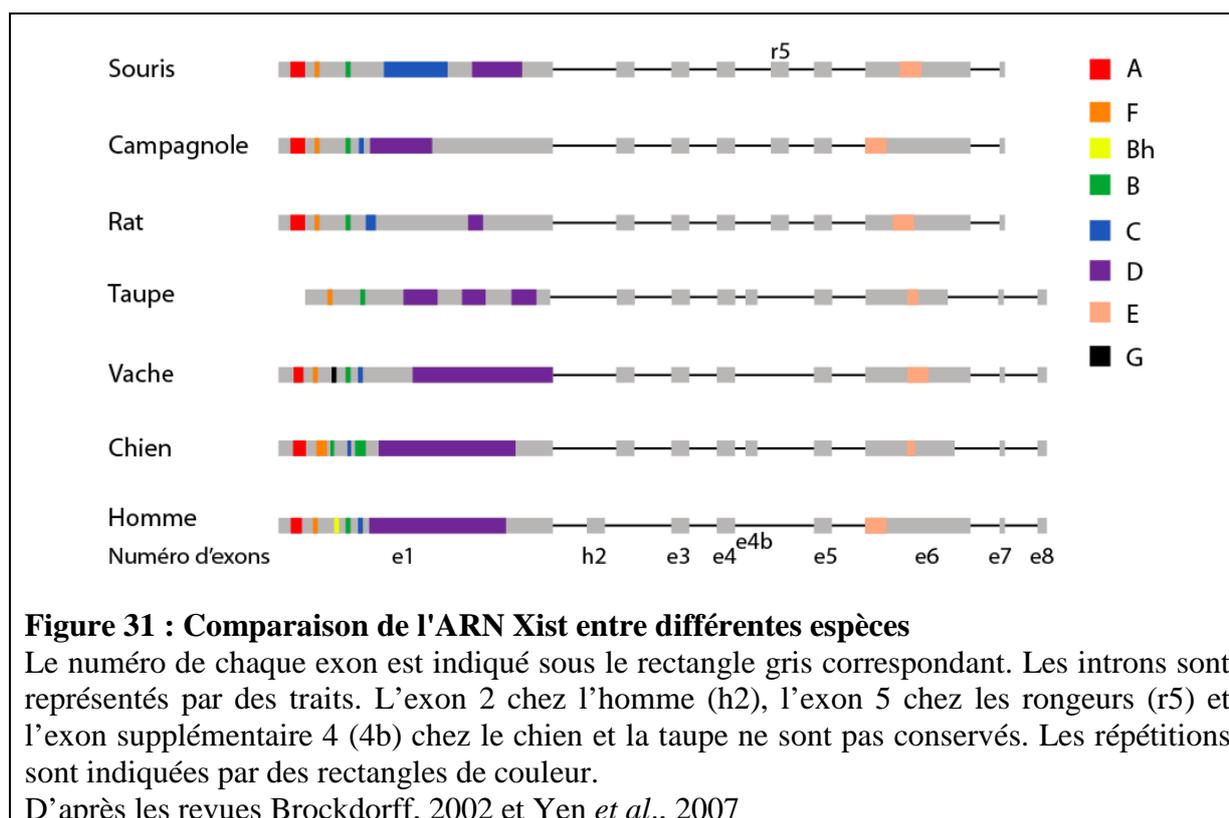
Une fois que l'état inactif est établi, il est maintenu tout au long des divisions cellulaires. Certaines des modifications épigénétiques sont d'ailleurs détectées sur le Xi dans des cellules métaphasiques (Boggs *et al.*, 2002 ; Chaumeil *et al.*, 2002 ; Mak *et al.*, 2002 ; Mermoud *et al.*, 2002 ; Peters *et al.*, 2002 ; Chadwick et Willard, 2004). De plus, une délétion de *Xist* dans des cellules somatiques (Brown et Willard, 1994 ; Csankovszki *et al.*, 1999) ou l'extinction transcriptionnelle de *Xist* après quelques jours de différenciation (Wutz et Jaenisch, 2000) n'induit pas une réactivation globale du Xi. Cette observation souligne l'importance de ces marques dans le maintien de l'état inactif au cours des divisions, même si elles ne suffisent pas puisque la localisation de macroH2A (Csankovszki *et al.*, 1999) et l'enrichissement en H3K27me3 (Zhang *et al.*, 2007) disparaissent du Xi et que certains gènes sont réactivés dans ces expériences (Csankovszki *et al.*, 2001 ; Zhang *et al.*, 2007).

La méthylation de l'ADN, quant à elle, est absolument nécessaire pour le maintien de l'état répressif, attendu que l'inactivation du gène *Dnmt1* codant l'enzyme responsable de cette modification provoque la réactivation du chromosome X dans l'embryon (Sado *et al.*, 2000).

#### **IV. Le gène *Xist* et son transcrit**

##### **IV.1 Description du gène *Xist***

Le gène *XIST* a été découvert chez l'homme dès 1991, puis chez la souris (*Xist*). Il a été le premier gène caractérisé comme n'étant exprimé qu'à partir du chromosome X inactif dans les cellules somatiques (Borsani *et al.*, 1991 ; Brown *et al.*, 1991). Son implication dans l'inactivation du chromosome X a pour cette raison vite été suggérée (Penny *et al.*, 1996). Le gène *Xist* est long de 35 kb et huit exons ont été caractérisés chez l'homme (Brockdorff *et al.*, 1992 ; Brown *et al.*, 1992 ; Chureau *et al.*, 2002), chez la vache (Chureau *et al.*, 2002), la souris (Simmler *et al.*, 1996 ; Sheardown *et al.*, 1997 ; Chureau *et al.*, 2002) et le campagnol (Nesterova *et al.*, 2001). En 2007, des études ont montré que la majorité des exons est conservée entre les espèces, malgré quelques exceptions (Yen *et al.*, 2007) (**Figure 31**).



De manière générale, la conservation de séquence des exons est importante : entre la souris et l'homme, elle est de 66 %, entre la souris et la vache de 62 %. En revanche, la séquence des introns l'est très peu (Chureau *et al.*, 2002).

L'expression de ce gène est sous le contrôle de trois promoteurs. Même si la majorité des transcrits sont issus du second promoteur (environ 75 % dans les cellules somatiques) (Johnston *et al.*, 1998), il a été montré que le troisième promoteur permet de produire un ARN de 1,6 kb (Zhao *et al.*, 2008).

Sa transcription se fait par l'ARN polymérase II, et il est mûri comme le serait un ARN messager, à savoir qu'il est épissé, qu'une coiffe lui est apposée à son extrémité 5' et qu'une queue poly(A) est synthétisée à son extrémité 3' (Brown *et al.*, 1991 ; Brockdorff *et al.*, 1992 ; Brown *et al.*, 1992 ; Hong *et al.*, 1999 ; Hong *et al.*, 2000). Par contre, malgré toutes ces caractéristiques, il n'est pas exporté dans le cytoplasme mais reste localisé sur le chromosome dont il est issu. Plusieurs variants de l'ARN Xist ont été observés. Certains correspondent à des formes non épissées. D'autres présentent une terminaison précoce de la transcription, ne contenant pas le ou les deux derniers exons (Brockdorff *et al.*, 1992 ; Hong *et al.*, 1999 ; Hong *et al.*, 2000). D'autres encore constituent des variants d'épissage alternatif dont l'extrémité 3' est modifiée (Chureau *et al.*, 2002). Nous reviendrons sur l'épissage de cet ARN par la suite.

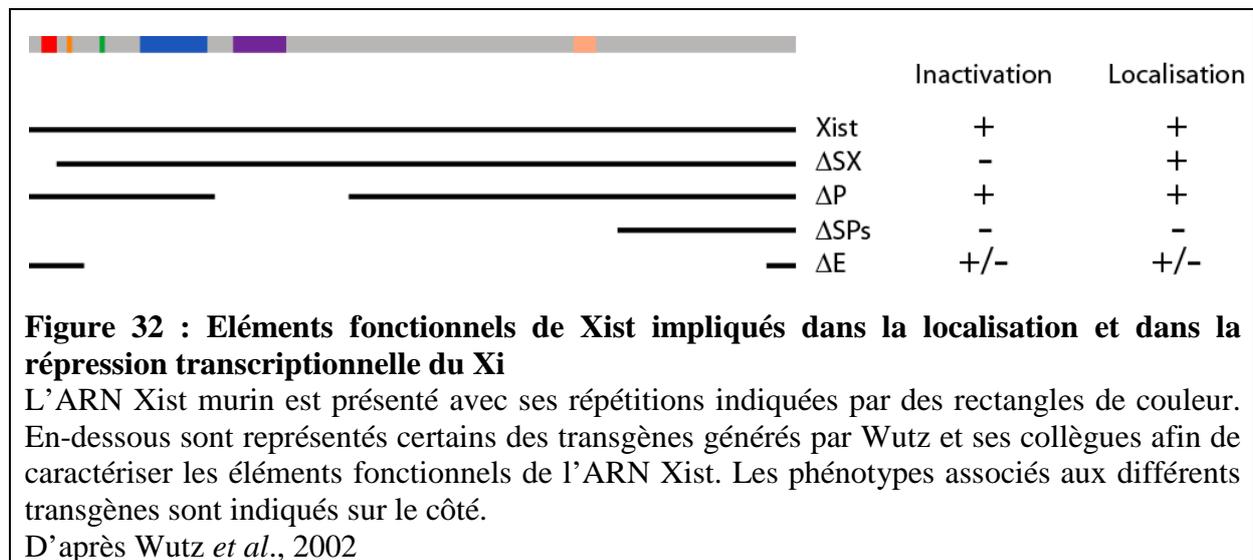
L'ARN *Xist* mature est composé de plusieurs régions contenant des séquences répétées en tandem. Leur conservation entre les espèces varie, certaines étant extrêmement divergentes, d'autres étant très conservées (Clemson *et al.*, 1996 ; Nesterova *et al.*, 2001) (cf. **Figure 31**). Nous verrons dans la partie suivante que certaines de ces régions sont connues pour être importantes pour les différentes fonctions de cet ARN. Nous discuterons notamment de la région des A-repeats, située à l'extrémité 5' de l'exon 1, qui est cruciale pour l'initiation de l'inactivation, sa délétion entraînant un défaut dans la mise en place du processus (Wutz *et al.*, 2002 ; Hoki *et al.*, 2009).

#### IV.2 Les fonctions de l'ARN *Xist*

La caractérisation fonctionnelle du gène a été réalisée par une série de délétions. Les délétions de la région promotrice et du premier exon de *Xist* dans des cellules ES femelles (Penny *et al.*, 1996), ainsi que les délétions de la majorité de l'exon 1 et des exons 2 à 5 (Marahrens *et al.*, 1997) provoquent un biais dans l'inactivation, c'est-à-dire que le chromosome inactivé est toujours celui ne portant pas la délétion. Il a donc vite été émise l'idée que le gène *Xist* joue un rôle clé dans l'inactivation. Cela a été confirmé par l'introduction d'un ADNc de *Xist* dans des autosomes sous le contrôle d'un promoteur inductible (Wutz et Jaenisch, 2000). L'induction de l'expression de *Xist* est suffisante pour inactiver le chromosome le portant, avant que la différenciation ne soit initiée, et cela même dans un contexte mâle. Cette étude a apporté la preuve formelle de l'implication primordiale de l'ARN *Xist* dans l'inactivation aléatoire. La même importance a été montrée dans l'inactivation soumise à empreinte parentale par le fait que des mutations dans ce gène transmises par le père provoquent une défaillance dans le développement des tissus extra-embryonnaires (Marahrens *et al.*, 1997 ; Hoki *et al.*, 2009).

Tandis que ces premières études de délétion ont permis de montrer l'implication de l'ARN *Xist* dans l'inactivation, Wutz et ses collègues ont généré de nombreux ADNc *Xist* porteurs de diverses délétions afin d'en déterminer les éléments fonctionnels. Placé sous le contrôle d'un promoteur inductible, chacun d'entre eux a été introduit par recombinaison au locus *Hprt* (hypoxanthine phosphoribosyltransferase) du chromosome X dans des cellules ES mâles (Wutz *et al.*, 2002). Après induction de l'expression de ces différents variants, ils ont étudié deux points : la localisation de l'ARN *Xist* par RNA-FISH, et l'efficacité d'inactivation par estimation de la mort cellulaire (rappelons que dans un contexte mâle, l'inactivation du seul

chromosome X entraîne la mort cellulaire). Certaines délétions et leurs effets sur l'inactivation et sur la localisation de Xist sont résumés dans la **Figure 32**.



La localisation est perdue si les deux tiers de l'ARN contenant les 6 régions conservées ne sont pas présents (transgène ΔSPs). Les auteurs ont proposés que les éléments composants cette région pourraient avoir des fonctions redondantes pour ce qui est de la localisation, ce qui expliquerait notamment la capacité de Xist à recouvrir le chromosome X même si des délétions apparaissent (transgène ΔP). C'est par exemple le cas de la délétion de l'exon 4, pourtant très conservé entre les espèces, qui ne provoque pas de délocalisation de Xist hors du Xi (Caparros *et al.*, 2002).

En ce qui concerne l'inactivation, elle n'a plus lieu avec le transgène ΔSPs, mais de manière plus intéressante, c'est également le cas lorsque seule la région des A-repeats est délétée (transgène ΔSX). Cependant, cette région n'est pas impliquée dans la localisation de Xist, ni dans la formation du domaine Xist, ni même la répression transcriptionnelle des séquences répétées du chromosome X (Wutz *et al.*, 2002 ; Chaumeil *et al.*, 2006 ; Clemson *et al.*, 2006 ; Hoki *et al.*, 2009). La région des A-repeats est donc nécessaire à la mise en place de l'inactivation, mais elle intervient dans une étape bien définie. Nous consacrerons une partie à cette région dans la suite de l'Introduction.

### IV.3 Les régulations de l'expression du gène *Xist*

Bien que la surexpression de *Xist* ait été identifiée comme l'événement déclencheur de l'initiation de l'inactivation, les bases moléculaires l'expliquant sont restées incomprises

pendant de longues années. Au début des années 2000, il a été proposé que cette augmentation de la quantité d'ARN *Xist* puisse être due à une régulation différente de sa demi-vie par dégradation/stabilisation. Ce n'est que récemment que les preuves d'une augmentation de la transcription de *Xist* par différentes régulations ont été découvertes.

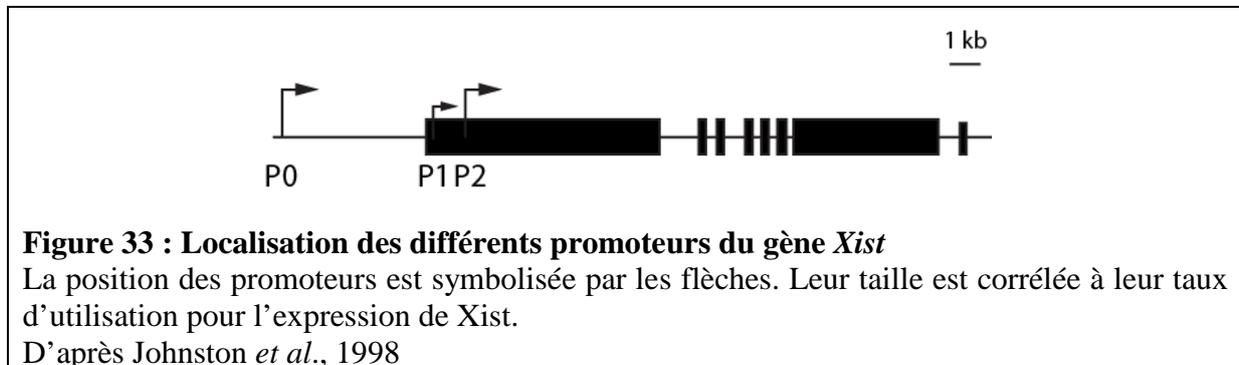
#### IV.3.1 La régulation de la localisation et de la demi-vie de l'ARN *Xist*

Comme nous l'avons déjà mentionné, l'ARN *Xist* est transcrit par l'ARN polymérase II, comme les ARN messagers codant des protéines et en possède les principales caractéristiques. Malgré ces marques distinctives qui promeuvent l'export hors du noyau par l'intermédiaire de facteurs qui recrutent les ARNm au niveau des pores nucléaires, *Xist*, lui, reste dans le noyau. Des études ont montré qu'il n'est pas même exporté dans le cytoplasme le temps d'être maturé avant d'être ré-importé dans le noyau ; il reste exclusivement nucléaire (Cohen et Panning, 2007). Une explication pourrait venir du fait que le facteur TAP/NXF1 ne se lie que faiblement à l'ARN *Xist* (Cohen et Panning, 2007). L'export de la majorité des ARNm est dépendant de ce facteur qui est recruté sur les ARN au niveau des jonctions d'exons générés par l'épissage. La faible fixation de facteurs permettant l'export du noyau semble donc contribuer à la rétention nucléaire de *Xist*, impliquant que tous les ARN *Xist* produits sont regroupés et peuvent recouvrir le chromosome X. Cependant les raisons de cette fixation limitée restent toujours inconnues à ce jour.

Avant la différenciation, *Xist* est exprimé en faible quantité à partir des deux allèles. Au moment de la différenciation, l'expression de *Xist* est augmentée sur le futur Xi tandis qu'elle est réprimée sur le futur Xa. Cette expression différentielle a longtemps été incomprise, jusqu'à ce qu'il soit suggéré que la transition d'une faible vers une forte expression soit due à une stabilisation de l'ARN *Xist* sur le Xi (Panning *et al.*, 1997). Ceci pourrait être la conséquence de facteurs augmentant la stabilité de l'ARN sur le Xi mais aussi à des facteurs bloquant l'expression de *Xist* sur le Xa. Cette idée était partagée par une autre équipe qui a également montré que la stabilisation de *Xist* sur le futur Xi était suffisante pour expliquer l'augmentation observée de l'ARN *Xist* (Sheardown *et al.*, 1997). De plus, la stabilisation sur le Xi se produit avant le silence sur le Xa.

L'accumulation sur le Xi par stabilisation de l'ARN serait due à une utilisation différentielle des promoteurs de *Xist* (Johnston *et al.*, 1998). Jusque là, seul un promoteur était connu, l'actuel P1. Ils ont découvert un second promoteur, situé 6,5 kb en aval du premier, qui

constitue en réalité le promoteur majeur de l'expression de *Xist* dans les cellules somatiques. Ils ont également mis en évidence un troisième promoteur en amont de P1, spécifique des cellules ES, désigné sous le nom de P0 (**Figure 33**). Ces trois promoteurs entraînent donc la transcription de trois isoformes différentes de l'ARN *Xist*.



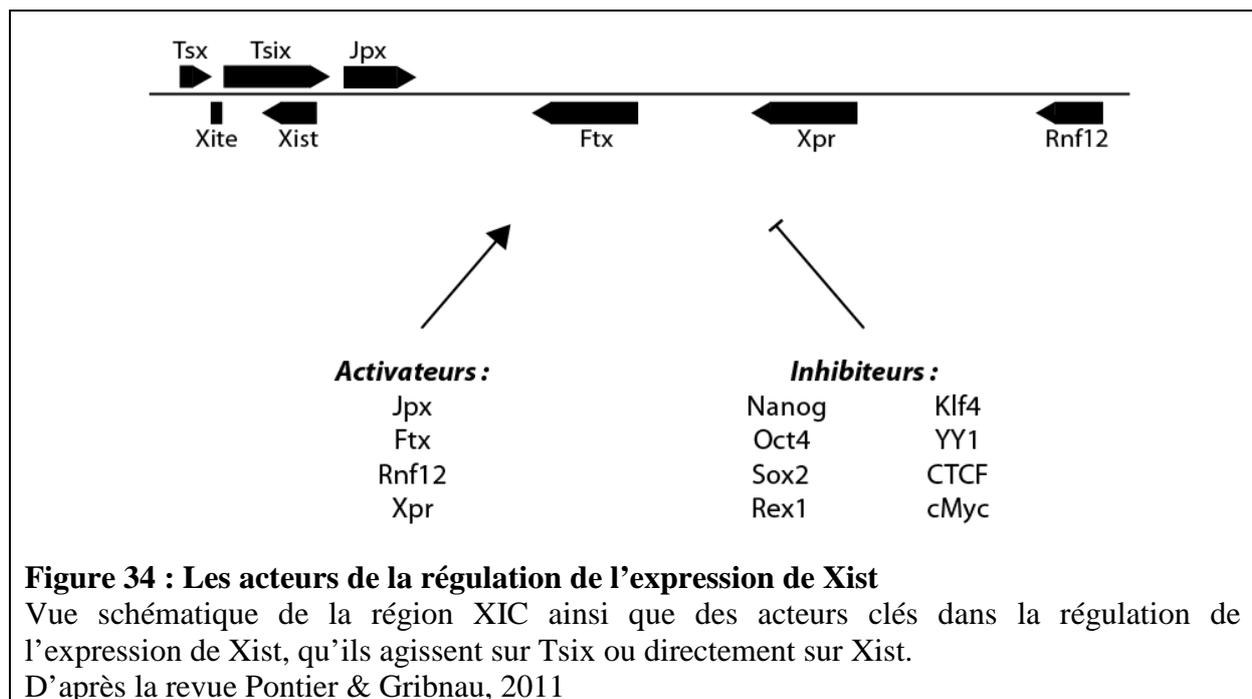
Ils ont ainsi déterminé que les transcrits instables de *Xist* étaient issus de la transcription à partir du promoteur P0 et que la transition vers une utilisation des promoteurs P1/P2 permettait de transcrire des ARN stables, que ce soit lors de l'inactivation par empreinte ou lors de l'inactivation aléatoire (Johnston *et al.*, 1998). Cependant, l'existence du promoteur P0 a été remise en cause peu de temps après (Warshawsky *et al.*, 1999).

Par ailleurs, dans le but d'identifier de nouvelles molécules impliquées dans l'inactivation du chromosome X, une analyse en série de l'expression des gènes (SAGE) a été réalisée sur des embryons murins mâles et femelles à 6,5 jours post-coïtum (Ciaudo *et al.*, 2006). A ce stade de développement, l'inactivation aléatoire vient de démarrer dans l'épiblaste des embryons femelles et le dimorphisme sexuel entre mâle et femelle est encore limité. Si des différences de fréquence de transcrits étaient révélées, un certain nombre devrait concerner la mise en place de l'inactivation. Cette analyse a ainsi permis d'identifier, parmi les plus de 200 gènes les plus surexprimés chez la femelle par rapport au mâle, le gène *Eif1*. Ce gène est impliqué dans l'initiation de la traduction et la dégradation des ARN. Une diminution de l'expression de ce gène par ARN interférence conduit à l'absence de formation de domaine transcriptionnellement silencieux ainsi qu'à l'absence d'inactivation du chromosome X. De manière similaire, les mêmes auteurs ont étudié les effets d'une diminution des gènes *Rent1* et *Exosc10*. Ces deux gènes codent pour des facteurs impliqués dans la dégradation des ARN. *Rent1* est un facteur essentiel à la machinerie de dégradation non-sens, tandis que *Exosc10* est un composant exclusivement nucléaire de l'exosome. Dans

les deux cas, l'ARN *Xist* n'est pas surexprimé et l'inactivation ne se met pas en place. Afin de compléter leur étude, ils ont montré que lorsque ces gènes sont réprimés, le taux de transcrits épissés de *Xist* chute fortement sans que celui de transcrits non épissés ne soit affecté ou que la stabilité de l'ARN ne soit diminuée (Ciaudo *et al.*, 2006). Il ressort de cette étude que la machinerie de dégradation nucléaire est impliquée dans la régulation du taux de transcrits d'ARN *Xist* épissés mais également dans l'initiation de l'inactivation.

#### IV.3.2 Les régulations transcriptionnelles de l'expression de *Xist*

L'augmentation de la transcription du gène *Xist* est un élément crucial dans la mise en place de l'inactivation. L'expression de ce gène doit donc être soumise à de nombreuses régulations afin d'assurer un processus correct (fenêtre de temps, nombre de chromosomes X à inactiver, réactivation du Xi, ...). La **Figure 34** présente les principaux acteurs de cette régulation.

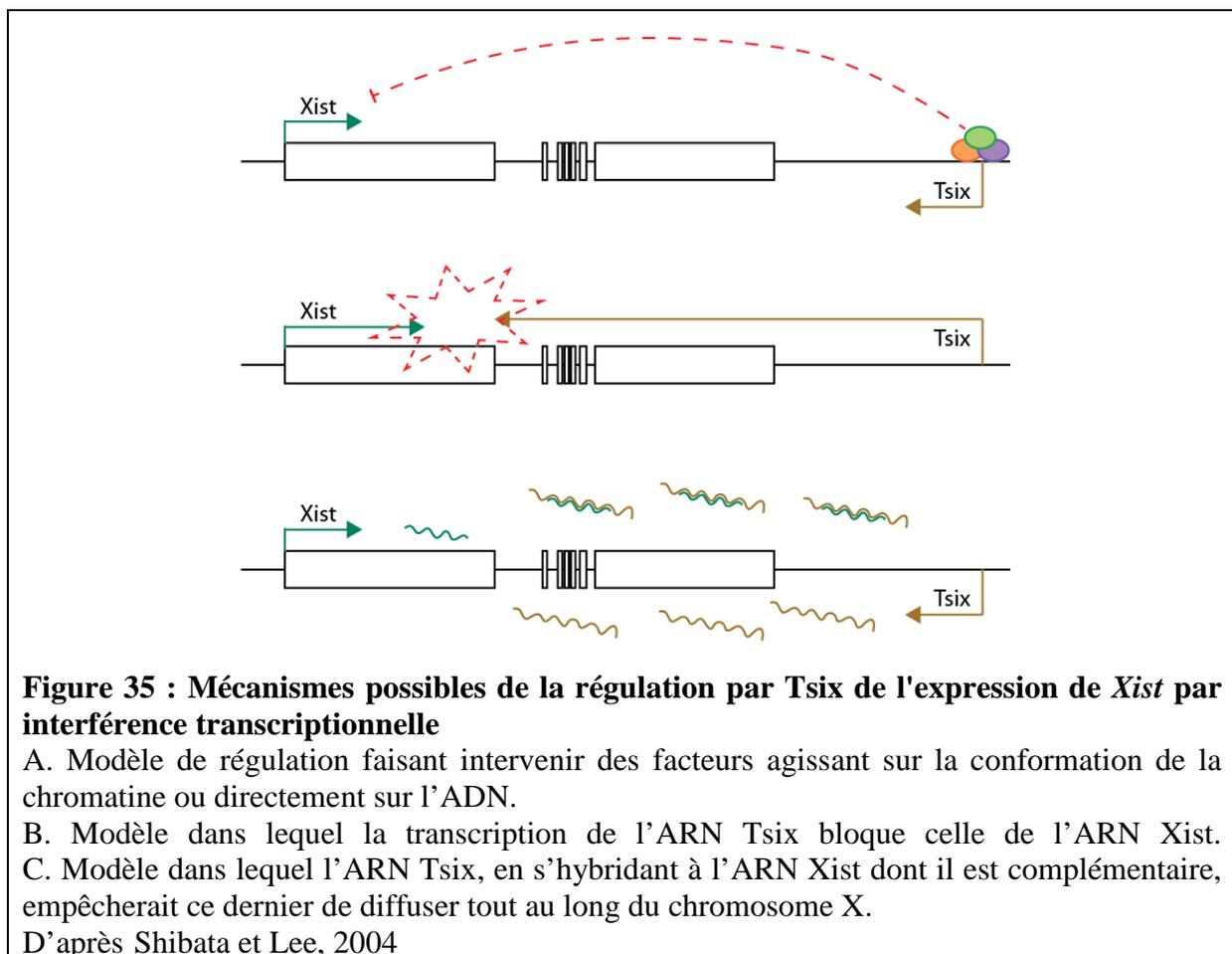


##### a. Les inhibiteurs transcriptionnels de *Xist*

###### - Régulation par l'ARN *Tsix*

Avant l'initiation de l'inactivation, *Tsix* est exprimé à partir des 2 chromosomes X. Au moment de l'initiation, son expression devient monoallélique, à partir du chromosome qui restera actif. L'ARN *Tsix* pourrait inhiber l'expression de *Xist* en *cis* par plusieurs mécanismes : (i) par interférence au niveau de la transcription. Ce serait la transcription de l'ARN *Tsix* chevauchant celle de *Xist* en antisens qui bloquerait cette dernière (Shibata et

Lee, 2004 ; Sado *et al.*, 2006) (**Figure 35.B et C**). (ii) Il a été montré que Tsix est capable d'influencer l'apposition de marques épigénétiques au niveau du promoteur de *Xist*. Il permettrait ainsi de moduler l'état hétérochromatinien de ce promoteur sur le Xa dans le but de bloquer son expression (Navarro *et al.*, 2005 ; Sado *et al.*, 2005 ; Sun *et al.*, 2006) (Figure 35.A). (iii) Tsix pourrait de même agir de concert avec des facteurs de pluripotence afin d'inhiber l'expression de *Xist*, notamment dans l'étape de comptage (Shibata et Lee, 2004, Navarro *et al.*, 2008 ; Nesterova *et al.*, 2011) (**Figure 35.A**). Nous allons justement détailler ces facteurs dans le prochain paragraphe.

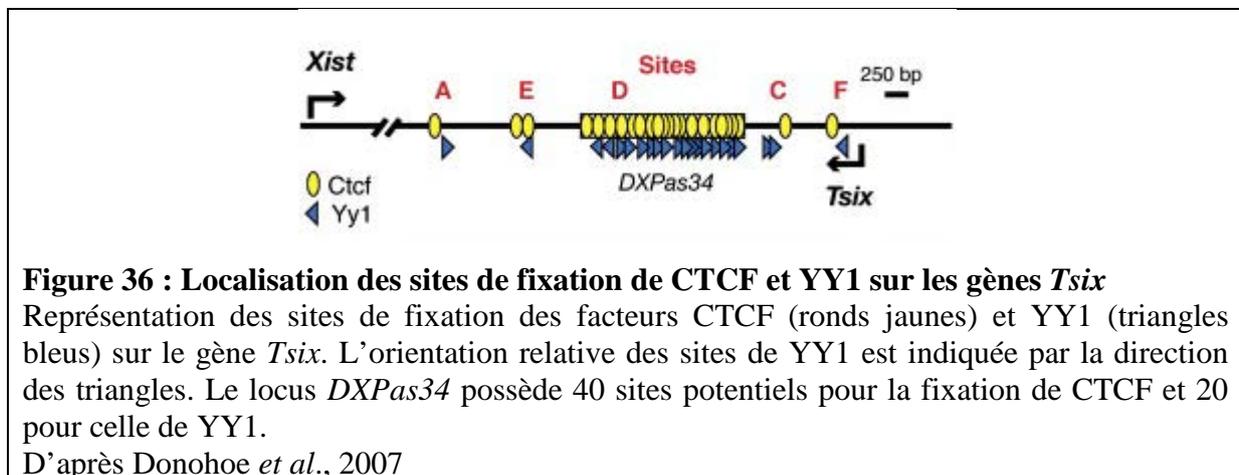


#### - Régulation par les facteurs de pluripotence

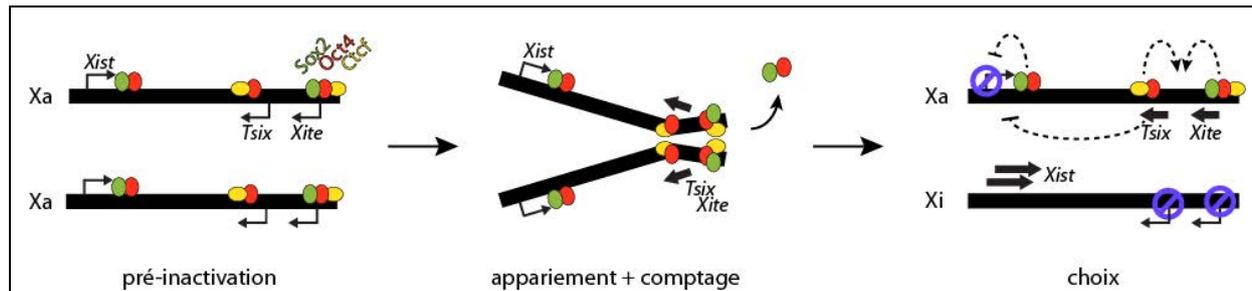
La régulation par l'ARN Tsix ne peut pas expliquer à elle seule la régulation de l'expression de *Xist*. En effet, l'inactivation de *Tsix* sur le chromosome X paternel ne bloque pas la réactivation du Xp dans les cellules de l'ICM (Kalantry et Magnuson, 2006). De même, la transcription de *Xist* reste réprimée dans des cellules ES indifférenciées malgré une mutation du gène *Tsix* (Navarro *et al.*, 2005). C'est alors qu'une corrélation entre les facteurs

de pluripotence et l'inactivation du chromosome X a été observée : dans l'ICM du blastocyste, la réversion de l'inactivation du X suit l'expression de Nanog (Mak *et al.*, 2004). Il a ainsi été montré que des facteurs clés comme Nanog, Oct4 ou encore Sox2 se liaient au niveau de l'intron 1 de *Xist* dans des cellules ES indifférenciées et qu'ils répriment l'expression de *Xist* (Navarro *et al.*, 2008). Ils ont également montré l'importance de la coopérativité entre ces facteurs.

Ce constat s'est poursuivi par une étude plus précise de la régulation de l'expression de *Xist* par ces facteurs. Outre l'action inhibitrice directe par fixation sur l'intron 1 de l'ARN *Xist*, Oct4 se fixe également de manière directe sur *Tsix* et *Xite* et se complexe avec le facteur CTCF (Donohoe *et al.*, 2009). Nous avons parlé brièvement de CTCF (cf. paragraphe I.2.1), un facteur de transcription qui peut être un activateur, un inhibiteur ou encore un acteur dans la régulation de l'expression de gènes soumis à empreinte. Ce facteur possède en outre des sites de fixation localisés dans la région activatrice de *Tsix* (Chao *et al.*, 2002). En revanche, ce facteur n'agit pas seul, et il a été découvert que les sites de fixation de CTCF sur cette région sont souvent accompagnés de sites pour un autre facteur de transcription, le facteur YY1 (Yin-Yang1) (Donohoe *et al.*, 2007) (**Figure 36**).



Le facteur Oct4 agit donc en coopération avec CTCF et YY1 afin d'activer l'expression de *Tsix* qui, lui, réprime celle de *Xist*. Cette action conjointe de Oct4, CTCF, YY1 sur *Xite* et *Tsix* pourrait se réaliser également sur les étapes de comptage et de choix (Donohoe *et al.*, 2009) (**Figure 37**). Par la suite, le facteur YY1 a été montré interagissant avec la région C de l'ARN *Xist* (Jeon et Lee, 2011), ce qui relierait encore plus la régulation de *Tsix* à celle de *Xist*.



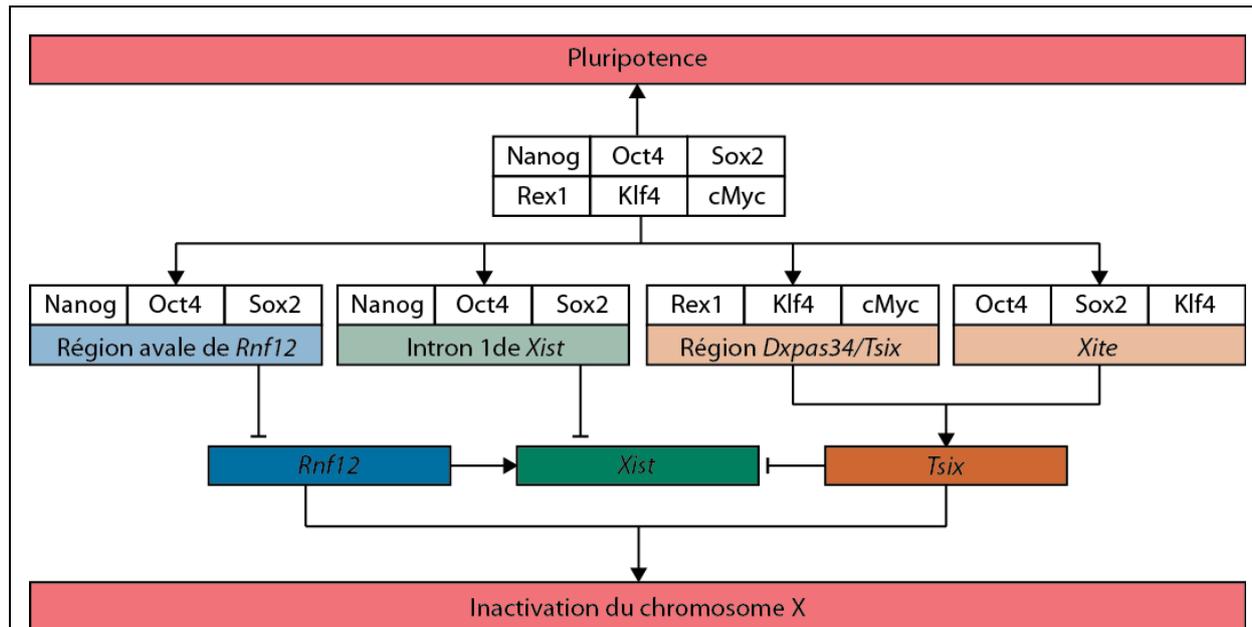
**Figure 37 : Modèle de la régulation dynamique et multifactorielle de l'inactivation du chromosome X par Oct4**

Selon ce modèle, Oct4 se fixe à l'extrémité 5' de *Xite* et *Tsix* afin d'activer la transcription de *Tsix*. En même temps, Oct4 se fixe aux sites qui chevauchent le gène *Xist*, réprimant l'expression de *Xist* directement, et via l'activation de *Tsix*. Lorsque la cellule se différencie, en coopération avec CTCF, Oct4 facilite l'appariement des deux chromosomes X en agissant sur *Tsix/Xite*. Sur le futur Xi, Oct4 ne se fixe plus puisque son taux dans la cellule diminue, ce qui réduit l'expression de *Tsix* et induit celle de *Xist*. Sur le futur Xa, le peu de fixation de Oct4 maintient l'expression de *Tsix* et l'inhibition de *Xist*.

D'après Donohoe *et al.*, 2009

Cependant, ces trois facteurs de pluripotence ne sont pas suffisants pour expliquer la régulation de *Tsix*. Les expériences sur ces facteurs ayant été réalisées en majorité dans des cellules souches pluripotentes induites (iPS cells), trois autres facteurs utilisés dans ce type de culture ont été étudiés, Rex1, Klf4 et cMyc. Il avait d'ailleurs été observé une diminution de l'expression de *Tsix* lors d'une délétion du facteur de pluripotence Rex1 (Masui *et al.*, 2008). Rex1 est un marqueur pour les cellules ES indifférenciées et il a été montré que Nanog est un activateur de sa transcription, et que cette activation est accentuée par Sox2 et Oct4 (Shi *et al.*, 2006). Ces trois facteurs ont été trouvés se fixant à la région 5' de *Tsix* dans les cellules mâles et femelles ES (Navarro *et al.*, 2010). Ils concluent de leurs expériences que *DXPas34* orchestre le recrutement de Rex1, Klf4 et cMyc afin d'activer *Tsix*, de manière à coupler étroitement l'expression de *Tsix* à la pluripotence. Ils ajoutent à cette conclusion le rôle de Rex1 dans l'élongation de la transcription de *Tsix* et celui de Klf4 dans le recrutement de la machinerie de transcription au niveau de son promoteur.

Les facteurs de pluripotence sont donc au cœur de la régulation de l'initiation de l'inactivation, tant par une régulation répressive directe de l'expression de *Xist* (lors de la réactivation du Xp) que par une régulation activatrice de l'expression de *Tsix* (lors de l'initiation de l'inactivation aléatoire) (**Figure 38**).



**Figure 38 : Réseau transcriptionnel couplant les régulateurs de pluripotence à l'inactivation du chromosome X**

Nanog, Oct4 et Sox2 se lient à l'intron 1 du gène *Xist*. La protéine associée à la pluripotence Rex1 se lie à la région 5' de *Dxpas34/Tsix*, conjointement aux facteurs de reprogrammation Klf4 et cMyc. Oct4, Sox 2 et Klf4 se lient également à *Xite*.

D'après Navarro *et al.*, 2011

#### b. Les activateurs transcriptionnels de *Xist*

Tandis que *Tsix* réprime l'expression de *Xist* sur le futur Xa, il existe des mécanismes d'activation de *Xist* sur le futur Xi.

##### - Régulation par la protéine RNF12

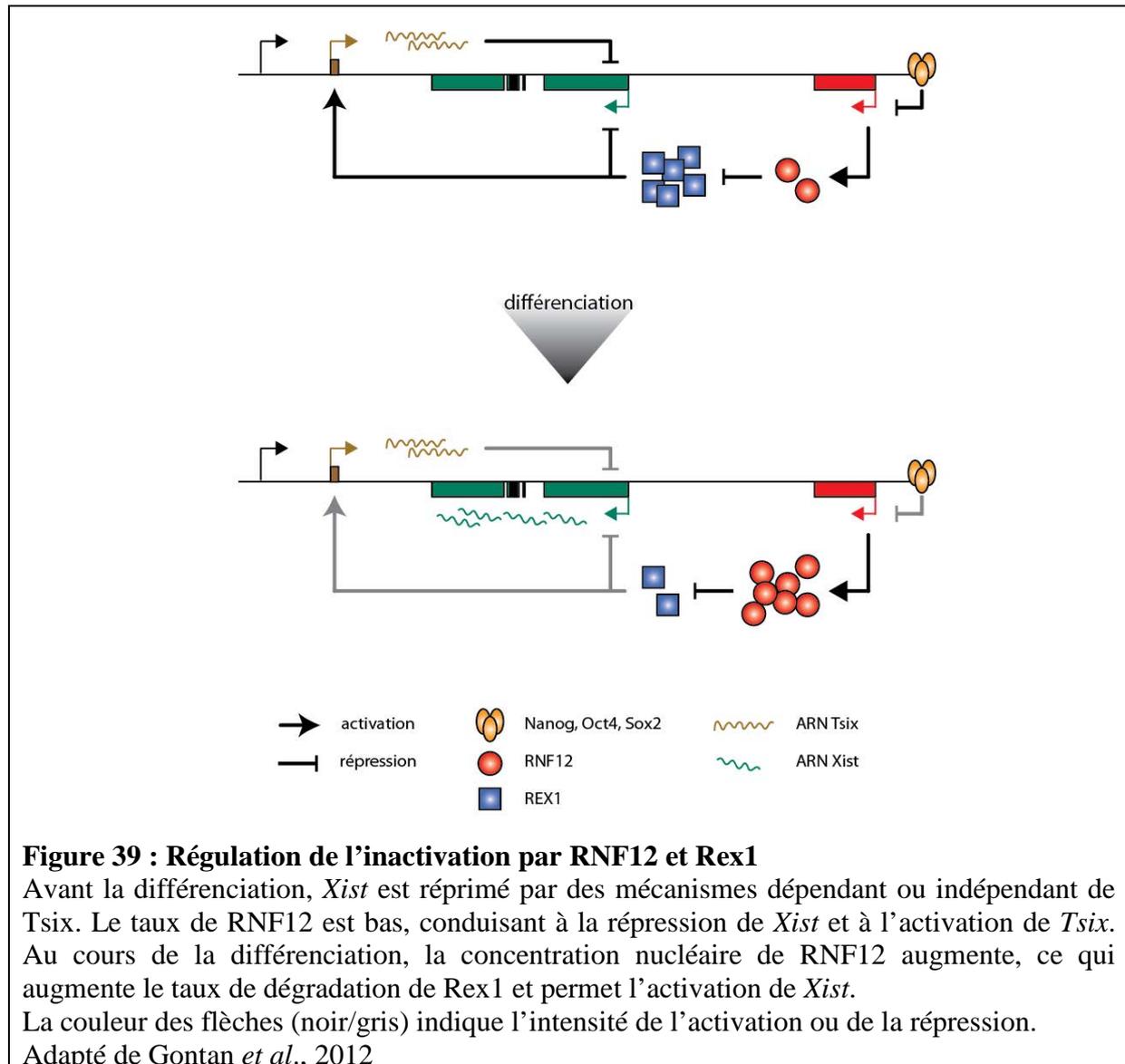
*Rnf12* est un gène localisé sur le chromosome X, plusieurs centaines de kilobases en aval de *Xist/Tsix* et codant pour une ubiquitine ligase E3 (RNF12 ou RLIM, RING finger LIM-domain-interacting protein). Il a tout d'abord été identifié comme un activateur de la transcription de *Xist* durant l'inactivation aléatoire (Jonkers *et al.*, 2009). Par la suite, son implication dans l'inactivation soumise à empreinte a été démontrée (Shin *et al.*, 2010).

Des cellules ES mâles surexprimant *Rnf12* induisent l'inactivation du seul chromosome X lors de la différenciation. De même, une large fraction de cellules ES femelles surexprimant ce gène inactive leurs deux chromosomes X. L'augmentation de la transcription de *Xist* dans les cellules femelles semble donc déclenchée par la présence de deux copies du gène *Rnf12*, au contraire de la seule copie présente chez les mâles (Jonkers *et al.*, 2009), ce qui représente une avancée majeure dans la compréhension du mécanisme de comptage du nombre de

chromosomes X dans la cellule. De manière cohérente, RNF12 a été montré comme agissant en *trans* sur le promoteur de *Xist*, sans avoir d'effet sur l'activité de Nanog, Oct4, Sox2 ou encore Tsix (Barakat *et al.*, 2011). Cependant, RNF12 est tout de même lié aux facteurs de pluripotence. En effet, dans des cellules ES femelles surexprimant le gène *Rnf12*, une proportion non négligeable de cellules présente déjà une accumulation de l'ARN *Xist* avant même la différenciation (Jonkers *et al.*, 2009). Cela signifie que pour assurer la réactivation du Xi dans les cellules pluripotentes, l'expression de *Rnf12* doit être diminuée. Il a ainsi été montré que les facteurs de pluripotence Nanog, Oct4 et Sox2 contrôlent également la répression de l'expression de *Rnf12*, en plus de leur action sur *Xist* et sur *Tsix* (Navarro *et al.*, 2011).

Très récemment, un autre lien entre facteurs de pluripotence et RNF12 a été montré. En effet, Gontan et ses collègues ont identifié le facteur Rex1 comme une cible de l'activité ubiquitine ligase de RNF12 (Gontan *et al.*, 2012). RNF12 catalyse l'ubiquitinylation de Rex1 et ainsi sa dégradation par le protéasome (cf. **Figure 38**).

Compte tenu de toutes les connaissances actuelles, l'hypothèse que l'on peut avancer est que lorsque les facteurs de pluripotence Nanog, Oct4 et Sox2 sont fortement exprimés dans la cellule, le gène *Rnf12* est réprimé, donc Rex1 n'est pas dégradé et il peut activer l'expression de *Tsix* qui lui réprime celle de *Xist*, ce qui permet la réactivation du Xi. A l'inverse, lorsque le taux de RNF2 atteint un certain seuil (via une plus faible concentration des facteurs Nanog, Oct4 et Sox2), la dégradation de Rex1 ne permet plus à ce dernier d'activer l'expression de *Tsix* de manière optimale, ce qui lève la répression de l'expression de *Xist* et déclenche l'initiation de l'inactivation (**Figure 39**).



#### - Régulation par l'ARN non-codant *Jpx*

Le gène non-codant spécifique des euthériens *Jpx/Enox* est localisé environ 10 kb en aval de *Xist* d'où il est transcrit en sens opposé (Chureau *et al.*, 2002 ; Johnston *et al.*, 2002 ; Chow *et al.*, 2003). *Jpx* ne contient pas de cadre ouvert de lecture mais il est relativement conservé dans ses 50 exons. Les premiers rapports sur ce gène indiquent qu'il n'est pas régulé au cours du développement et surtout qu'il n'est pas spécifique des mâles ou des femelles, ce qui ne le plaçait pas parmi les régulateurs potentiels de l'inactivation (Chureau *et al.*, 2002 ; Johnston *et al.*, 2002 ; Chow *et al.*, 2003). Pourtant, des expériences de capture de la conformation du chromosome (3C) ont identifié l'extrémité 5' de *Jpx* comme étant en interaction avec le promoteur de *Xist* (Tsai *et al.*, 2008). Des études plus poussées ont donc été entreprises afin de comprendre le rôle de cet ARN dans le processus d'inactivation.

Il est apparu que le taux de transcrit Jpx augmente 10 à 20 fois au moment de l'initiation de l'inactivation. La délétion de *Jpx* bloque l'inactivation et est létal dans les cellules femelles, et il en est de même si l'ARN est supprimé après sa transcription. De plus, le fait de fournir l'ARN Jpx en *trans* rétablit la viabilité (Tian *et al.*, 2010). Il en a donc été conclu que Jpx est acteur de l'inactivation qui agit en *trans* et qui fonctionne en tant qu'ARN non-codant. D'autre part, les effets de la délétion de *Jpx* disparaissent lorsque le gène *Tsix* est tronqué, ce qui indique une relation antagoniste entre ces deux ARN. Les auteurs penchent pour des voies parallèles de régulation, les ARN *Tsix* et *Jpx* contrôlant indépendamment la transcription de *Xist* (Tian *et al.*, 2010).

- Régulation par l'ARN non-codant *Ftx*

Récemment, un gène localisé dans la région 5' de *Xist*, dont la fonction était jusqu'alors inconnue, a été caractérisé (Chureau *et al.*, 2011). Il s'agit du gène *Ftx* (five prime to *Xist*) (cf. **Figure 25**). Ce gène produit un long ARN non-codant nucléaire, conservé entre la souris, l'homme et la vache. *Ftx* fait partie des gènes qui échappent partiellement à l'inactivation et il a été observé que son expression augmente au début de la mise en place de l'inactivation dans les cellules ES femelles, suivant un profil remarquablement proche de celui de *Xist*. La délétion de ce gène dans des cellules ES provoque une altération du taux de transcrits issus du XIC, en particulier une baisse très importante du taux d'ARN *Xist*, corrélée à une augmentation de la méthylation des îlots CpG du promoteur de *Xist*. *Ftx* a donc été caractérisé comme un nouveau régulateur positif de l'expression de *Xist* (Chureau *et al.*, 2011).

c. Régulation par la région des A-repeats de l'ARN *Xist* lui-même

La région des A-repeats située à l'extrémité 5' de l'exon 1 de l'ARN *Xist* régule elle aussi l'expression de *Xist*. La délétion de cette région sur le chromosome X paternel chez la souris conduit à une diminution de la quantité de l'ARN *Xist* (Hoki *et al.*, 2009). Cette délétion conduit également à l'activation de l'expression de la copie paternelle de *Tsix*. De plus, les îlots CpG de la région promotrice du gène *Xist* paternel présentent un taux significatif de méthylations sur le X paternel muté, alors qu'il devrait être essentiellement non méthylé. Il semblerait donc que la région des A-repeats soit un élément génomique essentiel pour la régulation transcriptionnelle correcte des gènes *Xist* et *Tsix*, et par conséquent à l'inactivation du chromosome X.

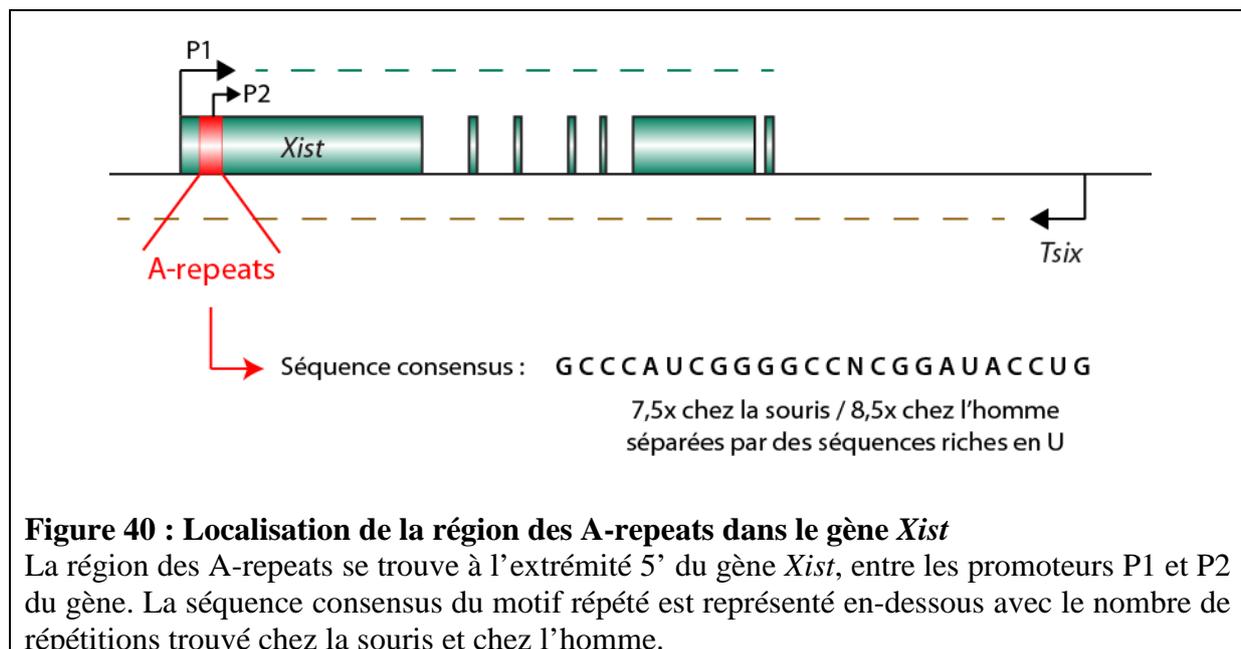
## V. La région des A-repeats

L'ARN Xist est long de 17 kb et comme nous l'avons déjà précisé, il comporte plusieurs régions de séquences répétées en tandem (Clemson *et al.*, 1996 ; Nesterova *et al.*, 2001). Nous avons également déjà parlé de la région des A-repeats, localisée à l'extrémité 5' du premier exon. Dans cette partie, nous allons voir que cette région est cruciale pour l'inactivation, et ce, à plusieurs niveaux (Wutz *et al.*, 2002 ; Hoki *et al.*, 2009).

### V.1 Présentation de la région des A-repeats

#### V.1.1 Caractérisation de la région des A-repeats

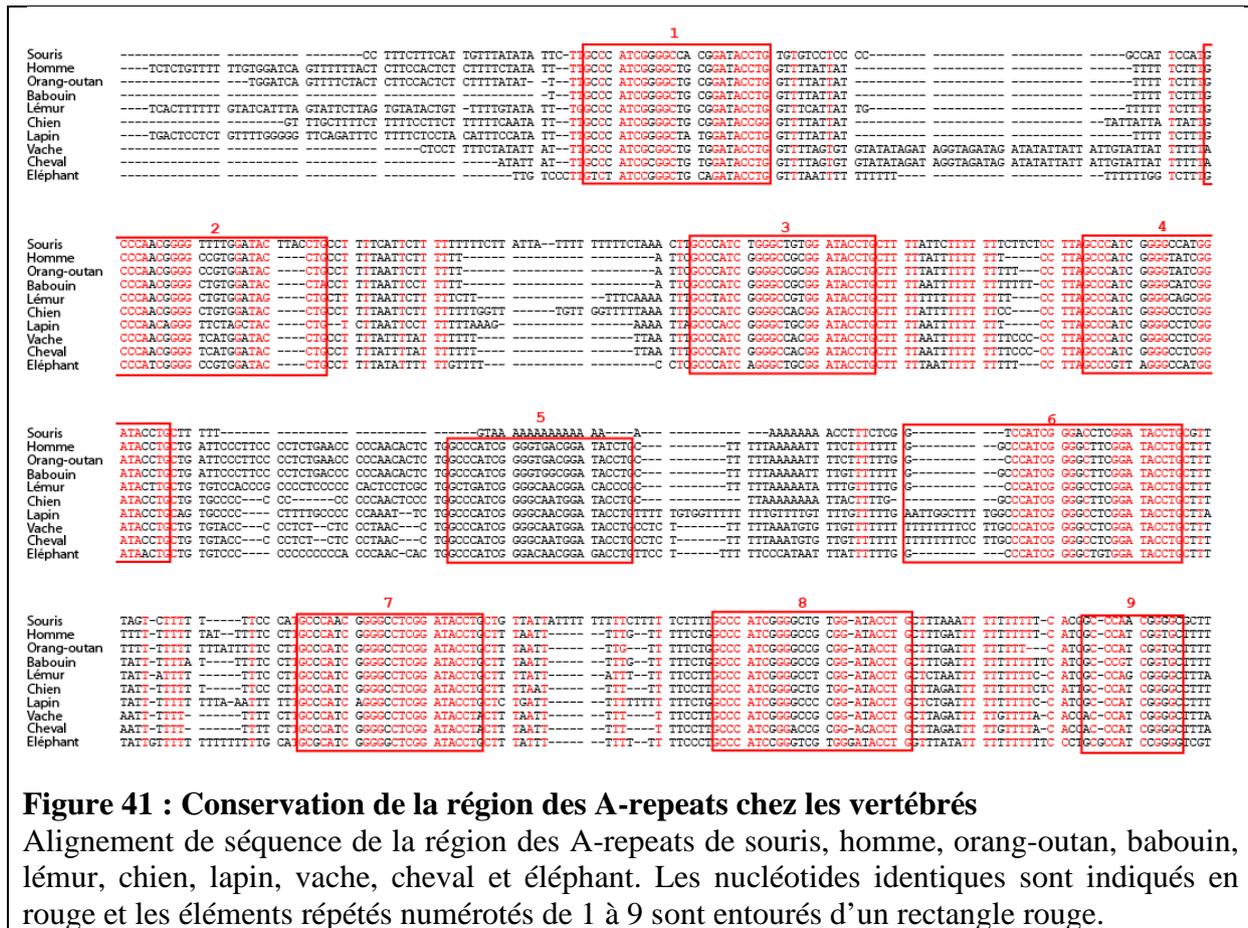
La région des A-repeats est localisée entre les deux promoteurs P1 et P2 de *Xist* (**Figure 40**). Chez la souris, elle est constituée de 419 nucléotides et comporte 7 répétitions du motif G(C/U)CCA(U/A)C(G/U)GGG(C/U)(C/U)N(C/U)GGAUAC(C/U)U(G/A) auxquels s'ajoute une demi-répétition (c'est-à-dire GCCCAACGGGGC). Ces répétitions sont séparées par des régions riches en résidus A et U. Chez l'homme, elle comporte 8 répétitions entières, séparées par des régions riches en C et U.



**Figure 40 : Localisation de la région des A-repeats dans le gène *Xist***

La région des A-repeats se trouve à l'extrémité 5' du gène *Xist*, entre les promoteurs P1 et P2 du gène. La séquence consensus du motif répété est représenté en-dessous avec le nombre de répétitions trouvé chez la souris et chez l'homme.

La région des A-repeats est la plus conservée des régions répétées de l'ARN Xist. La **Figure 41** présente la comparaison des séquences de cette région pour plusieurs espèces de vertébrés et met en évidence le degré de conservation entre espèces.



Son importance dans le processus d'inactivation a été découverte par des expériences de délétion de régions de l'ARN Xist afin d'en caractériser leurs fonctions respectives. Lorsque l'ARN Xist est transcrit à partir d'un ADNc dans lequel la région des A-repeats est déléte, il peut s'accumuler sur le chromosome X dans des cellules ES murines mâles, mais il est incapable d'en initier l'inactivation (Wutz *et al.*, 2002). De même, dans des cellules humaines HEK293, l'intégration de l'ADN correspondant à Xist sans la région des A-repeats en simple copie conduit à un blocage de la localisation de l'ARN Xist ainsi qu'à une absence d'inactivation. Il semble donc que cette région soit essentielle pour la localisation de Xist sur le Xi et pour la répression génique dans les cellules humaines (Chow *et al.*, 2007). Plus récemment, il a été montré que la délétion mono-allélique paternelle de la région des A-repeats chez l'embryon murin provoque un défaut d'inactivation dans les tissus extra-embryonnaires (qui, pour rappel, sont soumis à l'inactivation par empreinte). Cela met en évidence l'incapacité du chromosome X muté à être inactivé (Hoki *et al.*, 2009), et par là-même, l'importance de la région des A-repeats pour l'inactivation du chromosome X, qu'elle soit soumise à empreinte ou aléatoire.

### V.1.2 Régulation de l'expression des A-repeats

Nous avons abordé précédemment les mécanismes de régulation de l'expression de *Xist*. Il est évident que ces mécanismes de régulation régissent également l'expression de la région des A-repeats puisque cette région fait partie de l'ARN *Xist*. Toutefois, la région des A-repeats est localisée entre les deux promoteurs de *Xist*, P1 et P2. Ceci implique que les deux ARN *Xist* variants issus de l'usage alternatif de ces deux promoteurs ont des propriétés différentes. Seul le transcrit issu de P1 est théoriquement capable d'inactiver, puisqu'il comprend la région des A-repeats. Cette hypothèse est confirmée par des observations qui montrent (i) d'une part que l'utilisation différentielle des promoteurs est coordonnée avec l'avancement de la différenciation (ii) et que d'autre part, dans les cellules ES mâles en cours de différenciation, TFIIB est recruté sur P1, mais pas sur P2 lorsque l'expression de *Tsix* est interrompue (Navarro *et al.*, 2006). Le transcrit majoritaire dans des cellules somatiques est initié à partir de P2. Il est vraisemblablement inactif en termes de répression puisque la région des A-repeats participe à l'initiation de l'inactivation (Johnston *et al.*, 1998). L'utilisation alternative des deux promoteurs semble donc essentielle à la régulation de l'inactivation du chromosome X, toutefois l'intérêt de l'existence de ces deux promoteurs n'a pas été défini.

D'autre part, l'utilisation d'un système rapporteur a permis de démontrer que la région +79 à +320 du gène *Xist* permet de produire un ARN de 1,6 kb (Zhao *et al.*, 2008). Ce promoteur qui n'était pas décrit précédemment produit un petit transcrit contenant la région des A-repeats. Cet ARN a été nommé *Arep*. Il a été détecté dans les cellules ES mâles et femelles non différenciées, c'est-à-dire avant inactivation du chromosome X. Son rôle n'a pas été encore identifié.

## V.2 Structure de la région des A-repeats

### V.2.1 Prédiction informatique de Wutz *et al.*

Cette région étant essentielle pour la mise en place de l'inactivation comme nous l'avons mentionné précédemment, et comme la structure secondaire d'un ARN conditionne son action, la structure 2D potentielle de chaque répétition a tout d'abord été prédite par voie informatique. Selon le modèle obtenu, les séquences GCCC et GGGC formeraient une petite structure en tige-boucle et les triplets (C/U)GG et (C/U)U(G/A) de la partie 3' des répétitions en formeraient une autre (**Figure 42.A**) (Wutz *et al.*, 2002). Néanmoins, la seconde structure tige-boucle n'est pas thermodynamiquement stable et la structure 2D globale des A-repeats

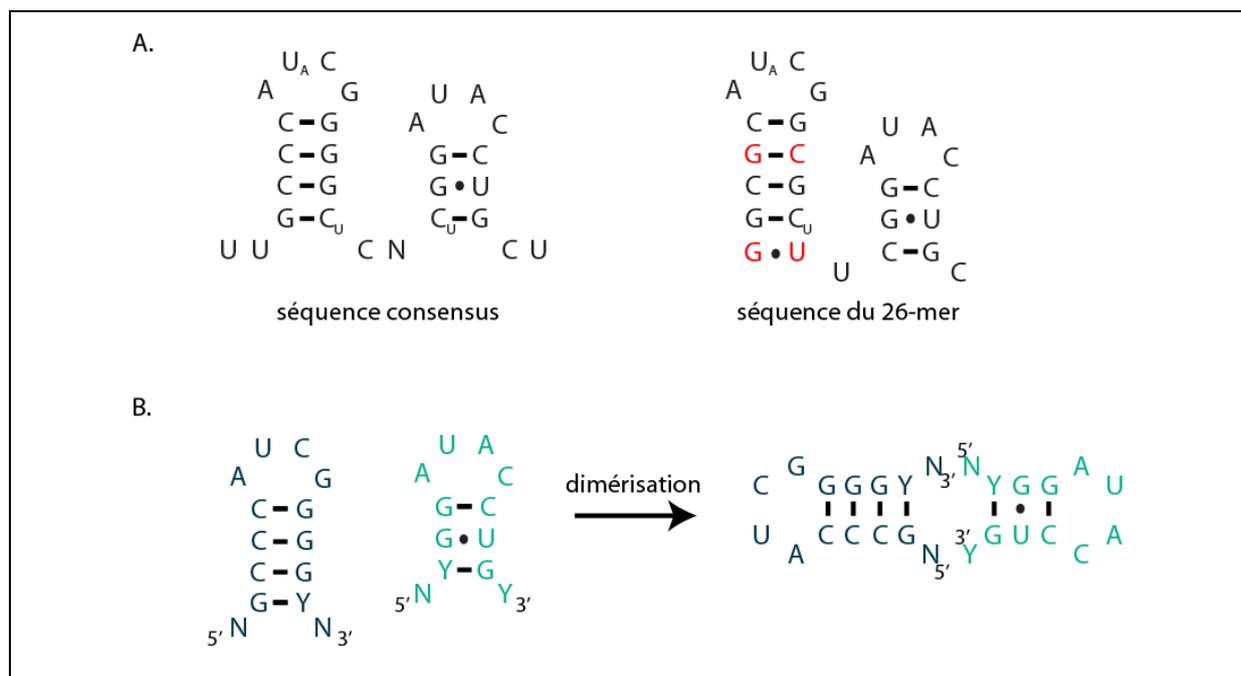


**Figure 42 : Analyse de la région des A-repeats**

A. Prédiction de la structure secondaire d'une répétition de la région des A-repeats  
 B. Séquence des monomères de l'élément de base de la région des A-repeats de type sauvage ou muté par Wutz et ses collègues. Le nombre de répétitions présentes dans les transgènes étudiés est également indiqué. La capacité des séquences à instaurer une inactivation du chromosome X est indiquée par les signes + et -.  
 D'après Wutz *et al.*, 2002

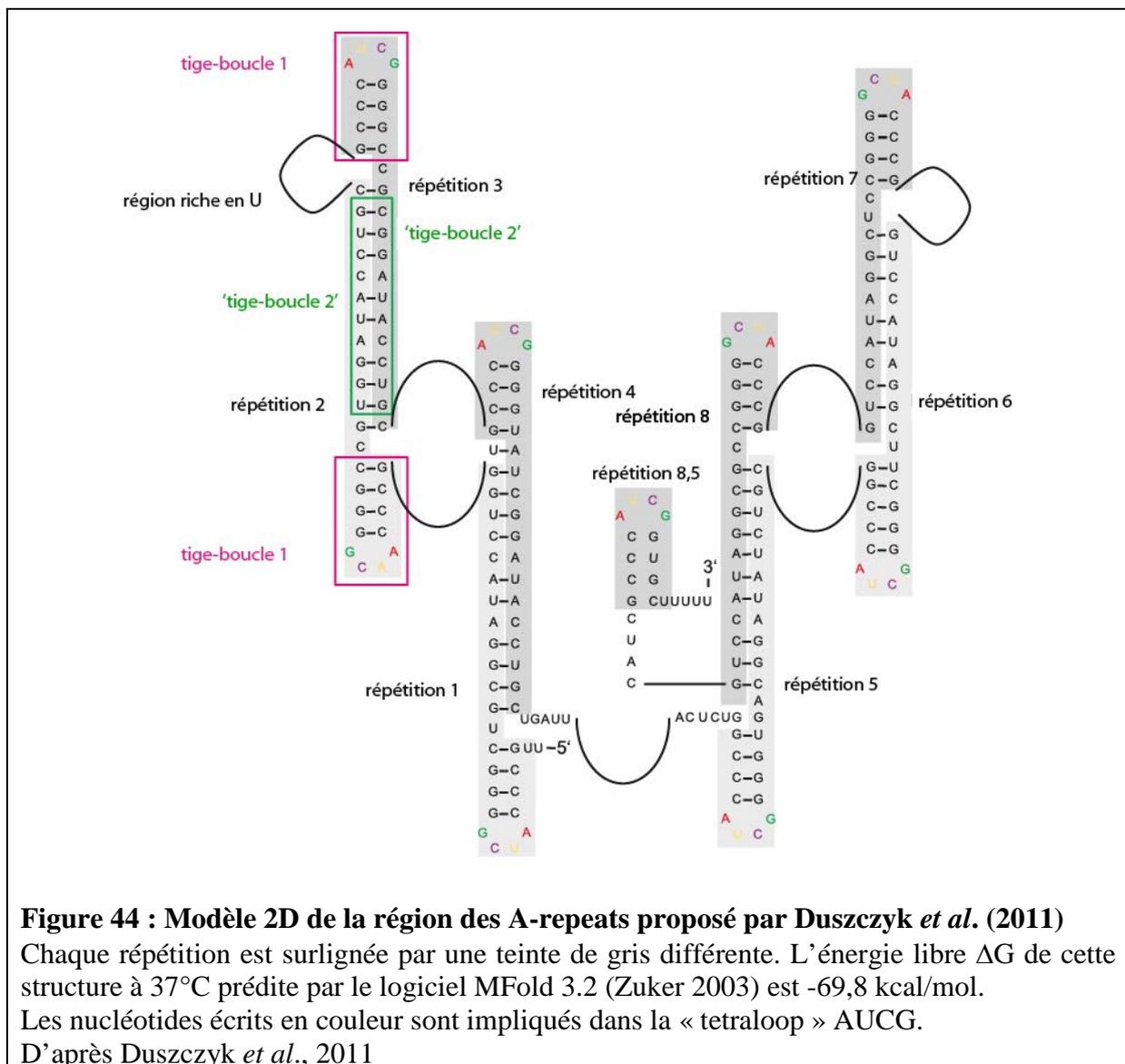
V.2.2 Modèle proposé par Duszczuk *et al.*

Récemment, il a été montré par résonance magnétique nucléaire qu'un élément répété de la région des A-repeats long de 26 nt est capable de former la première structure tige-boucle, alors que le deuxième élément dans cette situation *in vitro* participerait à la dimérisation de l'ARN (**Figure 43.B**) (Duszczuk *et al.*, 2008). En plus du fait que cette répétition ait été étudiée hors de son contexte, il faut remarquer que la séquence utilisée pour cette analyse ne correspondait pas à la séquence consensus des répétitions (**Figure 43.A**). En effet, elle comporte les séquences GGCGC et GCGCU au lieu des séquences GCCC et GGGC, ce qui pourrait influencer sur la structure observée.

**Figure 43 : Modèle de dimérisation des éléments répétés de la région des A-repeats proposé par Duszczuk *et al.* (2008)**

A. Comparaison entre la séquence consensus des éléments répétés de la région des A-repeats et celle utilisée par Duszczuk *et al.*, 2008. Les nucléotides différents sont indiqués en rouge.  
 B. Modèle de dimérisation de la partie 3' des éléments répétés de la région des A-repeats.  
 D'après Duszczuk *et al.*, 2008

Les membres de cette équipe ont poursuivi leur étude RMN et ils ont observé que la boucle composée des 4 nucléotides AUCG (« AUCG tetraloop ») formée par la première structure tige-boucle d'un élément répété de la région des A-repeats est remarquablement stable (Duszczek *et al.*, 2011). Ils soutiennent de nouveau que la partie 3' ne se structure pas en tige-boucle comme la prédiction le proposait, mais qu'elle est impliquée dans la formation d'un duplex inter-répétition. L'expression de transgènes de *Xist* portant des mutations déstabilisant cette dimérisation intermoléculaire induit l'absence d'inactivation du chromosome X *in vivo* (Duszczek *et al.*, 2011). Ils en concluent que la région des A-repeats semble être stabilisée par des interactions inter-répétitions, formant ainsi une plateforme de multimérisation et de reconnaissance des « tetraloops » par des facteurs agissant en *trans*. Le modèle de la structure secondaire de la région des A-repeats entière qu'ils proposent est présenté dans la **Figure 44**.

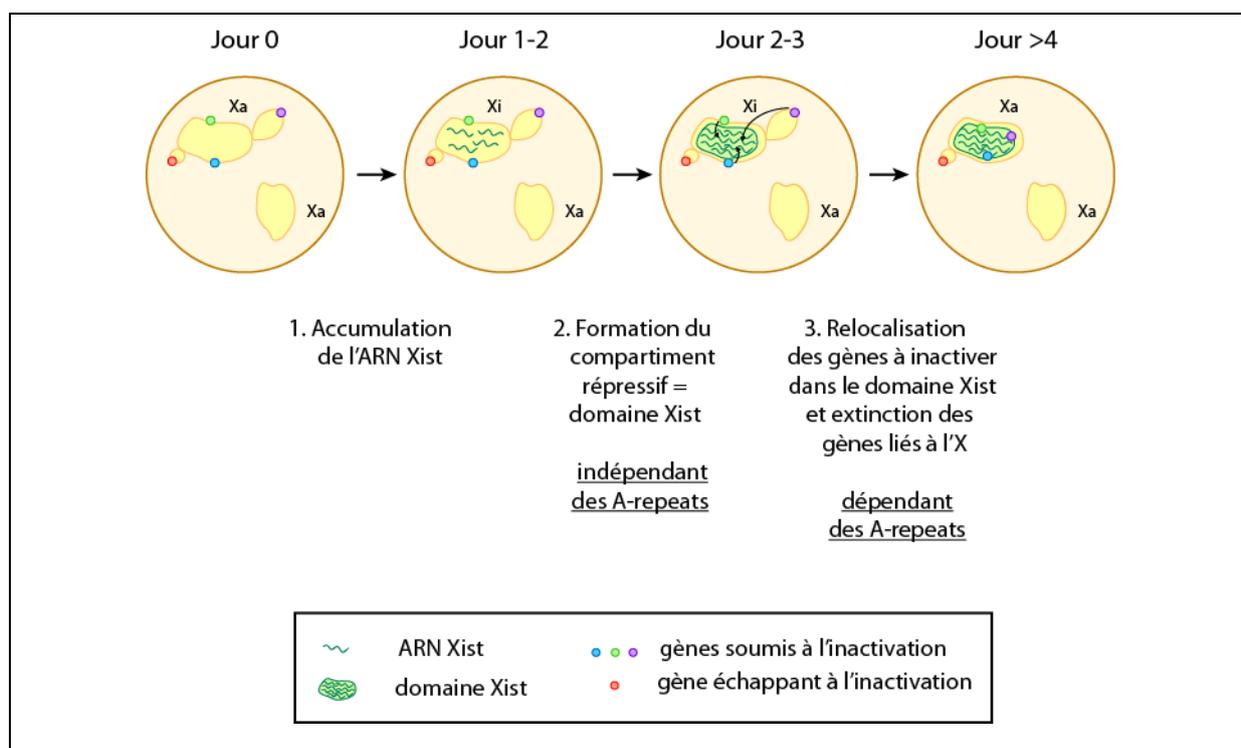


Entre les modèles de dimérisation (2008) et de la région entière (2011) de Duszcyk *et al.*, notre laboratoire a lui aussi proposé un modèle de structure 2D de la région complète des A-repeats. Elle fait également intervenir des interactions entre répétitions mais ne conduit pas tout à fait à la même structure finale. Nous y reviendrons dans la partie Résultats.

### V.3 Rôles de la région des A-repeats

#### V.3.1 Formation du domaine Xist

Nous avons déjà évoqué la formation d'un domaine transcriptionnellement silencieux, appelé domaine Xist (cf **Figure 29**). Chaumeil et ses collègues ont étudié la cinétique précise des premiers événements de sa mise en place lors de la différenciation *in vitro* de cellules ES murines femelles. Après exclusion rapide de l'ARN polymérase II et des facteurs généraux de transcription hors du domaine, la répression des gènes est liée à leur repositionnement vers l'intérieur du domaine Xist. Ils ont proposé un modèle dans lequel l'inactivation peut être divisée en 2 étapes. 1/ une phase de répression transcriptionnelle de séquences répétées non géniques ayant lieu de manière concomitante à la mise en place du domaine Xist. Cette étape ne nécessite pas la présence de la région des A-repeats. 2/ une 2<sup>ème</sup> phase de relocalisation des gènes dans ce compartiment et de leur inactivation. Cette étape est quant à elle dépendante de la région des A-repeats (**Figure 45**) (Chaumeil *et al.*, 2006). Il a été montré que les gènes échappant à l'inactivation peuvent être regroupés en clusters (Lopes *et al.*, 2011).



**Figure 45 : Rôle de la région des A-repeats dans la relocalisation des gènes à inactiver dans le domaine Xist**

Dans les cellules ES femelles indifférenciées, les gènes (ronds colorés) des chromosomes X (en jaune) sont actifs. Au début de la différenciation, l'ARN Xist commence à recouvrir le chromosome X choisi pour être inactivé. La formation du domaine Xist est associée à l'exclusion de la machinerie transcriptionnelle et à la répression des séquences répétées, étape indépendante de la région des A-repeats. Les gènes du chromosome X à inactiver sont ensuite relocalisés dans le domaine Xist, étape dépendante de la région des A-repeats. Les gènes échappant à l'inactivation (ronds rouges) sont maintenus hors du domaine.

D'après Chaumeil *et al.*, 2006

### V.3.2 Recrutement du complexe PRC2

Des expériences d'immuno-précipitation réalisées dans la perspective d'identifier les partenaires protéiques de la région des A-repeats ont permis de montrer que des composants du complexe PRC2 (Ezh2 et Suz12) se lient à l'ARN Xist (Zhao *et al.*, 2008). Dans les cellules ES non différenciées, le PRC2 interagirait uniquement avec la région des A-repeats. Après différenciation (et donc surexpression de *Xist*), il interagirait en plus avec la région 3' de Xist. Comme cette fixation à la région des A-repeats n'est pas retrouvée après le 3<sup>ème</sup> jour de différenciation, il a été suggéré que le complexe PRC2 soit recruté par la région des A-repeats avant et au début de la différenciation puis qu'il soit transféré sur la chromatine afin de catalyser la mise en place des H3K27me3.

Le recrutement du complexe PRC2 par la région des A-repeats a été confirmé par S. Maenner. Par des expériences de chromatographie d'affinité de RNP suivies d'identification par spectrométrie de masse, associées à des expériences immunoprécipitation et western-blot, il a démontré que la région des A-repeats est capable de s'associer à tous les composants du PRC2 (Suz12, Ezh2, Eed, RbAp46 et 48) dans des extraits nucléaires de cellules ES murines (Maenner *et al.*, 2010). Il a également analysé leur fixation en fonction du nombre de répétitions contenues dans la région. Un minimum de 4 répétitions est nécessaire à la fixation de tout le complexe, même s'il semble que la région entière permette un recrutement plus efficace de Suz12. Ce recrutement du complexe PRC2 nécessitant la présence de 4 répétitions amène la question de la structure de la région des A-repeats et de son importance dans les fonctions de celle-ci. Nous reviendrons sur ce point dans la partie Résultats.

### V.3.3 Rôle dans l'épissage de Xist

Comme nous en avons déjà parlé, la région des A-repeats a été identifiée comme une région cruciale pour la mise en place de l'inactivation (Wutz *et al.*, 2002). Plus récemment, il a été montré que la délétion monoallélique de cette région dans des cellules ES femelles

murines conduit toujours à l'inactivation du chromosome non délété (Royce-Tolland *et al.*, 2010), comme cela avait déjà été observé par Hoki *et al.* (2009). Cependant, l'équipe de B. Panning a également établi qu'il s'agit d'une inactivation non aléatoire primaire, c'est-à-dire que le chromosome muté n'est jamais sélectionné pour être inactivé, et non pas d'une mort des cellules dans lesquelles le chromosome muté aurait été choisi pour l'inactivation. Le mécanisme par lequel la cellule choisit le futur Xi est donc influencé par la région des A-repeats. Les membres de cette équipe ont alors étudié le taux d'ARN Xist issus du chromosome délété de la région des A-repeats. Ils ont observé que les transcrits Xist ainsi produits présentent un défaut de maturation. En effet, les produits de RT-PCR obtenus ne migrent pas à la taille attendue et l'analyse de leur séquence a révélé l'exclusion de certains exons dans les transcrits finaux (Royce-Tolland *et al.*, 2010). La région des A-repeats semble donc être impliquée dans la production ou l'accumulation des transcrits Xist correctement épissés.

Par ailleurs, des expériences d'immuno-précipitation d'une répétition consensus de la région des A-repeats avec un extrait nucléaire ont révélé la présence de la protéine ASF/SF2 (Royce-Tolland *et al.*, 2010). ASF/SF2 joue un rôle dans le processus d'épissage de Xist puisque sa délétion conduit à une augmentation du taux de transcrits non épissés et à une diminution de celui des transcrits épissés. L'interaction avec cette protéine semble être dépendante de la séquence de l'ARN et non de sa structure, et nécessiterait au moins la présence d'une répétition entière. La région des A-repeats comprenant 7 ou 8 répétitions (chez la souris et l'homme respectivement), il serait intéressant de comprendre le mode de recrutement de ASF/SF2 sur cette région. De plus, comme la région des A-repeats se trouve à l'extrémité 5' de l'exon 1 de Xist, et à quasiment 10 kb du premier intron, il faut se poser la question du mécanisme permettant de faire le lien entre le recrutement de ASF/SF2 sur cette région et son action sur l'épissage d'introns si éloignés.

Afin de mieux appréhender le rôle de la protéine ASF/SF2 dans l'épissage de l'ARN Xist, nous allons consacrer la partie suivante à la présentation du processus d'épissage et de ses régulations.

## **VI. L'épissage**

Les transcrits des gènes nucléaires sont produits par l'ARN polymérase II et sont soumis à des processus complexes de maturation pour convertir les pré-ARN messagers (pré-

ARNm) en ARN messagers (ARNm) fonctionnels. Avant qu'un ARNm puisse être transporté en dehors du noyau pour être traduit, il doit être coiffé au niveau de son extrémité 5', épissé de ses introns (séquences non codantes) et modifié au niveau de son extrémité 3' par un clivage et l'ajout d'une queue poly(A). L'épissage correspond à l'excision des introns du transcrit primaire puis à la jonction des exons. Il est dit constitutif lorsqu'un exon est systématiquement inséré dans le transcrit mature. A l'inverse, l'épissage alternatif permet de produire des transcrits différents à partir d'un seul gène, ceci du fait de l'existence de plusieurs sites potentiels d'épissage en compétition les uns avec les autres au sein du pré-ARNm. Certains exons seront choisis et inclus dans l'ARNm, alors que d'autres ne le seront pas. L'épissage alternatif constitue un mode particulier de régulation de l'expression des gènes. Ce processus contribue grandement à la diversité du protéome et fournit une part d'explication au différentiel entre le nombre de gènes et celui des protéines. Pour atteindre cette flexibilité, la sélection des sites d'épissage chez les eucaryotes supérieurs a évolué afin de pouvoir dépendre de multiples paramètres comme la force du site d'épissage, la présence ou l'absence de régulateurs d'épissage, la structure secondaire de l'ARN, l'architecture exon/intron et le processus de synthèse du pré-ARNm.

L'élimination des introns est un processus très complexe qui nécessite l'assemblage d'un macrocomplexe, le spliceosome, composé de 5 petits ARN nucléaires snRNA (U1, U2, U4, U5 et U6) et d'un très grand nombre de protéines (plus de 100 protéines différentes) autour de l'intron à éliminer (pour revue, Will et Luhrmann, 2001). Il existe en fait 2 types d'introns chez les eucaryotes supérieurs : les introns dits majeurs, éliminés par le spliceosome que nous décrirons par la suite et les introns dits mineurs, rencontrés moins fréquemment. Ces introns sont éliminés par un autre type de spliceosome, renfermant les snRNA U11 et U12 à la place des snRNA U1 et U2, les snRNA U4atac et U6atac à la place des snRNA U4 et U6, et le snRNA U5 (pour revue, Tarn et Steitz, 1997). Nous ne nous intéresserons ici qu'au spliceosome majeur.

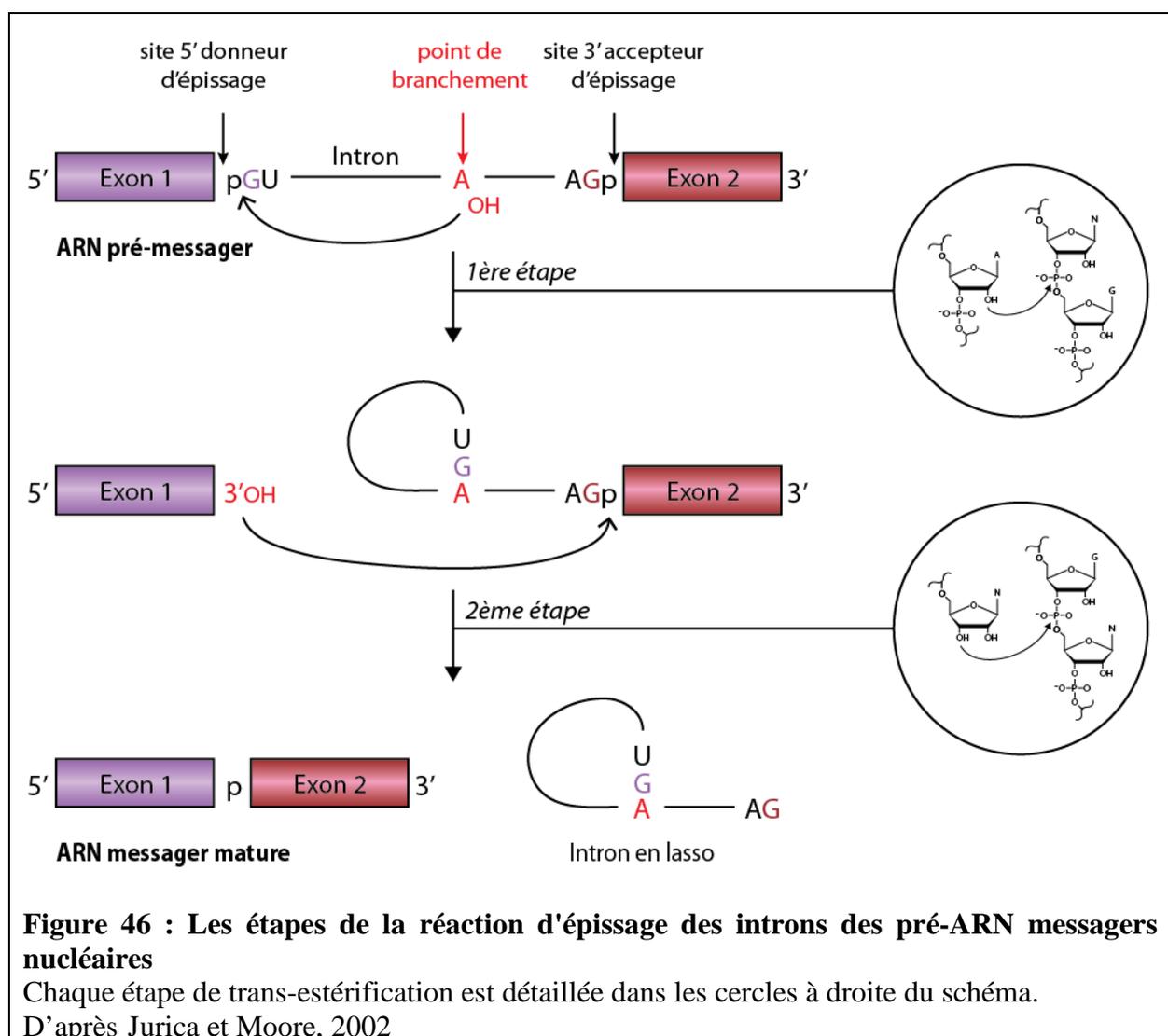
## **VI.1 Description générale de l'épissage des ARN**

### **VI.1.1 Les étapes de la réaction d'épissage**

La réaction d'épissage se déroule en deux étapes successives de trans-estérification (**Figure 46**). La première étape consiste en l'attaque nucléophile du phosphate en 5' de l'intron (site 5' donneur) par l'hydroxyle en 2' d'un résidu adénosine situé dans l'intron au niveau d'un élément de séquence conservé, la boîte de branchement. Deux intermédiaires

réactionnels sont alors produits, d'une part l'exon 1, avec une extrémité 3'OH libre, et d'autre part un intermédiaire contenant l'exon 2 et l'intron structuré en lasso, dont la guanosine 5' terminale est covalamment liée au point de branchement par une liaison 2'-5' phosphodiester.

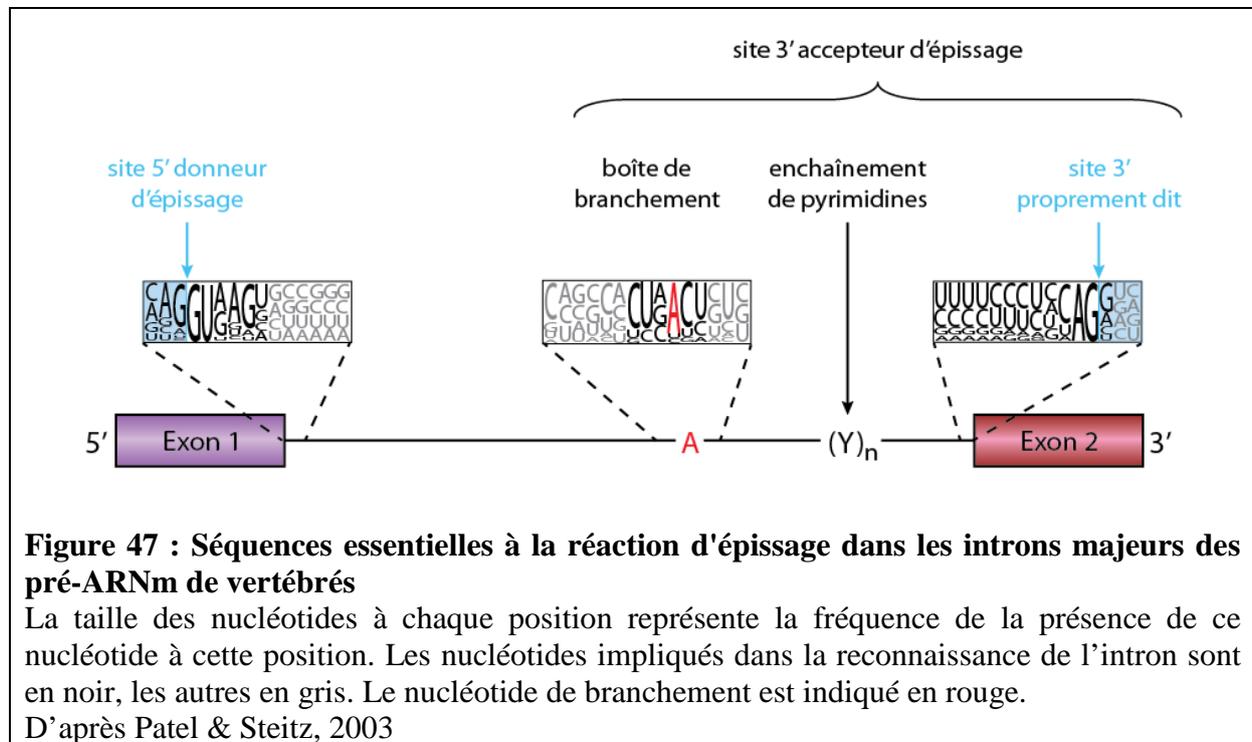
Dans la seconde étape de trans-estérification, l'extrémité 3'OH de l'exon 1 réalise une attaque nucléophile du phosphate situé à l'extrémité 3' de l'intron (site 3' accepteur d'épissage). Ceci conduit à la ligation des deux exons et à la libération de l'intron sous forme de lasso. Ce sont les produits finaux de la réaction d'épissage. L'intron est alors linéarisé et dégradé (Jurica et Moore, 2002).



### VI.1.2 Les séquences définissant les introns

L'élimination des introns nécessite un ensemble de séquences conservées permettant leur reconnaissance par la machinerie d'épissage. Malgré la grande variabilité des séquences

introniques des pré-ARNm des eucaryotes supérieurs, la comparaison des séquences situées aux extrémités des introns a permis de définir des séquences consensus pour les sites 5' et 3' d'épissage (**Figure 47**).



#### a. Le site 5' donneur d'épissage

La bordure exon/intron est appelée site 5' donneur d'épissage. La comparaison de ces bordures exon/intron a permis d'établir la séquence consensus: AG/GURAGU (R = purine et / = jonction exon/intron), dans laquelle le dinucléotide **GU** en 5' de l'intron est le motif le plus conservé (Moore et Sharp, 1993) (Figure 47). En effet, des mutations ponctuelles du site 5' d'épissage, notamment au niveau du dinucléotide GU très conservé, diminuent ou abolissent l'épissage normal. Ces mutations entraînent généralement l'utilisation d'un site 5' cryptique qui n'est pas utilisé normalement et qui est généralement proche du site 5' naturel. La reconnaissance d'un site 5' cryptique par le spliceosome n'est pas dépendante simplement de sa séquence mais plus d'un contexte global favorable à l'utilisation de ce site, ce qui reflète une grande flexibilité dans les pré-ARNm de mammifères.

#### b. Le site 3' accepteur d'épissage

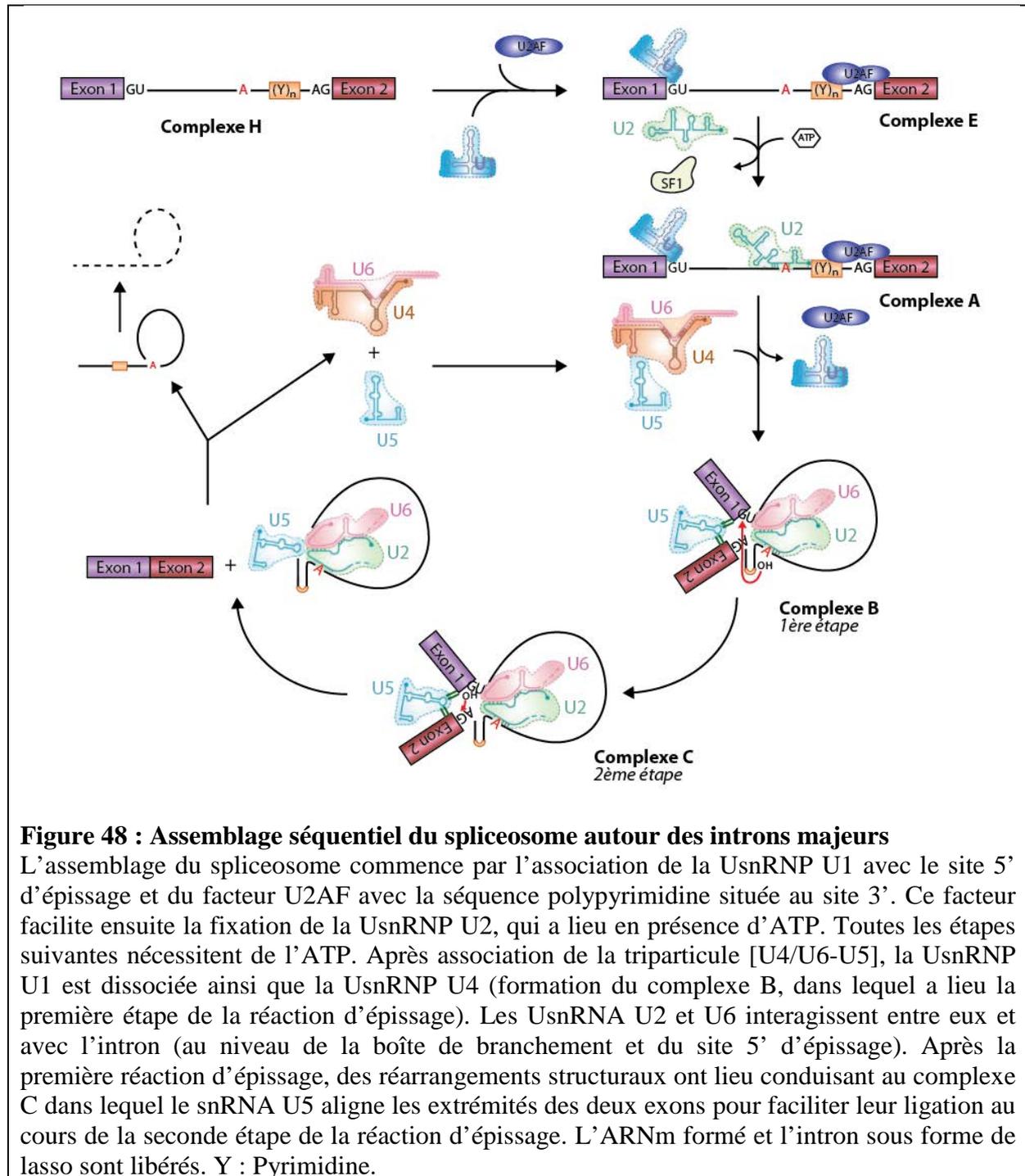
La définition des sites 3' accepteurs d'épissage est plus complexe. En effet, chez les vertébrés, ils renferment 3 éléments distincts : la boîte de branchement, une séquence polypyrimidine et le site 3' proprement dit (**Figure 47**).

La boîte de branchement est située 18 à 40 nts en amont du site 3' d'épissage. Sa séquence consensus établie est très dégénérée : YNYURAC (Y = pyrimidine). Le résidu adénosine conservé (en gras), correspondant au point de branchement, est responsable de la première réaction de trans-estérification. Dans quelques cas, ce résidu peut être une cytosine ou une uridine. La séquence polypyrimidine correspond à un enchaînement de 10 à 20 résidus pyrimidines, les résidus U étant plus favorables à sa fonction. Elle est localisée en amont du site 3' accepteur d'épissage. Enfin, le site 3' d'épissage, proprement dit, se situe à la bordure intron/exon et est caractérisé par une courte séquence conservée YAG/G.

#### VI.1.3 Les composants et l'assemblage du spliceosome

Le spliceosome est un macrocomplexe ribonucléoprotéique (RNP) architecturé par l'assemblage de 5 particules ribonucléoprotéiques appelées UsnRNP (snRNP U1, U2, U4, U5, U6). Chacune de ces UsnRNP s'architecture elle-même autour d'un ou plusieurs petits ARN nucléaires riches en uridine (ou UsnRNA), et comprend un ensemble de 7 protéines Sm ou Sm-like (Lsm) ainsi que plusieurs protéines spécifiques à chaque UsnRNP. Comme décrit par la suite, les UsnRNP et les protéines associées au spliceosome reconnaissent les séquences nécessaires à la réaction d'épissage décrites dans le paragraphe précédent. Des interactions ARN-ARN et ARN-protéines se forment alors de manière ordonnée et transitoire autour des sites d'épissage, conduisant à la formation du spliceosome au sein duquel ont lieu les 2 réactions de trans-estérification (pour revues, Hastings et Krainer, 2001 ; Will et Luhrmann, 2001).

L'assemblage du spliceosome autour de l'intron se fait de manière séquentielle. Ainsi chez les vertébrés, les premières étapes de l'épissage consistent à reconnaître les sites donneur et accepteur d'épissage. S'assemble ensuite le reste de la machinerie d'épissage qui confère au pré-ARNm une configuration spatiale adéquate pour permettre la réaction d'épissage. Les études *in vitro* de l'assemblage du spliceosome sur le pré-ARNm ont permis de mettre en évidence la formation successive de 5 complexes ayant une mobilité électrophorétique différente lorsqu'ils sont fractionnés en gel non dénaturant (Jurica *et al.*, 2002) (**Figure 48**).



Ces complexes H, E, A, B et C correspondent à l'assemblage ordonné des différentes UsnRNP et protéines associées sur le pré-ARNm. Ainsi, dès leur synthèse, les transcrits Pol II sont associés à un ensemble de protéines formant un complexe ribonucléoprotéique appelé complexe H. Le processus dynamique est ensuite initié par la reconnaissance et l'interaction de la snRNP U1 avec le site 5' donneur. Le complexe ainsi formé est appelé complexe d'engagement ou complexe E. En présence d'ATP, la snRNP U2 interagit avec la boîte de

branchement pour former avec la snRNP U1 le pré-spliceosome (complexe A). La tri-snRNP [U4/U6-U5] s'associe alors au complexe A pour former le complexe B\* qui subit des réarrangements structuraux nécessaires à la formation du complexe B capable de catalyser la 1<sup>ère</sup> étape de la réaction d'épissage (pour revue, Wahl *et al.*, 2009)). Lors de l'association de la tri-snRNP [U4/U6-U5], la particule snRNP U1 se dissocie, ce qui permet l'interaction du snRNA U6 avec le site 5' d'épissage. L'interaction entre les snRNA U4 et U6 est dissociée au profit d'une interaction entre les snRNA U2 et U6. Ces transitions structurales s'accompagnent d'un remaniement des protéines. La 1<sup>ère</sup> réaction a lieu. Elle est suivie de nouvelles transitions structurales. Enfin, la boucle terminale du snRNA U5 interagit avec les 2 extrémités des exons, ce qui permet de les rapprocher afin de les liguer.

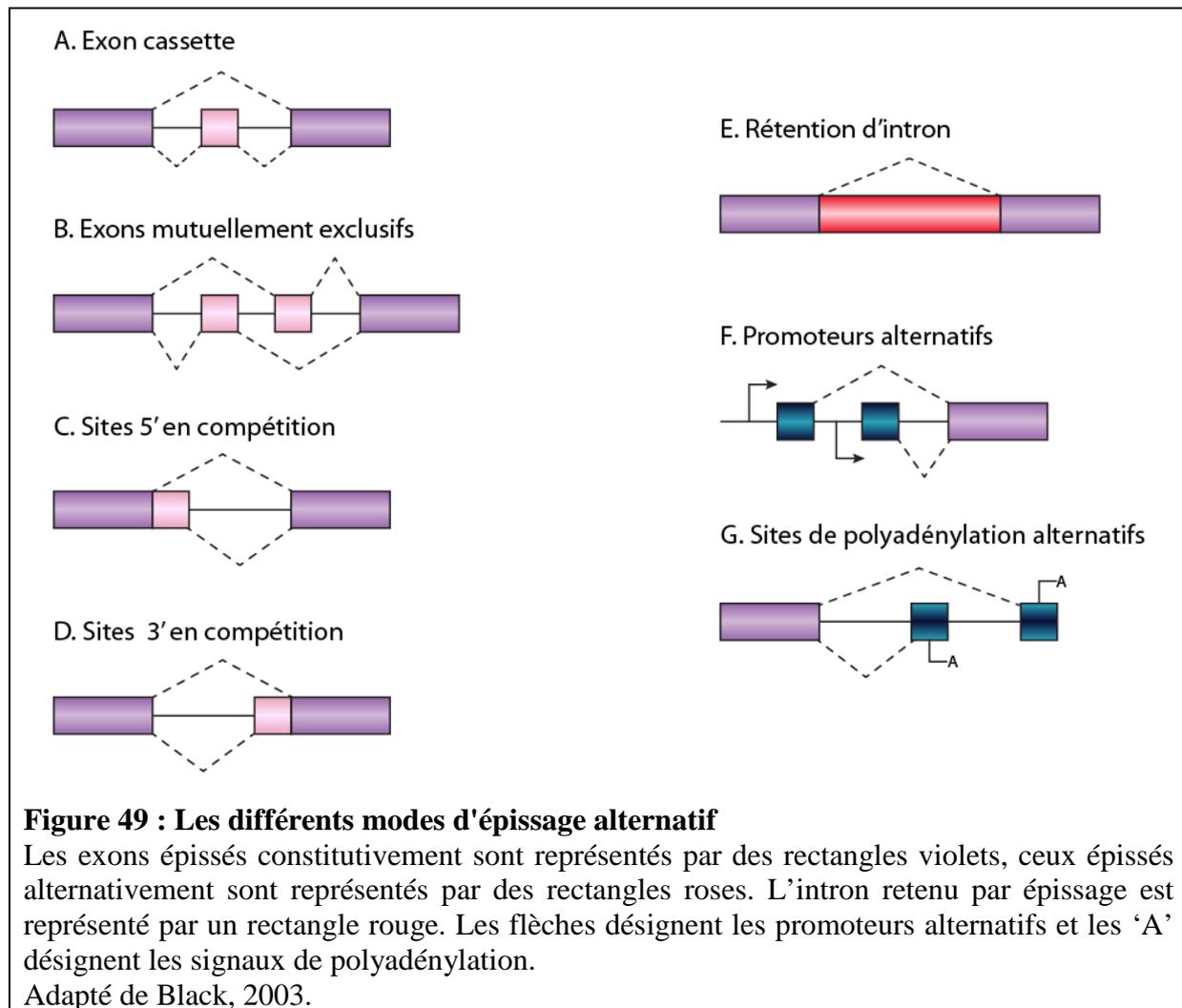
La fixation des UsnRNP sur le pré-ARNm fait intervenir un réseau complexe d'interactions ARN-ARN, ARN-protéine et protéine-protéine, impliquant des protéines appartenant ou non aux UsnRNP. Ainsi, au total, près de 300 protéines s'assemblent sur le pré-ARNm pour former le spliceosome (pour revue, Wahl *et al.*, 2009). Deux familles de protéines, les protéines SR et hnRNP jouent un rôle central dans l'ensemble du processus d'épissage, elles feront l'objet d'un paragraphe particulier.

## **VI.2 L'épissage alternatif**

Comme nous l'avons précisé précédemment, les eucaryotes supérieurs et les virus utilisent l'épissage alternatif pour accroître la capacité codante de leur génome. Chez l'Homme, l'épissage alternatif est un phénomène majeur, car il concerne plus de 60% des 30 000 gènes (Modrek et Lee, 2002).

### **VI.2.1 Les différents types d'épissage alternatif**

L'épissage alternatif s'appuie sur l'utilisation de sites 5' donneurs et 3' accepteurs d'épissage. Les profils d'utilisation de ces sites donnent lieu à plusieurs types d'épissage (**Figure 49**) (Black, 2003).



Les exons dits « cassettes » (**Figure 49.A**) sont les plus fréquemment rencontrés. Ces exons sont soit inclus, soit exclus de l'ARN messager mature. Ils sont en général de petite taille, exprimés de manière tissu-spécifique et codent un domaine supplémentaire permettant d'obtenir une isoforme de la protéine.

Les exons dits « mutuellement exclusifs » (**Figure 49.B**) ne sont jamais présents ensemble dans les ARNm matures, bien qu'ils aient des sites 5' et 3' fonctionnels.

Des sites d'épissage alternatifs en compétition les uns avec les autres peuvent être présents dans une séquence exonique et la sélection alternative de ces sites génère des isoformes où une ou plusieurs parties de l'exon sont incluses ou exclues de l'ARNm (**Figure 49.C et D**).

Parfois, l'intron peut ne pas être éliminé du transcrit mature (**Figure 49.E**). Dans ce cas, il fait partie de l'ARN messager qui est transporté dans le cytoplasme pour la traduction. Ce phénomène, généralement tissu-spécifique, résulte de l'échec de la machinerie d'épissage à reconnaître des sites d'épissage faibles flanquant les introns. La rétention d'intron est le type

d'épissage alternatif le moins connu et le moins étudié car l'ARNm obtenu est confondu avec des pré-ARNm partiellement ou non épissés.

L'initiation de la transcription au niveau de promoteurs différents, dits alternatifs, donne lieu à différents ARN messagers, de même l'épissage peut également conduire à l'utilisation de sites de polyadénylation différents (**Figure 49.F et G**).

Plusieurs types d'épissage alternatif peuvent être combinés au sein d'un même pré-ARNm. L'épissage alternatif est également régulé par la nature des sites d'épissage, la présence ou non de séquences régulatrices en *cis* et la concentration cellulaire des facteurs d'épissage (facteurs en *trans*) qui se lient à ces séquences régulatrices.

#### VI.2.2 La régulation de l'épissage alternatif

L'identification des exons au sein du génome est un véritable défi pour la machinerie d'épissage quand on considère la taille d'un exon (environ 140 nt chez l'homme) par rapport à celle d'un intron (en moyenne 3500 nt, mais pouvant aller jusqu'à 500 000 nt). D'autres informations et interactions sont donc nécessaires pour activer les sites d'épissage et définir les exons. Généralement, des protéines du spliceosome interagissent avec des séquences introniques ou exoniques, les éléments *cis* régulateurs, afin de stimuler ou inhiber l'excision des introns.

Les pré-ARNm présentant des sites d'épissage faibles, car très dégénérés, contiennent des éléments de séquence introniques et/ou exoniques localisés en amont ou en aval de l'exon régulé et qui agissent en *cis* pour stimuler ou au contraire, réprimer l'épissage. Ces séquences interagissent en général avec des facteurs protéiques *trans*-régulateurs.

Les séquences activatrices d'épissage peuvent être exoniques, ce sont alors des ESE (pour « Exonic Splicing Enhancer »), ou introniques, on parle alors d'ISE (pour « Intronic Splicing Enhancer »). A l'inverse, il existe aussi des séquences inhibitrices exoniques, appelées ESS (pour « Exonic Splicing Silencer »), ou introniques, appelées ISS (pour « Intronic Splicing Silencer »). Ces éléments régulateurs agissent par l'intermédiaire de leurs protéines partenaires. Les protéines hnRNP sont généralement associées aux éléments répresseurs, tandis que les protéines SR sont généralement associées aux éléments activateurs, mais nous verrons par la suite qu'il existe quelques exemples contraires.

### VI.3 Les régulateurs de l'épissage constitutif et alternatif

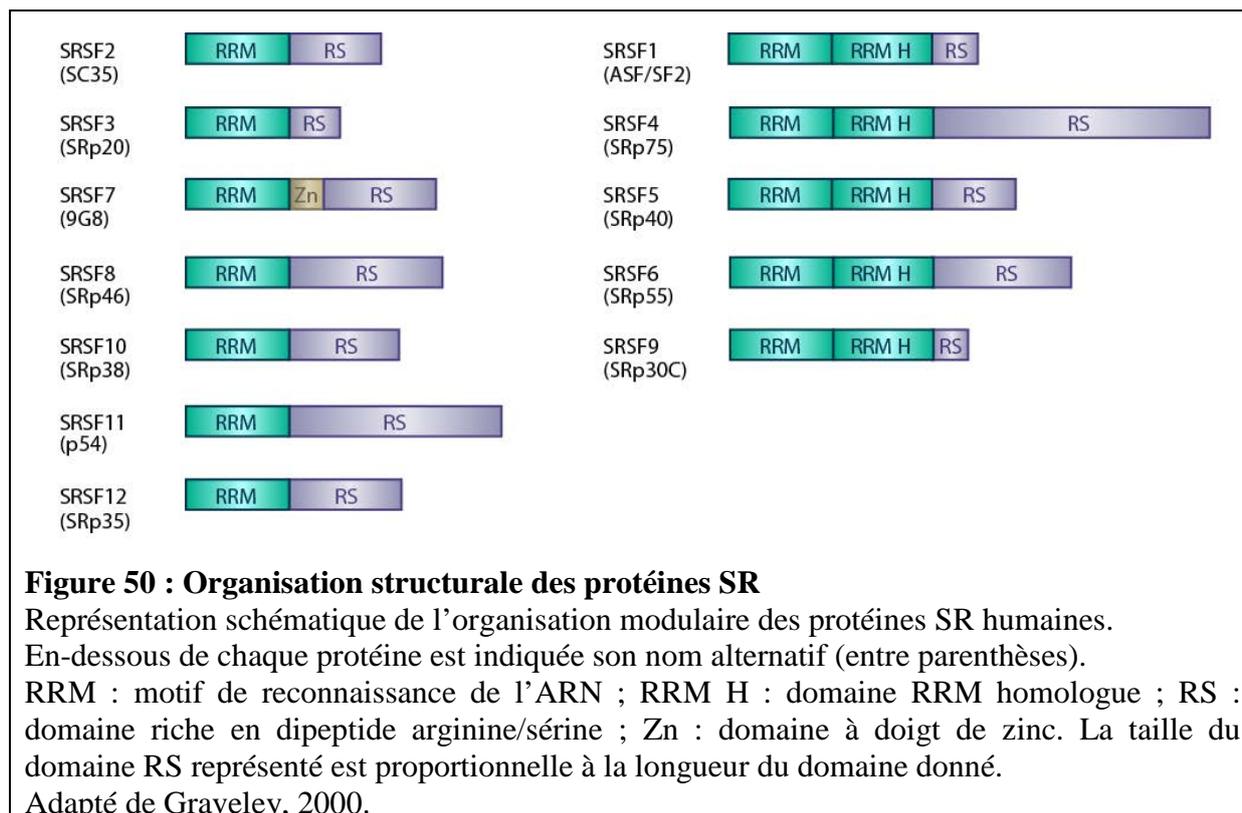
Certains facteurs d'épissage peuvent être impliqués aussi bien dans l'épissage constitutif qu'alternatif. D'autres, par contre, ne sont impliqués que dans l'épissage alternatif et sont nommés « facteurs régulateurs ». Certains facteurs d'épissage sont ubiquitaires, tandis que d'autres, et notamment les facteurs régulateurs, sont tissus spécifiques ou encore régulés au cours du développement. Dans tous les cas, il s'agit d'actions pour empêcher l'inclusion d'exons ou la promouvoir. On parle alors respectivement de facteurs répresseurs et activateurs. Nous n'allons nous intéresser ici qu'aux deux familles les plus importantes de protéines régulatrices, les protéines SR et les protéines hnRNP.

#### VI.3.1 Les protéines SR

Les protéines SR (Serine Arginine rich) constituent la famille la plus étudiée des facteurs d'épissage car elles sont essentielles à l'épissage constitutif. Ces protéines sont très conservées phylogénétiquement particulièrement chez les métazoaires et chez les plantes. Certains de ses membres sont impliqués aussi bien dans l'épissage constitutif qu'alternatif. D'autres sont considérés comme des facteurs régulateurs jouant des rôles spécifiques dans l'épissage alternatif (Long et Caceres, 2009). Cette famille comprend douze membres appelés SRSF1 à 12 (Zahler *et al.*, 1992 ; Long et Caceres, 2009).

##### a. Les caractéristiques communes des protéines SR

La caractéristique principale des protéines SR est leur ressemblance structurale (pour revue, Graveley, 2000). En effet, toutes sont composées d'un ou deux domaines de liaison à l'ARN de type RRM (pour « RNA Recognition Motif ») à l'extrémité N-terminale, suivi d'un domaine riche en dipeptides arginine-serine, appelé domaine RS à l'extrémité C-terminale (**Figure 50**).



Les protéines SR sont généralement associées à des séquences ESE. En effet, les protéines SR ont majoritairement une influence positive sur l'épissage et se fixent essentiellement dans les exons. Afin de comprendre les fonctions des protéines SR, plusieurs études, en particulier par la méthode SELEX (pour Systematic Evolution of Ligands by EXponential enrichment), ont tenté d'élucider les spécificités de liaison de chaque protéine SR. Le résultat global de ces expériences est que les protéines SR reconnaissent un vaste profil de séquences ARN. Les consensus des sites de fixation sont souvent dégénérés et dans beaucoup de cas, le motif reconnu par une protéine SR donnée est également reconnu par d'autres protéines SR (Tacke et Manley, 1995 ; Liu *et al.*, 1998).

Le domaine RS riche en enchaînements de sérines et d'arginines permet d'établir des interactions protéine-protéine par lesquelles les protéines SR s'associent entre elles ou avec les différents composants de la machinerie d'épissage contenant aussi un domaine RS (Wu et Maniatis, 1993). Néanmoins, il est apparu récemment que ce domaine pourrait aussi permettre une interaction avec l'ARN afin de définir les différents signaux d'épissage et faciliter leur reconnaissance par les composants du spliceosome (Shen et Green, 2004).

b. Les rôles des protéines SR dans l'épissage

- Effet activateur sur l'épissage

Les données actuelles montrent que le rôle des protéines SR dans l'épissage constitutif se situe à différents niveaux : (i) avec la snRNP U1, elles participent à la définition du site 5' (Crispino *et al.*, 1994) ; (ii) en se fixant au site 3', les protéines SR facilitent également la fixation de la snRNP U2 sur la séquence de branchement (Tarn et Steitz, 1995) ; (iii) en se fixant aux parties exoniques des sites 5' et 3', elles pourraient rapprocher ces extrémités pour faciliter la réaction d'épissage (Staknis et Reed, 1994).

De même, les protéines SR peuvent intervenir dans la régulation de l'épissage alternatif. En effet, certains éléments ESE sont capables de fixer des protéines SR différentes et de favoriser ainsi le choix d'un site par rapport à un autre.

Un autre modèle pour comprendre l'effet globalement activateur des protéines SR sur l'épissage est celui du phénomène d'antagonisme par rapport aux protéines hnRNP. Le rôle des protéines SR consiste alors à contrecarrer l'effet inhibiteur des protéines hnRNP liées à une séquence inhibitrice d'épissage ESS chevauchante ou juxtaposée à l'élément ESE. Dans ce cas, l'activation de l'épissage est tout à fait indépendante des domaines RS, seuls les domaines RRM des protéines SR sont nécessaires. Nous reparlerons de ce phénomène dans le paragraphe consacré aux protéines hnRNP.

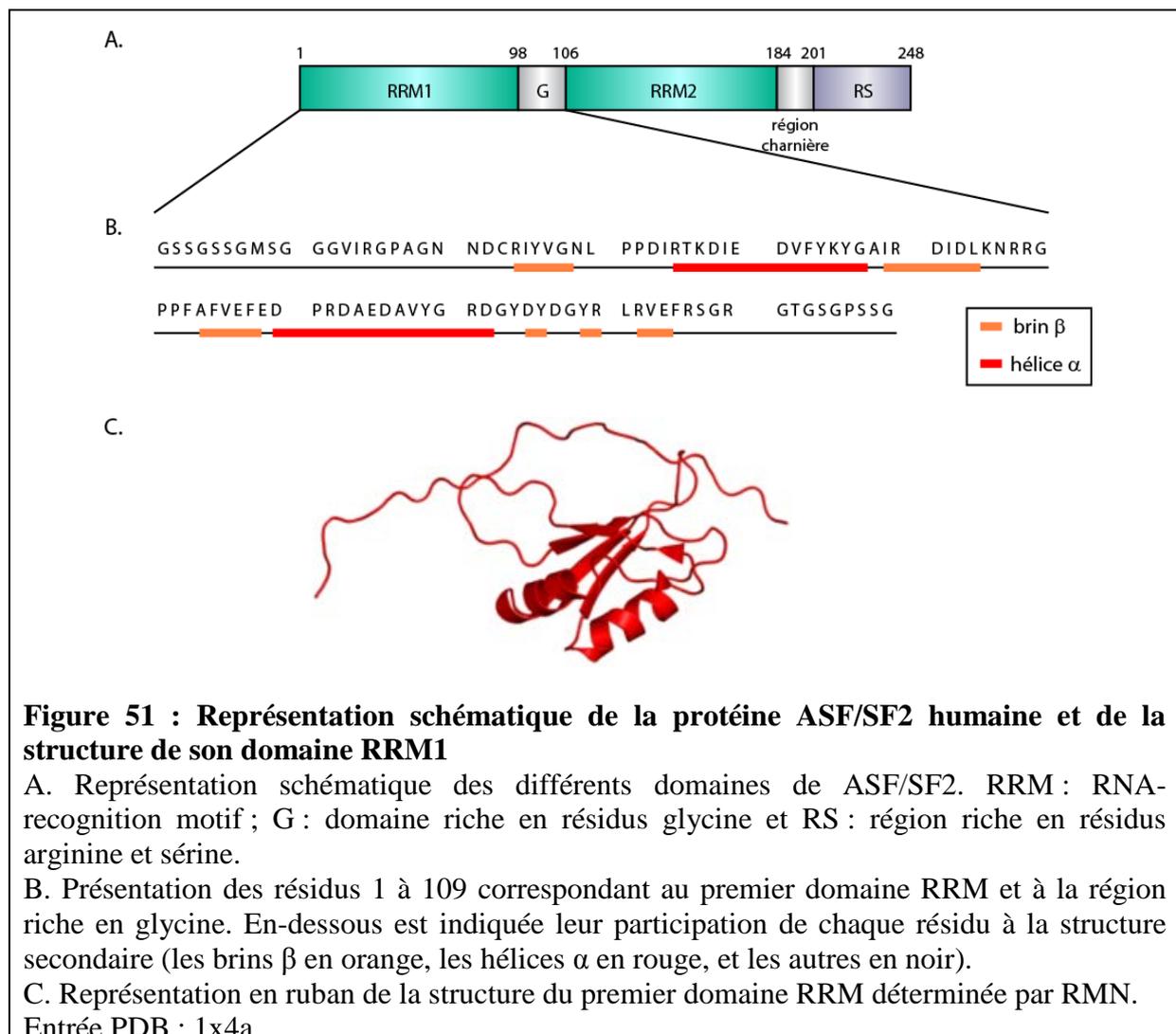
- Effet inhibiteur sur l'épissage

Bien que les protéines SR aient en majorité un effet activateur sur l'épissage, elles peuvent parfois avoir des effets inhibiteurs du fait de leur fixation sur des éléments de séquences requis pour l'assemblage des éléments du spliceosome ou de leur interaction avec certains de ces composants (pour revue, Pozzoli et Sironi, 2005). Plusieurs exemples montrent l'influence de certaines protéines SR dans l'exclusion d'exons et la protéine SRSF1 semble souvent impliquée : exclusion de l'exon IIIa de E1A de l'adénovirus en phase précoce d'infection par SRSF1 (pour revue, Akusjarvi et Stevenin, 2003) ; exclusion de l'exon 9 de CFTR par SRSF1, 4 et 6 (Buratti *et al.*, 2001) ; exclusion de l'exon 4 de CD45 par SRSF1 et 2 (Lemaire *et al.*, 1999).

c. La protéine ASF/SF2 (SRSF1)

La protéine ASF/SF2 a tout d'abord été identifiée comme un polypeptide nécessaire à l'épissage des pré-ARNm *in vitro*, notamment dans les étapes de clivage au niveau du site 5' d'épissage et de la formation du lasso (Krainer et Maniatis, 1985), ce qui lui a valu le nom de SF2 (Splicing Factor). Cependant, il a été montré qu'elle est également impliquée dans l'épissage alternatif, en influençant l'utilisation relative de deux sites 5' d'épissage en compétition d'un pré-ARNm du virus SV40 (Ge et Manley, 1990). C'est pourquoi elle porte désormais le nom complet d'ASF/SF2 (Alternative Splicing Factor/Splicing Factor 2).

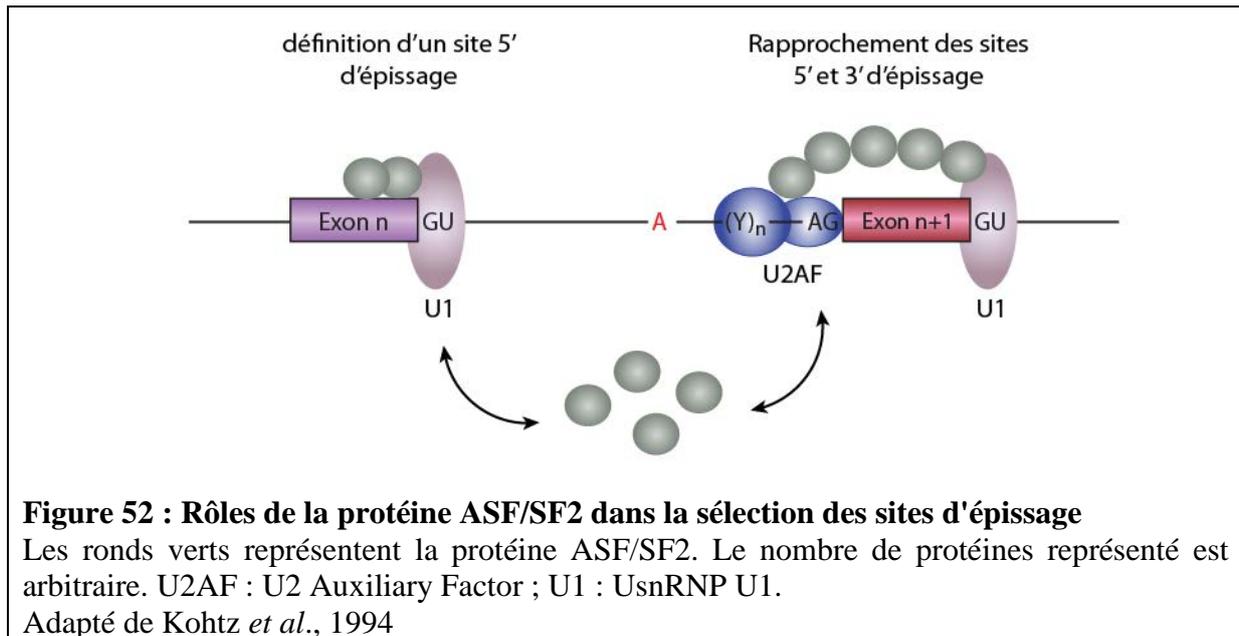
ASF/SF2 est une protéine de 33 kDa possédant 2 domaines de fixation à l'ARN et un domaine RS (**Figure 51**). Il a été montré que chaque domaine RRM possède une capacité individuelle de fixation à l'ARN, bien qu'elle soit réduite comparée à celle de la protéine native comprenant les 2 domaines (Caceres et Krainer, 1993 ; Zuo et Manley, 1993).



Il semblerait que les deux domaines RRM coopèrent afin de déterminer la spécificité de fixation à l'ARN de la protéine entière (Tacke et Manley, 1995). Plusieurs études ont été réalisées afin de permettre la caractérisation fonctionnelle de son domaine RS. Cependant, les résultats sont parfois en contradiction quant à la nécessité stricte de ce domaine pour sa fonction dans l'épissage (Caceres et Krainer, 1993 ; Zhu et Krainer, 2000). Il semblerait donc que la nécessité de la présence du domaine RS soit spécifique du substrat et qu'elle soit corrélée à la force des signaux d'épissage.

La protéine ASF/SF2 intervient à plusieurs niveaux dans le processus d'épissage et dans sa régulation. D'une part, elle est un facteur essentiel à la réaction d'épissage. Elle interagit notamment avec la snRNP U1 via son domaine RS, formant un complexe stable avec le pré-ARNm. Cette interaction assure la reconnaissance des sites 5' d'épissage, par collaboration entre les 2 protéines (Kohtz *et al.*, 1994). Elle a également été montrée comme augmentant la force de fixation de la snRNP U1 sur les sites 5' d'épissage (Eperon *et al.*, 1993). De la même manière, elle interagit avec le facteur U2AF au niveau du site 3' d'épissage grâce à son domaine RS (Wu et Maniatis, 1993). Ces interactions ASF/SF2-snRNP U1 au site 5' et ASF/SF2-U2AF au site 3' favorisent la juxtaposition des sites d'épissage, et donc un épissage efficace (**Figure 52**). Par ailleurs, l'action d'ASF/SF2 est influencée par son état de phosphorylation. En effet, cette modification post-traductionnelle est essentielle pour son activité puisqu'elle augmente sensiblement son interaction avec la snRNP U1 (Xiao et Manley, 1997).

D'autre part, elle agit également comme facteur régulateur de l'épissage alternatif. En effet, des expériences *in vivo* ont mis en évidence qu'un excès d'ASF/SF2 induit l'utilisation d'un site 5' donneur d'épissage cryptique (Caceres *et al.*, 1994 ; Wang et Manley, 1995). Cette découverte a également permis de montrer que les variations de concentration des facteurs d'épissage constituent aussi un élément de régulation de l'épissage alternatif.



Cette protéine a été récemment impliquée dans l'inactivation du chromosome X. En effet, l'identification de protéines associées à une répétition de la région des A-repeats incubée avec un extrait nucléaire (de cellules HeLa ou de cellules ES) a permis de mettre en évidence la présence d'ASF/SF2 (Royce-Tolland *et al.*, 2010). Son rôle dans l'épissage de l'ARN Xist est désormais une question à laquelle il faudra répondre. Nous y reviendrons dans la partie Résultats.

### VI.3.2 Les protéines hnRNP

Les protéines hnRNP (pour « heterogeneous nuclear ribonucleoproteins ») forment une famille de protéines très abondantes dans la cellule et structurellement variées. Elles ont initialement été décrites à travers leur capacité à lier les ARN nouvellement transcrits par l'ARN polymérase II, entraînant la formation des particules hétérogènes appelés hnRNP (Dreyfuss, 1986 ; Dreyfuss *et al.*, 1988). Cette famille comprend au moins une vingtaine de protéines et compte parmi ses membres hnRNP A1, B, C, F, G, H, K, et la hnRNP I, aussi appelée PTB (pour « Polypyrimidine Tract Binding ») dont nous reparlerons par la suite. Ces membres jouent des rôles variés, qui, au delà de la régulation de l'épissage, peuvent s'étendre à la transcription (Du *et al.*, 1998), à la maturation des extrémités 3' des ARNm (Kessler *et al.*, 1997 ; Bagga *et al.*, 1998) et à la traduction (Ostareck-Lederer et Ostareck, 2004).

#### a. Les caractéristiques communes des protéines hnRNP

Les protéines hnRNP se fixent à l'ARN par leurs domaines de type RRM, RGG ou KH (Dreyfuss *et al.*, 1993). Elles ont toutes une affinité assez large pour l'ARN, avec néanmoins une affinité plus forte pour certains types de séquences : séquences riches en C pour les hnRNP E, I et K, séquences riches en G pour les hnRNP A/B, E, H, H', 2H9, M et U et séquences riches en U pour les protéines hnRNP C et PTB. Certaines protéines hnRNP ont, en plus, la capacité de reconnaître avec une forte affinité des séquences spécifiques (hnRNP A/B, M, 2H9, PTB) (pour revue, Dreyfuss *et al.*, 2002).

Comme ces protéines se fixent préférentiellement aux séquences introniques des pré-ARNm (pour revue, Nakielny et Dreyfuss, 1999), elles éviteraient aux ARN non épissés d'être exportés dans le cytoplasme (Pinol-Roma et Dreyfuss, 1992). Ces protéines assureraient donc une rétention nucléaire des pré-ARNm.

#### b. Les rôles des protéines hnRNP dans l'épissage

Du fait de leur association aux pré-ARNm dès leur synthèse et jusque dans le cytoplasme, les protéines hnRNP jouent des rôles variés dans les différentes étapes de synthèse, maturation, transport, traduction et dégradation des ARNm (Dreyfuss *et al.*, 1993). Cette affinité pour l'ARN fait des protéines hnRNP des acteurs majeurs de l'épissage alternatif.

De part leur capacité à oligomériser le long de l'ARN, les protéines hnRNP jouent la plupart du temps le rôle d'inhibiteur de l'épissage en empêchant la fixation de la machinerie d'épissage. Cette inhibition prend effet sur les sites accepteurs et donneurs sans distinction. Les éléments sur lesquels se fixent les hnRNP peuvent être de nature exonique (ESS) ou intronique (ISS).

L'inhibition de l'épissage par les protéines hnRNP peut aussi provenir de leur capacité à contrecarrer l'effet activateur des protéines SR liées à une séquence activatrice d'épissage. Cet antagonisme provient du fait que leurs sites de fixation sur l'ARN sont en général juxtaposés voire chevauchants, ce qui empêche par encombrement stérique la fixation simultanée des 2 types de protéines (Smith et Valcarcel, 2000). Ce phénomène a été maintes fois observé et il a conduit à l'hypothèse largement répandue que l'épissage alternatif peut être régulé de manière tissu-spécifique grâce à des variations des taux relatifs de protéines SR et hnRNP.

De la même façon que les protéines SR présentent un effet minoritaire d'inhibiteurs de l'épissage, certaines protéines hnRNP, dans certains contextes, peuvent jouer un rôle positif sur l'épissage. Cependant, les mécanismes sont encore assez mal connus.

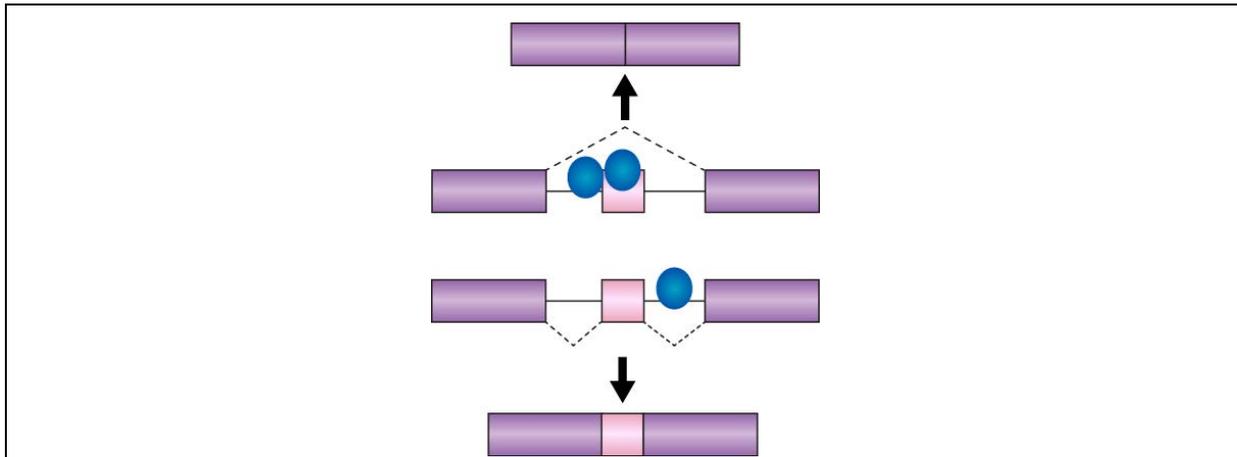
c. La protéine PTB (hnRNP I)

La protéine PTB est impliquée dans de nombreux processus cellulaires. Elle régule négativement l'épissage (Shen *et al.*, 2004), la polyadénylation des ARNm (Lou *et al.*, 1999 ; Castelo-Branco *et al.*, 2004), la stabilité des ARN (Wollerton *et al.*, 2004) et l'initiation de la traduction (Jang et Wimmer, 1990). Plus récemment, il a été proposé qu'elle induise en se fixant sur certains ARN, un changement de structuration qui permettrait ou au contraire empêcherait la fixation de facteurs protéiques (Kafasla *et al.*, 2009).

En ce qui concerne son action dans l'épissage, elle a tout d'abord été purifiée comme un polypeptide qui se lie aux régions riches en pyrimidines qui précèdent typiquement les sites 3' d'épissage des introns des ARNm eucaryotes (Garcia-Blanco *et al.*, 1989). Elle a ensuite été identifiée en tant que répresseur de l'épissage dans de nombreux modèles. Son action peut s'effectuer à plusieurs niveaux du processus : (i) par compétition directe avec le facteur U2AF65 pour la séquence riche en pyrimidines ; (ii) par propagation entre sites de haute affinité ; (iii) par formation de boucles d'ARN entre sites de haute affinité ; (iv) par inhibition de l'assemblage de complexes d'épissage le long d'introns ou d'exons (Wagner, 2001 ; Spellman et Smith, 2006).

Cependant, des études récentes par puces sensibles à l'épissage ont permis l'analyse d'un grand nombre d'événements d'épissage alternatif. Il en ressort qu'une minorité non négligeable sont des événements dans lesquels PTB joue un rôle activateur (Boutz *et al.*, 2007 ; Llorian *et al.*, 2010). Par comparaison entre ces événements, il a été possible de caractériser les différences entre les exons-cassette réprimés ou activés par PTB (Llorian *et al.*, 2010). Les exons réprimés par PTB sont associés à une fixation de PTB en amont ou dans l'exon régulé. En revanche, les exons activés par PTB sont associés à une fixation de PTB en aval de l'exon

**Figure 53.**



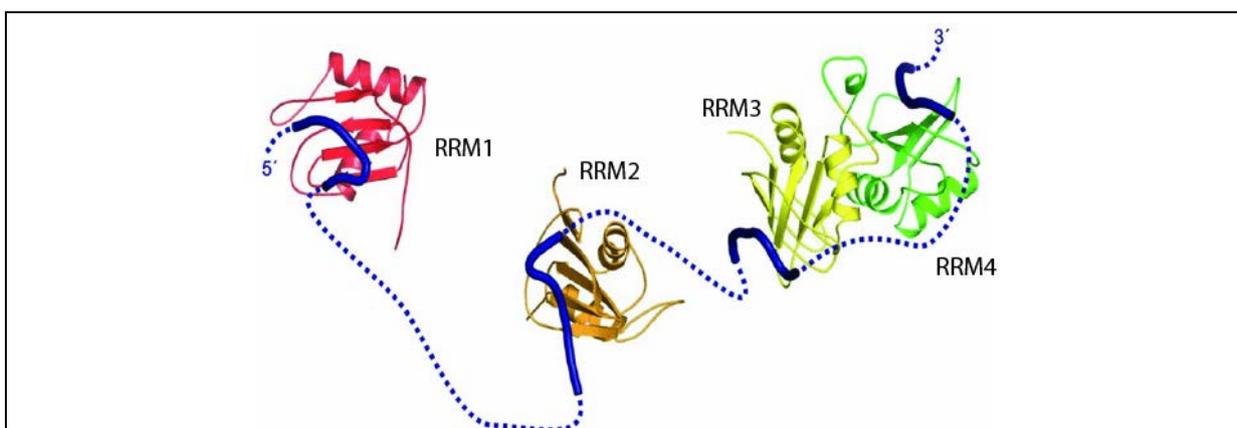
**Figure 53 : Représentation schématique des cibles de PTB et de la régulation de l'épissage associée**

En haut : les exons réprimés par PTB sont associés à une fixation de PTB sur l'intron en amont et/ou dans l'exon. En bas : les exons activés par PTB sont associés à une fixation de PTB en aval, à proximité immédiate de l'exon.

L'exon régulé est représenté par un rectangle rose, les rectangles violets représentent des exons constitutifs. PTB est indiquée par un ovale bleu.

D'après Kafasla *et al.*, 2012

Le poids moléculaire de PTB est de 57 kDa. Elle renferme 4 domaines de fixation à l'ARN de type RRM (RRM 1-4), qui sont chacun capables de fixer des séquences de type UCUU(C), UUCUCU ou CUCUCU, séquences présentes dans la région des A-repeats (Perez *et al.*, 1997). Les domaines RRM3 et RRM4 interagissent ensemble et forment un domaine compact et globulaire, alors que le reste de la protéine est linéaire, les domaines RRM1 et RRM2 se structurant selon le mode canonique de structuration des RRM (**Figure 54**).

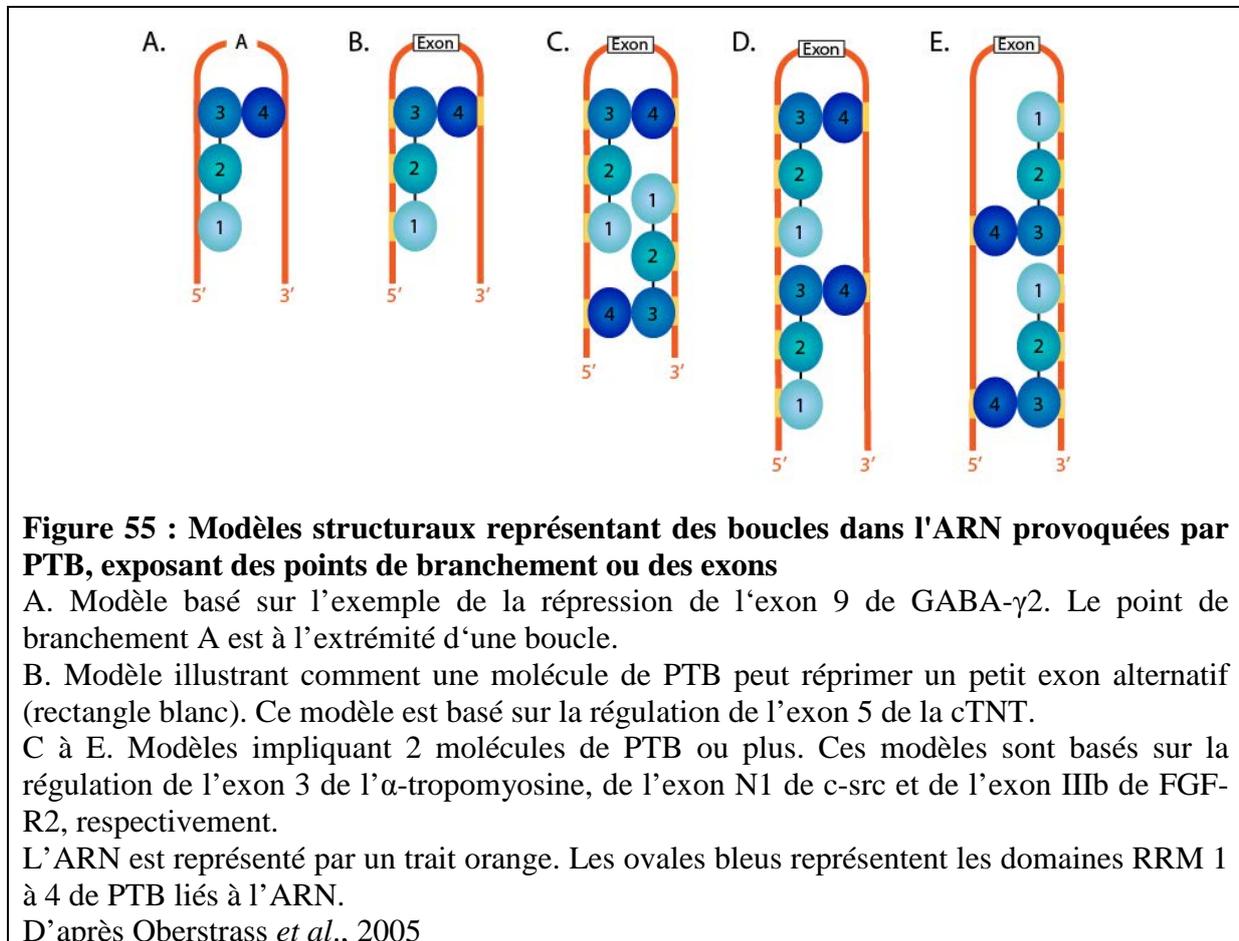


**Figure 54 : Modèle structural de PTB et son interaction avec l'ARN**

Structure et arrangement spatial des quatre domaines RRM de la protéine PTB. La conformation de l'ARN fixé à chaque domaine RRM est indiquée par une ligne bleue pleine. Les lignes en pointillé indiquent la connexion possible entre ces motifs.

D'après Sawicka *et al.*, 2008

Par la fixation de ces multiples domaines de liaison à une même molécule d'ARN, particulièrement à des séquences éloignées dans la structure primaire, PTB peut induire une restructuration de l'ARN et générer de grandes boucles dans l'ARN (**Figure 55**) (Mitchell *et al.*, 2003 ; Oberstrass *et al.*, 2005 ; Petoukhov *et al.*, 2006).



D'autre part, cette protéine, outre son rôle dans l'épissage, est connue pour intervenir dans de nombreux processus cellulaires, dont la stabilité des ARNm (Tillmar *et al.*, 2002 ; Kosinski *et al.*, 2003 ; Coles *et al.*, 2004).

Des expériences de chromatographie d'affinité de RNP avec des extraits nucléaires de cellules ES murines ont révélé la présence de PTB en association avec la région des A-repeats (Maenner *et al.*, 2010). Il serait intéressant de découvrir si elle y joue un rôle dans l'épissage comme ASF/SF2 ou si elle a une autre fonction au niveau de cette région, comme influencer sur sa stabilité via la formation ou la stabilisation de boucles.



---

# OBJECTIFS

---



## Objectifs de la thèse

---

Le processus d'inactivation du chromosome X est constitué d'un ensemble de mécanismes complexes parfois interdépendants. Les régulations sont nombreuses en même temps que la fenêtre de temps dans laquelle il peut se produire est limitée. L'étude de ce phénomène se révèle donc particulièrement difficile. Lorsque je suis arrivée au laboratoire, Sylvain Maenner avait commencé l'étude de la région des A-repeats avec Athanase Visvikis sur la base de la mise en évidence de son rôle crucial dans ce processus. Il avait entamé une analyse structurale de cette région, étant donné que la structure 2D des ARN joue en général un rôle important dans leur mécanisme d'action. Son étude avait abouti à trois modèles structuraux possibles pour la région des A-repeats. Mon premier objectif a été de discriminer parmi ces modèles afin d'identifier celui réellement formé. Pour cela, j'ai réalisé des expériences de FRET appliquées à l'ARN, en collaboration avec S. Maenner et M. Blaud. Les résultats de ces expériences sont présentés dans le premier paragraphe des Résultats. De plus, ces travaux ont fait l'objet en partie d'un article qui est paru dans PLoS Biology et présenté dans le paragraphe II des Résultats.

Par ailleurs, il avait été montré que la région des A-repeats recrute des facteurs importants pour la mise en place de l'inactivation du chromosome X. En parallèle de l'étude structurale, il était donc intéressant de chercher les partenaires protéiques de cette région ainsi que leur rôle dans le processus d'inactivation. L'identification de plusieurs dizaines de protéines par chromatographie d'affinité suivie d'analyse par spectrométrie de masse réalisée par S. Maenner m'a permis de me concentrer sur l'étude de l'interaction des A-repeats avec certaines d'entre elles. Mon deuxième objectif a été de compléter ces analyses en me focalisant plus particulièrement sur la protéine KSRP qui est impliquée aussi bien dans la régulation de la dégradation des ARN qu'à l'épissage différentiel des ARNm.

J'ai ainsi caractérisé l'interaction entre la protéine KSRP et la région des A-repeats par des expériences de retard sur gel, accompagnées d'expériences de protection de l'ARN contre les RNases en présence de la protéine afin d'identifier ses sites de fixation sur l'ARN. Les résultats issus de ces expériences ont révélé des sites de fixation similaires entre la protéine KSRP et une autre protéine identifiée comme partenaire protéique de la région des A-repeats,

la protéine PTB. Puisque les sites de fixation de ces deux protéines sont multiples, elles peuvent interagir avec l'ARN de manière simultanée ou séquentielle. J'ai donc entrepris d'étudier si ces deux protéines étaient en compétition pour se fixer à l'ARN, dans le but de comprendre leur rôle dans le processus d'inactivation. Cette partie de l'étude est présentée dans le paragraphe III des Résultats.

Au cours de ces travaux, une publication a apporté une information qui a retenu notre attention. Il a été montré que la protéine ASF/SF2 interagit également avec la région des A-repeats, ce facteur d'épissage n'ayant pas été identifié lors de nos analyses par spectrométrie de masse. Le troisième objectif de ma thèse a été de caractériser tant les interactions de cette protéine avec la région des A-repeats qu'à déterminer sa fonction.

Etant donné que les A-repeats recrutent effectivement la protéine ASF/SF2 de manière, il était ensuite intéressant d'étudier son rôle dans l'inactivation. L'ARN Xist étant mûré comme un ARN messenger, l'étude s'est naturellement portée sur la fonction d'ASF/SF2 dans son épissage. Pour cela, il a tout d'abord été nécessaire de réaliser des constructions permettant l'examen du profil d'épissage des exons de l'ARN Xist. La région des A-repeats se situant à l'extrémité 5' de l'exon 1 long de 10 kb, il était impossible de construire un plasmide contenant la séquence couvrant à la fois les A-repeats et les différents exons. Nous avons donc dû tronquer certaines régions, tout en essayant de conserver l'ensemble des éléments pouvant réguler l'épissage. Ces différentes constructions nous ont permis d'analyser l'épissage des exons de l'ARN Xist en fonction de la présence ou de l'absence de la région des A-repeats, mais également en fonction de la surproduction de certains facteurs régulateurs de l'épissage. Cette partie de mon travail est présentée dans le paragraphe IV des Résultats.

---

# MATERIEL & METHODES

---



# Matériel et Méthodes

## I. Matériel

### I.1 Souches et plasmides utilisés

#### I.1.1 Les souches d'*E. Coli*

**Souche DH5 $\alpha$**  (*supE44*,  $\Delta$ *lacU169* ( $\Phi$ 80 *lacZ* $\Delta$ M15), *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*) (Hanahan, 1983) ; souche utilisée pour la production et la purification de plasmides.

**Souche BL21 codon plus (DE3)** (*F-*, *ompT*, *hsdS<sub>B</sub>* (*r<sub>B-</sub>*, *m<sub>B-</sub>*), *dcm*, *gal*,  $\lambda$ (DE3), pLysS, Cm<sup>r</sup>) (Studier and Moffatt, 1986) ; souche utilisée pour la production de protéines recombinantes.

#### I.1.2 Les lignées de cellules de mammifères en culture

**Cellules HeLa** : N°ATCC® : CCL-2™ (Scherer, 1953) ; lignée cellulaire cancéreuse provenant d'un prélèvement effectué sur une patiente atteinte d'un cancer du col de l'utérus, et décédée en 1951, **Henrietta Lacks**.

**Cellules 293 (HEK-293)** : N°ATCC® : CRL-1573™ (Graham *et al.*, 1977) ; lignée de cellules embryonnaires humaines de rein.

**Cellules ES XX de type HP3-10** : lignée de cellules ES de souris hétérozygote pour le locus XIC (*Mus musculus/Mis domesticus*, forme allélique 129/Pgk du XIC) (Clerc and Avner, 1998).

#### I.1.3 Les plasmides utilisés

a. Plasmides recombinants utilisés pour la production *in vitro* d'ARN

Les plasmides sont listés dans le **Tableau 1**.

**Tableau 1: Liste de plasmides pour la production *in vitro* d'ARN**

Dénomination du plasmide	Séquence insérée	Position du fragment dans le gène <i>Xist</i>
pUC18-Xist-MS2	région des A-repeats humaine	+332 à +797
pUC18-TBIH-MS2	répétitions 1 à 4 de la région humaine	+332 à +560
pUC18-TBIIH-MS2	répétitions 5 à 9 de la région humaine	+547 à +797
pUC18-1/2H-MS2	répétitions 1 et 2 de la région humaine	+ 332 à +454
pUC18-3/4H-MS2	répétitions 3 et 4 de la région humaine	+435 à +560
pUC18-5'Xist-MS2	région amont de la région humaine	+72 à +331

pUC18-3'Xist-MS2	région aval de la région humaine	+798 à +1056
pUC18-Xim-MS2	région des A-repeats murine	+278 à +760
pUC18-TBIM-MS2	répétitions 1 à 4 de la région murine	+278 à +549
pUC18-TBIIM-MS2	répétitions 5 à 8 de la région murine	+530 à +760
pUC18-1/2M-MS2	répétitions 1 et 2 de la région murine	+278 à +421
pUC18-3/4M-MS2	répétitions 3 et 4 de la région murine	+414 à +549
pUC18-1RM-MS2	répétition 1 de la région murine	+335 à +358
pUC18-5'Xim-MS2	région amont de la région murine	+1 à +300
pUC18-3'Xim-MS2	région aval de la région murine	+761 à +1022

La description de la séquence insérée ainsi que sa position dans le gène *Xist* (humain ou murin selon la séquence) sont indiquées à droite du nom du plasmide.

b. Vecteurs utilisés pour la production de protéines recombinantes

Le plasmide pET28 a été utilisé dans le but de produire les protéines KH1-4, KH1-2, KH3-4 et PTB et permet leur purification grâce à une étiquette « 6xHis » fusionnée à l'extrémité N-terminale (Novagen). La séquence codant la protéine PTB humaine (+23 à +1710) a été clonée dans le pET28b. La séquence codant les 4 domaines KH de la protéine KSRP humaine (+387 à +1703) a également été clonée dans le pET28b.

Le plasmide pMBP-MS2 fourni par le Dr Lurhmann a permis de produire et purifier la protéine MBP/MS2.

c. Plasmides utilisés pour la transfection et l'expression dans des cellules eucaryotes

- Les plasmides pSpliceExpress (fourni par S. Stamm) et pGL3 ont été utilisés pour l'expression d'ARN dans les cellules eucaryotes.

- à partir d'ADN génomique de souris, nous avons amplifié la région +12114 à +14670 du gène *Xist*, correspondant aux exons 2 à 6, en utilisant des amorces (couple 8256/8257) permettant la recombinaison sites-spécifique avec le plasmide pSpliceExpress. Après recombinaison, le plasmide créé a été nommé pSplice-ex2/6.

- à partir d'ADN génomique de souris, nous avons amplifié la région -332 à +278 du gène *Xist*, grâce au couple d'amorces 8065/8066 permettant d'insérer des sites de restriction : *Xma*I à l'extrémité 5' du fragment amplifié et *Avr*II à l'extrémité 3'. Ces sites sont compatibles avec le clonage dans le vecteur pGL3. De même, la région +724 à +1080 a été amplifiée grâce au couple d'amorces 8067/8068, apportant les sites de restriction *Hind*III (à l'extrémité 5' du fragment) et *Nco*I (à l'extrémité 3'), permettant de l'insérer dans le pGL3, en

aval de la région précédente. Le plasmide ainsi généré a alors été appelé pGL3- $\Delta$ Arep-luci. Différents fragments de la région des A-repeats ont ensuite été amplifiés : (i) la région entière (couple 8199/8200) ; (ii) les répétitions 1 à 4 (couple 8199/8201) et (iii) les répétitions 1 et 2 (couple 8199/8202). L'amorce 8199 comporte un site *AvrII* en 5' tandis que les 3 autres possèdent en 5' un site *HindIII*, ce qui a permis le clonage de ces séquences entre les mêmes sites précédemment créés dans le plasmide pGL3- $\Delta$ Arep-luci, générant les plasmides pGL3-Arep-luci, pGL3-4rep-luci et pGL3-2rep-luci.

Ces plasmides ont ensuite été utilisés pour 2 stratégies de clonage différentes :

- la séquence codant l'ARN C2 d'HIV-1 a été clonée à partir du plasmide pGL3-C2 dans chacun de ces plasmides en utilisant les sites *BamHI* et *XbaI*, remplaçant ainsi les séquences codant la luciférase et générant les plasmides : pGL3- $\Delta$ Arep-C2, pGL3-Arep-C2, pGL3-4rep-C2 et pGL3-2rep-C2.

- la région +9201 à +9720 a été amplifiée grâce au couple d'amorces 8069/8219 apportant les sites de restriction *NcoI* en 5' et *EcoRI* et *BclI* en 3', permettant de l'insérer dans les plasmides précédents en utilisant les sites *NcoI* et *BclI*. Ensuite, la région des positions +12114 à +14670 a été amplifiée en utilisant le couple d'amorces 8071/8072 apportant le site *EcoRI* en 5' du fragment amplifié et le site *XbaI* en 3', ce qui a permis de l'insérer en aval de la région précédemment clonée. Les plasmides ainsi générés sont : pGL3- $\Delta$ Arep-ex1/6, pGL3-Arep-ex1/6, pGL3-4rep-ex1/6 et pGL3-2rep-ex1/6.

- Le plasmide pRGB (Promega) porte la séquence codant le gène de la luciférase Renilla de reniformis. Il a été utilisé comme vecteur de contrôle de transfection.
- Le plasmide pXJ41 a été utilisé pour la surexpression des protéines ASF/SF2 (fourni par J. Stevenin), 9G8 et PTB.
- La séquence codant les 4 domaines KH de la protéine KSRP a été clonée dans le plasmide pcDNA3 portant une étiquette Flag. Elle a été amplifiée avec les amorces 7434 et 4670, permettant un clonage entre les sites *BamHI* et *XhoI*, et générant le plasmide pcDNA3-Flag-KH1-4.

Les séquences des oligonucléotides utilisés pour réaliser ces constructions sont indiquées dans le **Tableau 2**.

**Tableau 2 : Listes des oligonucléotides utilisés pour la construction des plasmides d'expression dans des cellules eucaryotes**

Constructions	Séquence des oligonucléotides
pSplice-ex2/6	8256 – GGGGACAAGTTTCTACAAAAAAGCAGGCTGGCCTTTGAAATGTAGACATGC 8257 – GGGGACCACTTTGTACAAGAAAGCTGGGTAAGAATGGTAGGGCACCAAG
pGL3-ΔArep-luci	8065 – CCCGGGCGGAGCGATACATGGTTTGT 8066 – CCTAGGCCAAAAGGCACAACAAAGAAATATTCC 8067 – AAGCTTGCTTGGTGGATGGAAATATGG 8068 – CCATGGCAACCCCGCAAATGCTACCA
pGL3-Arep-luci pGL3-4rep-luci pGL3-2rep-luci	8199 – AACCTAGGGTTGATACTTGTGTGTGTATGG 8200 – CCAAGCTTGCCCCGTTGGGCCGTGAA 8201 – CCAAGCTTTTTTTTTTTTTTTTTTTTACAAAAAGCAGG 8202 – CCAAGCTTAAGAATGAAAAGGCAGGTAAGTATCC
pGL3-ΔArep-ex1/6 pGL3-Arep-ex1/6 pGL3-4rep-ex1/6 pGL3-2rep-ex1/6	8069 – CCATGGGCGCAGTCATTCATTCACATGG 8219 – GGTGATCATTGAATTCCTCATTTTAGTGTACCCAGC 8071 – GAATTCGGCCTTTGAAATGTAGACATGC 8072 – TCTAGAAAGAATGGTAGGGCACCAAG
pcDNA3-Flag-KH1-4	7434 – AAAGGATCCAATCAGTTCTCAACTTGGACCCAT 4670 – CCGCTCGAGCGGGCTGCAGCTGCTTTGCTTGG

La séquence des oligonucléotides est indiquée dans le sens 5' vers 3'.

## I.2 Milieux de culture, tampons et solutions d'usage courant

**Milieu LB (Luria Bertani)** : peptone pancréatique animale 5 g/L ; peptone tryptique de caséine 5 g/L ; extrait de levure 5 g/L ; NaCl 10 g/L.

Le pH est ajusté à 7,5 avec NaOH 10 N. Le milieu de culture est ensuite autoclavé 20 min à 121°C sous une pression de 1 bar.

**Milieu solide LB** : l'agar est ajouté à raison de 17 g par litre de milieu avant stérilisation à l'autoclave.

**Milieu LBA** : l'ampicilline (100 mg/mL) est ajoutée après autoclave dans le milieu de culture LB à une concentration finale de 100 µg/mL.

**Milieu LBK** : la kanamycine (100 mg/mL) est ajoutée après autoclave dans le milieu de culture LB à une concentration finale de 100 µg/mL.

**TE** : Tris-HCl 10 mM, pH 7,4 ; EDTA 1 mM.

**TBE** : Tris borate 89 mM, pH 8,3 ; EDTA 2 mM.

**TBM** : Tris borate 89 mM, pH 8,3 ; MgCl<sub>2</sub> 2 mM.

**PBS 10X** : NaCl 1,37 M ; KCl 27 mM ; KH<sub>2</sub>PO<sub>4</sub> 11,5 mM ; Na<sub>2</sub>HPO<sub>4</sub> 64 mM.

**Solution de transfert** : Tris 50 mM ; glycine 14,5 mM ; éthanol 20 % (v/v).

**Solution de polyacrylamide dénaturante** : Acrylamide/Bisacrylamide (23,75/1,25) 25 % (p/v) ; TBE 1X ; urée 8M.

**Solution de polyacrylamide native** : Acrylamide/Bisacrylamide (38/2) 40 % (p/v) ; TBE 1X.

**Solution de dépôt agarose (6X)** : EDTA 125 mM ; SDS 2 % (v/v) ; glycérol 25 % (v/v) ; bleu de bromophénol 0,12 % (p/v).

**Solution de bleu de formamide** : EDTA 20 mM ; bleu de bromophénol 0,05 % (p/v) ; bleu de xylène cyanol 0,05 % (p/v) dans de la formamide ionisée.

**Solution de dépôt non dénaturant pour les expériences de retard sur gel (bleu CSB)** : HEPES-KOH 20 mM, pH 7,9 ; glycérol 40 % (v/v) ; bleu de bromophénol 0,05 % (p/v) ; bleu de xylène cyanol 0,03 % (p/v).

**Tampon d'éluion** : Tris-HCl 10 mM, pH 7,5 ; NaCl 300 mM ; EDTA 1 mM ; SDS 1 %.

**Tampon D** : HEPES-KOH 20 mM, pH 7,9 ; MgCl<sub>2</sub> 1,5 mM ; EDTA 0,2 mM ; KCl 150 mM ; glycérol 10 % (p/v).

**Tampon SSC 20X** : NaCl 3 M ; citrate de sodium 0,3 M.

### I.3 Anticorps

Tous les anticorps utilisés au cours de ce travail sont répertoriés dans le **Tableau 3**.

**Tableau 3 : Liste des anticorps utilisés au cours de ce travail**

Dénomination de l'anticorps	Clonalité	Utilisation	Origine
anti-PTB	polyclonal de mouton	Western-blot (3000°)	Abcam (ab5642)
anti-KSRP	polyclonal de souris	Western-blot (1000°)	D. Black, University of California
anti-Sin3A	polyclonal de lapin	Western-blot (500°)	Santa Cruz (sc-994)
anti-Rif1	polyclonal de lapin	Western-blot (500°)	Abnova (PAB10275)
anti-Pds5a	polyclonal de lapin	Western-blot (750°)	Abnova (H00023244-B01)
anti-Smc1	polyclonal de lapin	Western-blot (1000°)	Bethyl Laboratories (A300-055A)
anti-Smc3	polyclonal de lapin	Western-blot (1000°)	Bethyl Laboratories (A300-060A)
anti-ASF/SF2	monoclonal de souris	Western-blot (2500°)	J. Stevenin, IGBMC
anti-9G8	monoclonal de souris	Western-blot (3000°)	Euromedex (9G8-1C6)
anti-souris IgG (H+L)	mouton	Western-blot (20000°)	Jackson Immunoresearch (515035003)
anti-lapin IgG (H+L)	mouton	Western-blot (20000°)	Pierce (31460)
anti-mouton (H+L)	singe	Western-blot (20000°)	Santa-Cruz (sc-2020)

L'hôte ayant servi à la production, la dilution utilisée pour les Western-blots et le fournisseur sont indiqués pour chaque anticorps.

IgG : immunoglobuline G ; H : chaîne lourde (heavy) ; L : chaîne légère (light).

## II. Méthodes

### II.1 Méthode utilisée pour l'analyse de la structure secondaire d'un ARN par la technique de FRET

#### II.1.1 Préparation des ARN

Dans le cadre de la production de transcrits non-radioactifs, le mélange réactionnel suivant est réalisé :

- Plasmide linéarisé	500 ng
- Tampon 5X (Tris-HCl 200 mM, pH 7,5 ; MgCl <sub>2</sub> 30 mM ; spermidine 10 mM ; NaCl 50 mM)	6 µL
- rNTP (UTP/ATP/GTP/CTP 25 mM)	5 µL
- ARN T7 polymérase	3 µL
- H <sub>2</sub> O	qsp 30 µL

Ce mélange est incubé pendant 3h à 37°C. La matrice ADN est ensuite dégradée par ajout de 1µL de DNase RQ1 (1 UE/µL, Promega) et le mélange est incubé 30 min à 37°C. Les ARN sont ensuite purifiés par électrophorèse sur gel de polyacrylamide dénaturant 5 % à 10 % selon la taille de l'ARN produit. Après migration, les bandes du gel contenant l'ARN sont détectées par « UV-shadowing » ( $\lambda = 254$  nm). Ces bandes sont ensuite découpées et placées dans 400 µL de tampon d'éluion. Les ARN sont alors élués à 4°C pendant 15h sous agitation constante puis précipités à l'éthanol 96 % en présence d'acétate de sodium 0,3 M final, lavés à l'éthanol 70 %, séchés, repris dans de l'eau stérile et quantifiés.

#### II.1.2 Technique de FRET permettant d'apprécier la distance relative entre 2 zones ARN d'intérêt

Le mélange suivant est réalisé :

- ARN	9,3 µg (0,38 µM final)
- O-Cy3	0,38 µM
- tampon SSC 20x	8 µL (1x final)
- MgCl <sub>2</sub>	3,125 mM
- H <sub>2</sub> O	qsp 160 µL

Ce mélange est incubé 5 min à 65°C pour dénaturer l'ARN puis 10 min à température ambiante. Puis il est incubé 4 heures à 4°C pour l'hybridation de l'oligonucléotide à l'ARN.

Le rendement de l'hybridation est déterminé par électrophorèse en gel acrylamide non dénaturant 7 %. Le gel est scanné un appareil Typhoon 9410 GE Healthcare et la fluorescence est quantifiée à l'aide du logiciel ImageQuant.

Un volume de 150 µL du mélange est transféré dans une cuve de contenance 160 µL placée dans un spectrofluoromètre flx (SAFAS) maintenu à une température de 4°C. L'appareil est paramétré sur 750 V pour le photomultiplicateur et sur 3 nm pour les bandes passantes

d'excitation et d'émission. Les mesures de fluorescence sont réalisées à une longueur d'onde d'excitation de 515 nm et les spectres d'émission sont enregistrés entre 500 et 750 nm. La mesure comprend 10 spectres successifs dont la moyenne est ensuite calculée.

L'oligonucléotide O-Cy5 est alors ajouté à une concentration finale de 0,38  $\mu\text{M}$  et le mélange est incubé 4 heures à 4°C. Dix spectres sont à nouveau enregistrés et leur moyenne est calculée. Le FRET est alors calculé par comparaison des 2 moyennes de spectres, en tenant compte du ratio d'oligonucléotides libres et liés à l'ARN.

Voici le calcul réalisé :

- soit  $P_C$  la proportion d'oligonucléotide complexé à l'ARN et  $P_L$  la proportion d'oligonucléotide libre. L'oligonucléotide n'existant pas sous une autre forme, nous avons  $P_C + P_L = 100 \%$ .

- soit  $I_i$  l'intensité de fluorescence du groupement Cy3 avant l'ajout de O-Cy5 et. Elle peut être exprimée selon ses 2 composantes,  $I_L^i$  (représentant l'intensité de fluorescence issue du O-Cy3 libre avant ajout de O-Cy5) et  $I_C^i$  (l'intensité de fluorescence issue du O-Cy3 complexé à l'ARN avant ajout de O-Cy5).  $I_i$  est fixée comme référence à 100 %. Comme l'intensité de fluorescence d'une forme est directement liée à sa proportion, alors en pourcentage nous avons :  $I_L^i = P_L$  et  $I_C^i = P_C$ .

- soit  $I_f$  l'intensité de fluorescence de Cy3 après hybridation de O-Cy5. Elle est également composée de 2 variables,  $I_L^f$  (l'intensité de fluorescence issue du O-Cy3 libre après ajout de O-Cy5) et  $I_C^f$  (l'intensité de fluorescence issue du O-Cy3 complexé à l'ARN après ajout de O-Cy5). Or l'intensité de fluorescence issue du O-Cy3 libre ne varie pas lors de l'ajout de O-Cy5. Donc  $I_L^i = I_L^f = I_L = P_L$  et par conséquent  $I_f = P_L + I_C^f$ . Donc  $I_C^f = I_f - P_L$ .

- dans notre cas, l'efficacité de transfert E représente la variation d'intensité de fluorescence du Cy3, soit  $I_C^i - I_C^f$ . Afin de l'exprimer en pourcentage, voici la formule appliquée :

$$E = 100 - \frac{100 \times (I_f - P_L)}{P_C}$$

Les oligonucléotides qui ont servis à cette technique sont présentés dans le **Tableau 4**.

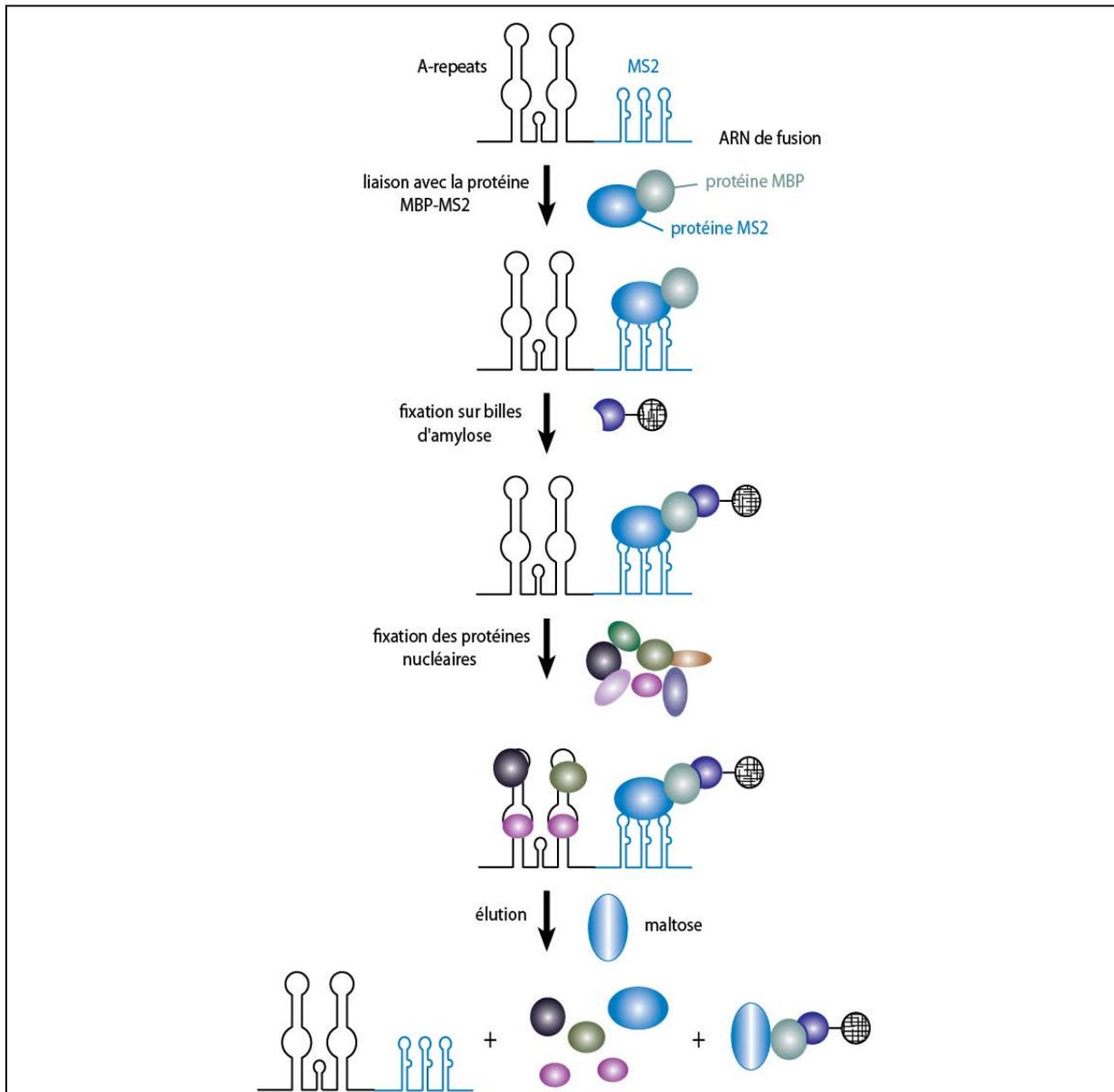
**Tableau 4 : Liste des oligonucléotides utilisés pour les expériences de FRET**

Dénomination	Séquence des oligonucléotides
P1	GAATATATAAACAATGAAAGAAAGG-Cy3
P2	Cy3-GCGGGGAGGACACACAGGTA
P3	Cy3-AATGAAAAGGCAGGTAAGTA
P4	GAAGAAAAAAGAATAAAAG-Cy5
P5	Cy5-ACAAAAGCAGGTATCCATG
P6	GCAATAACTCACAAAACCAT-Cy3
P7	Cy5-GGTTTTTTTTTTTTTTTTTT
P8	Cy5-AATGAAAAGGCAGGTAAGTA

La séquence de chaque oligonucléotide est indiquée de l'extrémité 5' vers l'extrémité 3'.

## II.2 Purification des complexes RNP par chromatographie d'affinité MS2/MBP

Les ARN utilisés pour cette méthodologie contiennent trois sites de fixation de la protéine MS2 du phage R17 (phage R17-MS2 coat protein). La protéine de fusion MS2/MBP est composée de la protéine MBP (Maltose Binding Protein) et de la protéine MS2. L'ARN (100 pmol) est repris dans 50  $\mu$ L de tampon D, dénaturé pendant 10 min à 65°C, puis renaturé 10 min à température ambiante et placé dans la glace. La protéine MS2/MBP (500 pmol) est ensuite ajoutée et le volume final est ajusté à 100  $\mu$ L avant d'être incubé 20 min dans la glace. Le complexe formé [ARN-MS2/MBP] est ensuite fixé sur 40  $\mu$ L de billes d'amylose pendant 1h à 4°C sous agitation constante. Les billes sont isolées par centrifugation 1 min à 11000 rpm, puis lavées à 3 reprises avec 500  $\mu$ L de tampon D. Les complexes formés [billes-ARN MS2/MBP] sont incubés 15 min à 4°C sous agitation constante, en présence d'1 mg d'extrait nucléaire et 5  $\mu$ M d'ARNt de levure. Les billes sont à nouveau isolées par centrifugation d'1 min à 11000 rpm, puis lavées 3 fois avec 500  $\mu$ L de tampon D pendant 10 min. Les complexes RNP sont élués avec 75  $\mu$ L de tampon D complété en maltose à raison de 10 mM final lors d'une incubation de 30 min sous agitation constante à température ambiante. Puis, 20  $\mu$ L de bleu SBL (Tris 250 mM, pH 6,8 ; SDS 8 % ;  $\beta$ -mercaptoéthanol 20 % ; bleu de bromophénol 0,04 %) sont ajoutés à 50 % de l'éluat et les protéines sont fractionnées sur gel SDS-PAGE 10 % Acrylamide-Bisacrylamide (29:1). Après migration, le gel est coloré au bleu colloïdale (sulfate ammonium 10 % ; bleu de Coomassie G250 0,1 % ; ac. Orthophosphorique 1,6 % ; EtOH 20 %). Dix pourcent des éluats sont analysés par Western-blot (**Figure 56**).



**Figure 56 : Stratégie de purification de complexes formés entre la région des A-repeats et les protéines d'un extrait nucléaire de cellules ES murines**

L'ARN d'intérêt fusionné co-transcriptionnellement à 3 motifs MS2 est incubé en présence de la protéine de fusion MBP/MS2. Ce complexe est ensuite mis en contact avec des billes d'amylose. Le complexe ternaire ainsi formé est ensuite incubé en présence d'un extrait nucléaire. Après plusieurs lavages successifs, l'ajout de maltose permet l'élution des complexes formés. Une partie de l'éluat est déposée sur gel SDS-PAGE afin de fractionner les protéines contenues dans ces complexes. Après coloration au bleu de Coomassie, les bandes du gel sont découpées et analysées par spectrométrie de masse. Le reste de l'éluat est analysé par Western-Blot.

## II.3 Production et purification de la protéine KH1-4-6His chez *E. Coli*

### II.3.1 Transformation et culture bactérienne

Des bactéries *E. coli* BL21(DE3) sont transformées avec 10 ng de plasmide pET28b-KH1-4 portant la séquence codant les 4 domaines KH de la protéine KSRP. Les bactéries sont cultivées dans le milieu LB en présence d'un antibiotique de sélection. La production de la protéine est induite par l'ajout d'IPTG (1 mM final) pendant 4 h à 37°C sous agitation.

### II.3.2 Préparation des extraits cellulaires

La culture est centrifugée pendant 20 min à 4000 rpm à 4°C et le culot est repris dans 10 mL de tampon de sonication (Tris-HCl 20 mM, pH 7,8 ; NaCl 200 mM) par litre de culture. Les cellules subissent ensuite une étape de sonication (2 fois 2,5 min ; Duty Cycle 50 % ; Output 5, entrecoupé d'arrêt de 1 min dans la glace). Le sonicat est centrifugé à 4°C pendant 40 min à 18000 rpm. Le surnageant est ensuite complété avec 20 mM d'imidazole et d'un mélange d'inhibiteurs de protéases 1X puis il est filtré à 0,45 µM.

### II.3.3 Purification de la protéine KH1-4-6His

Sont ensuite ajoutés 1,5 mL de billes agarose Ni-NTA (Qiagen) et l'ensemble est incubé pendant 1 h à 4°C sur roue. Après décantation des billes et un lavage avec 10 mL de tampon de sonication contenant 100 mM d'imidazole, les protéines retenues sur le gel sont éluées avec 10 mL de tampon de sonication contenant 500 mM d'imidazole, récupérées par fractions de 500 µL, et analysées sur gel SDS-PAGE à 10 %. Les fractions sont ensuite concentrées dans un volume de 500 µL selon le kit « Amicon ultra » et dialysées contre du tampon D.

### II.3.4 Estimation de la quantité et de la qualité de l'échantillon protéique

La concentration finale en protéine est évaluée par la méthode de dosage Quick Start™ Bradford (BioRad). La qualité de l'échantillon est vérifiée d'une part, par dépôt sur gel SDS-PAGE révélé par coloration au bleu de Coomassie (bleu de Coomassie 0,25 %, éthanol 50 % et acide acétique 10 %) ; et d'autre part par dichroïsme circulaire. Les aliquots protéiques sont conservés à -80°C.

## II.4 Analyses des interactions ARN/protéines in vitro

### II.4.1 Analyses des complexes ARN/Protéines par expérience de retard sur gel

#### a. Préparation des ARN

Le mélange réactionnel pour préparer les ARN uniformément radiomarqués est le suivant :

- Plasmide linéarisé (500 ng/μL)	1 μL
- Tampon 5X (MgCl <sub>2</sub> 30 mM ; Spermidine 10 mM ; NaCl 50 mM ; Tris-HCl 200 mM, pH 7,5)	6 μL
- rNTP (25 mM ATP/CTP/GTP + 8 mM UTP)	5 μL
- [ $\alpha$ - <sup>32</sup> P]-UTP (800 Ci/mmol, 10 mCi/mL, Perkin Elmer)	1 μL
- ARN polymérase T7	3 μL
- H <sub>2</sub> O	qsp 30 μL

Le mélange réactionnel est incubé pendant 3 h à 37°C. La matrice ADN est ensuite dégradée par ajout de 1 μL de DNase RQ1 (1 UE/μL, Promega) et incubation 30 min à 37°C. Les ARN sont ensuite purifiés par électrophorèse sur gel de polyacrylamide dénaturant. Après migration, les bandes du gel contenant l'ARN sont détectées par fluorographie. Ces bandes ont ensuite découpées et placées dans 400 μL de tampon d'éluion. Les ARN sont alors élués 4°C pendant 15 heures sous agitation constante puis précipités à l'éthanol à 96 % en présence d'acétate de sodium 0,3 M final, lavés à l'éthanol 70 %, séchés et repris dans de l'eau stérile.

#### b. Formation des complexes ARN/protéines

Pour la formation des complexes ARN/protéines, les mélanges réactionnels suivant sont réalisés :

- Transcrit non radiomarqué (1,2 pmol/μL)	1 μL
- Transcrit radiomarqué (2500 cpm/μL)	1 μL
- ARNt de levure (5 μg/μL)	1 μL
- Tampon D	3,6 μL
- H <sub>2</sub> O	qsp 13 μL

Ce mélange est incubé 10 min à 65°C pour la dénaturation de l'ARN

- MgCl <sub>2</sub> (62,5 mM)	1 μL
-------------------------------	------

Une incubation 10 min à température ambiante permet de renaturer l'ARN.

Le mélange est ensuite incubé en présence de protéines recombinantes en quantité variable (mais toujours ajusté à un volume réactionnel de 20 μL) pendant 30 min à 4°C afin de permettre la formation des complexes ARN/protéines. Après avoir ajouté 3 μL de bleu de dépôt CSB, l'échantillon est ensuite déposé sur un gel de polyacrylamide natif 5 %

(Acrylamide/Bisacrylamide 38/2), en présence de glycérol 5 % et de TBE 0,5X. L'électrophorèse est réalisée à 100 V/cm et à 4°C. Le résultat de la réaction de retard sur gel est visualisé après une nuit de fluorographie.

#### II.4.2 Analyses des complexes ARN/Protéines par expérience d'empreinte aux RNases

La RNase T1 hydrolyse spécifiquement les ARN en 3' des résidus guanines, préférentiellement dans les segments non appariés, lorsqu'elle est utilisée en conditions limitantes. La RNase T2 hydrolyse préférentiellement les régions en simple-brin, lorsqu'elle est utilisée en conditions limitantes. Cette enzyme agit sans spécificité vis-à-vis des nucléotides et, comme la RNase T1, libère des fragments avec des extrémités 3'-phosphate (3'-P) et 5'-hydroxyle (5'-OH). La RNase V1 hydrolyse préférentiellement les régions en double-brin et les régions structurées par empilement de bases, ceci avec peu de spécificité vis-à-vis de l'identité des résidus. Elle libère des fragments avec des extrémités 3'-OH et 5'-P.

##### a. Préparation des ARN

La préparation des ARN et la formation des complexes ARN/protéines suivent le protocole présenté dans le paragraphe II.4.1. Le mélange est ensuite incubé avec l'enzyme (T1 : 0,02 ou 0,0375 U/ $\mu$ L ; T2 : 0,025 ou 0,0375 U/ $\mu$ L et V1 :  $2.5 \times 10^{-3}$  ou  $5 \times 10^{-3}$  U/ $\mu$ L) dans un volume réactionnel ajusté à 20  $\mu$ L avec du tampon D. Les réactions sont arrêtées après 10 min d'incubation à 30°C. Un témoin sans enzyme est également réalisé. L'arrêt des réactions par ajout de 20  $\mu$ g d'ARNt de levure est suivi par une extraction au phénol/chloroforme et par une précipitation à l'éthanol 96 % en présence de glycogène (10  $\mu$ g) et de 0,3 M d'acétate de sodium. Le culot d'ARN est alors lavé avec 200  $\mu$ L d'éthanol à 70 %, séché, puis resuspendu dans 7  $\mu$ L d'eau. L'analyse des coupures enzymatiques se fait par extension d'amorce avec un oligonucléotide radiomarké sur 1  $\mu$ L d'ARN clivés, ce qui permet une révélation par fluorographie après électrophorèse sur gel dénaturant.

##### b. Analyse par extension d'amorce

###### • Marquage de la sonde

- ATP [ $\gamma^{32}$ P] (Perkin Elmer, 5 mCi/mL, 3000 Ci/mmol)	1 $\mu$ L
- Oligonucléotide à marquer (15 pmol/ $\mu$ L)	1 $\mu$ L
- Tampon de phosphorylation 10X (Tris/HCl 0,5 M, pH 7,6 ; MgCl <sub>2</sub> 100 mM ; DTT 50 mM ; Spermidine 1 mM ; EDTA 1mM )	1 $\mu$ L
- T4 polynucléotide kinase (MBI Fermentas 10 U/ $\mu$ L)	1 $\mu$ L

- H<sub>2</sub>O stérile qsp 10 µL

Le mélange réactionnel est incubé 45 min à 37°C. Les amorces marquées sont purifiées par électrophorèse en gel de polyacrylamide dénaturant à 10 %, éluées une nuit à 37°C dans 400 µL de tampon d'éluion sous agitation constante et précipitées comme décrit précédemment. Les amorces sont ensuite reprises dans de l'eau stérile à 100 cps/µL.

• Extension d'amorce

Hybridation de l'amorce

- Oligonucléotide marqué (100 cps/µL) 1 µL
- Transcrit hydrolysé 1 µL
- Tampon de l'enzyme AMV 10X 0,25 µL  
(Tris-HCl 50 mM pH, 8,3 ; KCl 40 mM ; MgCl<sub>2</sub> 6 mM)
- H<sub>2</sub>O stérile qsp 2,5 µL

Ce mélange est incubé 10 min à 65°C, puis 10 min dans la glace.

• Elongation et séparation

- dNTP (1,25 mM UTP/ATP/GTP/CTP) 0,1 µL
- Tampon de l'enzyme AMV 10X 0,25 µL
- AMV Réverse transcriptase (Quantum Appligene, 2 U/µL) 0,25 µL
- H<sub>2</sub>O stérile qsp 2,5 µL

Pour réaliser une extension d'amorce, 2,5 µL de ce dernier mélange sont ajoutés aux 2,5 µL du mélange d'hybridation. L'ensemble est incubé 45 min à 42°C, plongé dans la glace et la réaction d'extension est arrêtée par l'ajout de 4 µL de bleu de dépôt dénaturant. Dans le cas d'un séquençage, 0,5 µL de l'un des quatre didésoxynucléotides à 0,5 mM sont ajoutés au mélange de la réaction d'extension. Les échantillons sont dénaturés 2 min à 96°C, refroidis dans la glace, puis analysés par électrophorèse en gel de polyacrylamide/urée dénaturant à 7%. Après migration et séchage, le gel est placé en fluorographie une nuit à -80°C.

### II.4.3 Analyses des complexes ARN/Protéines par expérience de pontage covalent aux UV

#### a. Préparation des ARN

Le mélange réactionnel pour préparer les ARN uniformément radiomarqués est le suivant :

- Plasmide linéarisé (500 ng/µL) 1 µL
- Tampon 5X 6 µL  
(MgCl<sub>2</sub> 30 mM ; Spermidine 10 mM ; NaCl 50 mM ; Tris-HCl 200 mM, pH 7,5)
- rNTP (25 mM ATP/CTP/GTP + 8 mM UTP) 5 µL
- [α-<sup>32</sup>P]-UTP (800 Ci/mmol, 10 mCi/mL, Perkin Elmer) 1 µL
- ARN polymérase T7 3 µL
- H<sub>2</sub>O qsp 30 µL

Le mélange réactionnel est incubé pendant 3 h à 37°C. La matrice ADN est ensuite dégradée par ajout de 1 µL de DNase RQ1 (1 UE/µL, Promega) et incubation 30 min à 37°C. Les ARN sont ensuite purifiés par électrophorèse sur gel de polyacrylamide dénaturant. Après migration, les bandes du gel contenant l'ARN sont détectées par fluorographie. Ces bandes sont ensuite découpées et placées dans 400 µL de tampon d'éluion. Les ARN sont alors élués à 4°C pendant 15 heures sous agitation constante puis précipités à l'éthanol à 96% en présence d'acétate de sodium 0,3 M final, lavés à l'éthanol 70 %, séchés et repris dans de l'eau stérile.

b. Préparation des puits

Une plaque de 96 puits est utilisée. Les puits sont saturés avec une solution contenant 20 µg de BSA dans 25 µL de tampon D. Une incubation de 30 min à 4°C est nécessaire, puis les puits sont lavés à 2 reprises en tampon D. La plaque est ensuite placée sur un lit de glace.

c. Préparation du mélange réactionnel

Le mélange réactionnel est le suivant :

- Transcrit uniformément radiomarqué	500000 cpm
- ARNt de levure (20 µg/µL)	1 µL
- Tampon D	2 µL
- H <sub>2</sub> O	qsp 9 µL

Il est incubé 5 min à 65°C pour dénaturer l'ARN.

- MgCl <sub>2</sub> (62,5 mM)	1 µL
-------------------------------	------

Le mélange est alors incubé 5 min à température ambiante pour renaturer l'ARN puis 5 min à 4°C.

- Extrait nucléaire	30 µg
- Inhibiteur de RNases (RNase Inhibitor, 40 U/µL, Fermentas)	1 µL
- Tampon D	qsp 30 µL

L'incubation a lieu pendant 30 min à température ambiante sous agitation constante.

d. Pontage covalent aux UV entre les protéines et les ARN

Chaque échantillon est transféré dans un puits de la plaque placée ensuite sous une source UV ( $\lambda = 254$  nm ; Biolink) à 3000 Joules/cm<sup>2</sup> pendant 10 min. L'échantillon est transféré à nouveau dans un tube Eppendorff dans lequel sont ajoutés 15 µg de RNase A et incubé 45 min à 50°C. De cette solution est prélevé 1 µL auquel sont ajoutés 10 µL d'eau et

3  $\mu$ L de bleu SBL 4X, et l'échantillon est ensuite fractionné par électrophorèse en gel SDS-PAGE.

## **II.5 Méthodes utilisées pour la culture cellulaire**

### II.5.1 Culture cellulaire

#### a. Lignées HEK293 et HeLa

Les lignées cellulaires HEK293 et HeLa sont cultivées dans le milieu DMEM (Dulbecco's Modified Eagle Medium) supplémenté avec 10 % de sérum de veau fœtal (FBS), 10 U/mL de pénicilline/streptomycine et 2,9 mg/mL de glutamine (Gibco). Les cellules sont cultivées en flasques ou en boîtes de différents diamètres (en fonction de l'utilisation) à 37°C, sous une atmosphère humide et à 5 % de CO<sub>2</sub>.

#### b. Lignée cellulaire ES XX de type HP3-10

Cette lignée est cultivée dans le milieu DMEM glutamax supplémenté avec 20 % de sérum de veau fœtal (Invitrogen), 10 U/mL de pénicilline/streptomycine et 2,9 mg/mL de glutamine (Gibco) et en présence de 0,5 M de 2- $\beta$ -mercapto-éthanol. Les cellules sont cultivées en flasques ou en boîtes dont la surface de culture a été gélatinisée. Cette étape consiste en une incubation de 20 min à température ambiante des supports plastiques avec une solution de gélatine 0,1 % autoclavée. Les boîtes et les flasques sont ensuite rincées au PBS 1X. Les cellules peuvent y être placées et mises en culture dans un incubateur à 37°C, sous une atmosphère humide et à 8 % de CO<sub>2</sub>. Pour que ces cellules restent indifférenciées, on ajoute au milieu du LIF (Invitrogen) à raison de 1000 U/mL. Lorsque les cellules sont destinées à se différencier, un milieu de culture sans LIF remplace le précédent lors de leur passage.

### II.5.2 Entretien des cellules

L'entretien et l'amplification des cellules s'effectuent par culture dans des boîtes de 10 cm<sup>2</sup>. La confluence des cellules est vérifiée par microscopie. Le milieu de culture est retiré et 1 mL d'une solution de trypsine (0,5 g/L) / EDTA (0,2 g/L) est ajouté. Les cellules se décollent après une incubation de 5 min à 37°C. Puis, 8 mL de milieu sont ajoutés, le mélange est récupéré et les cellules sont centrifugées 2 min à 1000 g. Elles sont ensuite resuspendues dans 10 mL de milieu etensemencées dans une nouvelle boîte de culture avec 1 mL de cette suspension et 9 mL de milieu. Elles sont mises en culture selon les conditions présentées ci-avant jusqu'à confluence (2 à 3 jours).

### II.5.3 Préparation d'extrait nucléaire (adapté d'après Dignam, 1983)

Toutes les étapes doivent se dérouler à 4°C. Après lavages au PBS 1X, les cellules sont grattées, récoltées puis centrifugées pendant 10 min à 1850 g. Après élimination du surnageant, le volume cellulaire (packed cell volume ou pcv) est mesuré. Le culot cellulaire est ensuite suspendu dans 5 pcv de tampon hypotonique puis à nouveau centrifugé pendant 5 min à 1850 g ce qui permet de laver les cellules. Le surnageant est éliminé et le culot est resuspendu dans un volume final de 3 pcv de tampon hypotonique. Le tampon hypotonique ainsi que l'incubation des cellules pendant 10 min dans la glace permet d'augmenter le volume des cellules, ce qui les rend plus sensibles à la casse mécanique. Les cellules sont ensuite transférées dans un Dounce et 18 mouvements de va-et-vient avec un piston A permettent de rompre les cellules sans endommager les noyaux cellulaires. Le lysat est transféré dans un tube à centrifugation et les noyaux sont collectés après une centrifugation de 15 min à 3300 g. Le volume nucléaire (packed nuclear volume ou pnv) est mesuré. Les noyaux sont suspendus dans un volume de tampon de basse concentration saline égal à ½ du pnv. Goutte à goutte, ½ du pnv en tampon de haute concentration saline est ajouté afin de détruire les membranes nucléaires. L'ensemble est homogénéisé avec le dounce. L'échantillon est incubé pendant 30 min à 4°C sous agitation constante puis centrifugé pendant 30 min à 16000 g. Le surnageant est récupéré puis dialysé pendant 1h contre 50 volumes de tampon de dialyse. Le dialysat est centrifugé pendant 20 min à 16000 g. Les protéines contenues dans l'extrait nucléaire recueilli sont dosées, puis l'échantillon est aliquoté et conservé à -80°C.

### II.5.4 Analyse d'épissage *in cellulo*

#### a. Transfection

Des cellules HEK293 sont comptées grâce à une cellule de Mallassez puis ensemencées à 200 000 cellules par puits dans des plaques de 6 puits, 24 heures avant la transfection. Les cellules sont alors transfectées avec 1 µg de mini-gène (pGL3c et pGL3-Arep-ex1/6, pGL3-4rep-ex1/6, pGL3-2rep-ex1/6 et pGL3-ΔArep-ex1/6) par puits en suivant les instructions du fournisseur du kit JetPEI® (PolyPlus). Pour certaines expériences, les cellules sont aussi transfectées par des plasmides codant les protéines ASF/SF2, PTB, KH1-4, 9G8, SC35 ou hnRNP A1. La production des protéines recombinantes a été vérifiée par Western-blot.

Le lendemain, le milieu est renouvelé ; et 48 heures après transfection, les ARN sont isolés avec le kit GenElute™ Mammalian Total RNA Miniprep (Sigma-Aldrich), selon le protocole

indiqué par le fournisseur. Ils sont ensuite quantifiés par mesure de la densité optique à 260 et 280 nm (NanoDrop).

b. Analyse par RT-PCR

• Traitement des ARN à la DNase

- ARN	2 µg
- Tampon 10X (Tris-HCl 400 mM pH 8,0 ; MgSO <sub>4</sub> 100 mM ; CaCl <sub>2</sub> 10 mM)	1 µL
- DNase RQ1 (Promega, 1 U/µL)	2 µL
- H <sub>2</sub> O	qsp 10 µL

Après une incubation 30 min à 37°C, la réaction est stoppée par ajout de 10 µL d'EDTA 30 mM. Le mélange est alors incubé 10 min à 65°C puis à 4°C.

• Hybridation de l'oligonucléotide polyT à l'extrémité 3' des ARN

- ARN	10 µL
- Amorces dT (100 µM)	1 µL

L'hybridation est réalisée par incubation du mélange 5 min à 65°C.

• Réaction de transcription inverse

- Tampon 5X (Tris-HCl 250 mM pH 8,3 ; KCl 375 mM ; MgCl <sub>2</sub> 15 mM ; DTT 50 mM)	4 µL
- dNTP (10 mM de chaque)	2 µL
- MgCl <sub>2</sub> (50 mM)	2 µL
- Transcriptase inverse (MMLV, 200 U/µL, Promega)	1 µL
- H <sub>2</sub> O	1 µL

La réaction se déroule pendant 1h à 42 °C puis le mélange est incubé 4°C.

• Amplification des ADN complémentaires synthétisés

- Matrice	2 µL
- Tampon 10X (Tris-HCl 10 mM, pH 8,8 ; KCl 50 mM ; Triton X-100 0.1 %)	2,5 µL
- dNTP (10 mM)	4 µL
- Amorces (20 mM de chaque)	0,5 µL
- DreamTaq (Fermentas, 5 U/µL)	0,25 µL
- H <sub>2</sub> O	qsp 25 µL

Après une dénaturation de 5 min à 95°C, les cycles d'amplification des ADN ont été établis pour chaque couple d'amorces utilisé :

- pour le couple de la β-D-microglobuline (O-3253/O-3254) :  
 { 30 s à 95°C ; 45 s à 55°C ; 30 s à 72°C }<sub>30x</sub> ;
- pour les couples d'amorces correspondant à une amplification des ADNc de l'ARN Xist : { 30 s à 95°C ; 45 s à 55°C ; 1 min 10 à 72°C }<sub>35x</sub> ;

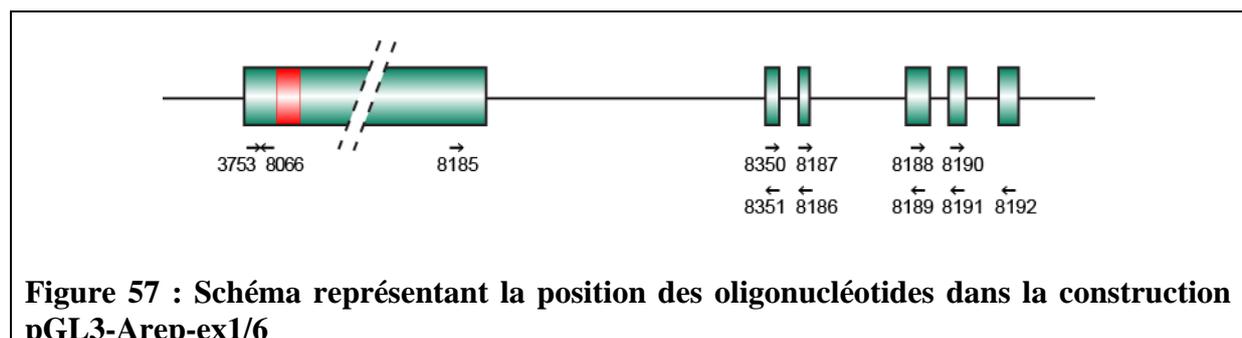
- pour le couple d'amorces amplifiant la forme épissée de l'ARN C2 (O-6519/O-6246) : {30 s à 95°C ; 45 s à 50°C ; 45 s à 72°C }<sub>35x</sub> ;
- pour le couple d'amorces amplifiant les formes épissée et non épissée de l'ARN C2 (O-6519/O-6187) : {30 s à 95°C ; 45 s à 50°C ; 45 s à 72°C }<sub>30x</sub>.

Tous les cycles se terminent par une incubation de 10 minutes à 72°C. Les produits PCR amplifiés sont ensuite analysés par électrophorèse en gel agarose 1 % en présence de bromure d'éthidium et analysés par le système d'images GeneGenius (Syngene). Les séquences des oligonucléotides utilisés pour les expériences de RT-PCR sont répertoriées dans le **Tableau 5**.

**Tableau 5 : Liste des oligonucléotides utilisés pour les expériences de RT-PCR**

Oligonucléotides	Utilisation
<b>dT15</b> : TTTTTTTTTTTTTT	Amorce antisens pour la RT
<b>O-8185</b> : GCTGGTTCGTCTATCTTGTTGGG <b>O-8186</b> : CAGAGTAGCGAGGACTTGAAGAG <b>O-8350</b> : GAAGGATTGCCTGGATTTAGAGG <b>O-8351</b> : CCTCTAAATCCAGGCAATCCTTC <b>O-8187</b> : ACTGCCAGCAGCCTATACAG <b>O-8188</b> : AATAGCTCACCACCACTGGG <b>O-8189</b> : CCCAGTGGTGGTGAGCTATT <b>O-8190</b> : ACCTTCGAGGAAGCCATTCT <b>O-8191</b> : AGAATGGCTTCCTCGAAGGT <b>O-8192</b> : ATGTTGATCCTCGGGTCATTAT <b>O-3753</b> : CGGCTTGCTCCAGCCATGTT <b>O-8066</b> : CCTAGGCCAAAAGGCACAACAAAGAAATATTCC	PCR sur les exons de <i>Xist</i>
<b>O-6519</b> : GGTCTCTCTGGTTAGACCAG <b>O-6246</b> : TATAGCAGAGTCAGTCGC <b>O-6187</b> : CGCTTTCAAGTCCCTGTT	PCR sur C2
<b>O-6164</b> : TACCTGTGGAGGCACGTCGT <b>O-6165</b> : CTCAGGCTCCAGTTTCCGCA	PCR sur la Renilla
<b>O-3253</b> : ACCCCCACTGAAAAAGATGA <b>O-3254</b> : ATCTCAAACCTCCATGATG	PCR sur la β-D-microglobuline

La séquence des oligonucléotides est indiquée dans le sens 5' vers 3'. Leur utilisation est précisée dans la partie droite du tableau.



**Figure 57 : Schéma représentant la position des oligonucléotides dans la construction pGL3-Arep-ex1/6**

### II.5.5 Mesure de l'activité luciférase

Des cellules HEK293 sont ensemencées à 200 000 cellules par puits dans des plaques de 6 puits, 24 heures avant la transfection. Les cellules sont alors transfectées avec 1 µg de mini-gène (pGL3c et pGL3-Arep-luci, pGL3-4rep-luci, pGL3-2rep-luci et pGL3-ΔArep-luci) par puits en suivant les instructions du fournisseur du kit JetPEI® (PolyPlus). Le lendemain, le milieu est renouvelé : 48 heures après transfection, les cellules sont rincées au PBS 1x puis traitées selon les indications du fournisseur du kit Dual-Luciferase® Reporter Assay (Promega). Les cellules sont lysées par ajout de 500 µL de tampon PLB 1x (le PLB 5x est fourni dans le kit) et mises sous agitation douce pendant 15 min à température ambiante. Un volume de 20 µL est ensuite transféré dans une plaque 96 puits à fond blanc. L'appareil SAFAS Monaco est muni d'auto-injecteurs et il est paramétré pour dispenser 100 µL de réactif LARII (Luciferase Assay Reagent II) qui permet la mesure de l'activité de la luciférase Firefly. La réaction est arrêtée en même temps qu'est déclenchée la mesure de l'activité de la luciférase Renilla, par ajout de 100 µL de réactif Stop & Glo dans le même échantillon.

### II.5.6 Analyse du temps de demi-vie des ARN

Des cellules HEK293 sont ensemencées à 200 000 cellules par puits dans des plaques 6 puits. Après 24 heures de culture dans du milieu DMEM, les cellules sont transfectées avec 1 µg de pGL3-ΔArep-ex1/6, pGL3-2rep-ex1/6, pGL3-4rep-ex1/6, pGL3-Arep-ex1/6 respectivement et 1 µg de phRGB-SV40, en suivant les instructions du fournisseur du kit JetPEI® (PolyPlus). Après 48 heures, les cellules sont traitées en changeant le milieu de culture par du milieu contenant 5 µg/mL d'actinomycine D (Sigma). Les ARN sont ensuite extraits après 0, 3 et 6 heures de traitement avec le kit GenElute™ Mammalian Total RNA Miniprep (Sigma-Aldrich). La quantité d'ARN a été analysée par expériences de RT-PCR, selon les conditions décrites précédemment, avec les couples 3753/8066 pour l'ADNc de Xist et 6164/6165 pour l'ADNc de la rénilla. Les produits sont déposés sur gel d'agarose et les bandes sont quantifiées à l'aide du logiciel GeneTools de SynGene.

### II.5.7 Quantification et significativité

Toutes les quantifications ont été effectuées à l'aide du logiciel GeneTools de SynGene. Pour les expériences réalisées 3 fois, les valeurs ont été moyennées et les écarts-types calculés. Que ce soit pour les quantifications d'intensité de bandes PCR ou pour les activités relatives de la luciférase, la significativité des valeurs a été calculée par le test de

Student, avec une distribution unilatérale et de type homoscédastiques (les échantillons étant de variance égale).

---

# RESULTATS

---

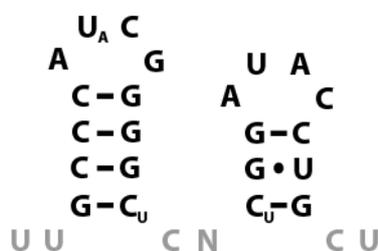


## Résultats

### I. Analyse de la structure 2D de la région des A-repeats par FRET

#### I.1 Introduction

La détermination de la structure de la région des A-repeats de l'ARN Xist est primordiale dans l'objectif de comprendre son mécanisme d'action. Néanmoins, son étude est difficile en raison de la présence d'éléments répétés et de régions riches en résidus A et U. Comme nous l'avons mentionné dans l'Introduction, au moment où j'ai débuté ma thèse, l'équipe d'A. Wutz avait proposé une structure pour chaque élément répété qui compose cette région (**Figure 58**) (Wutz *et al.*, 2002).



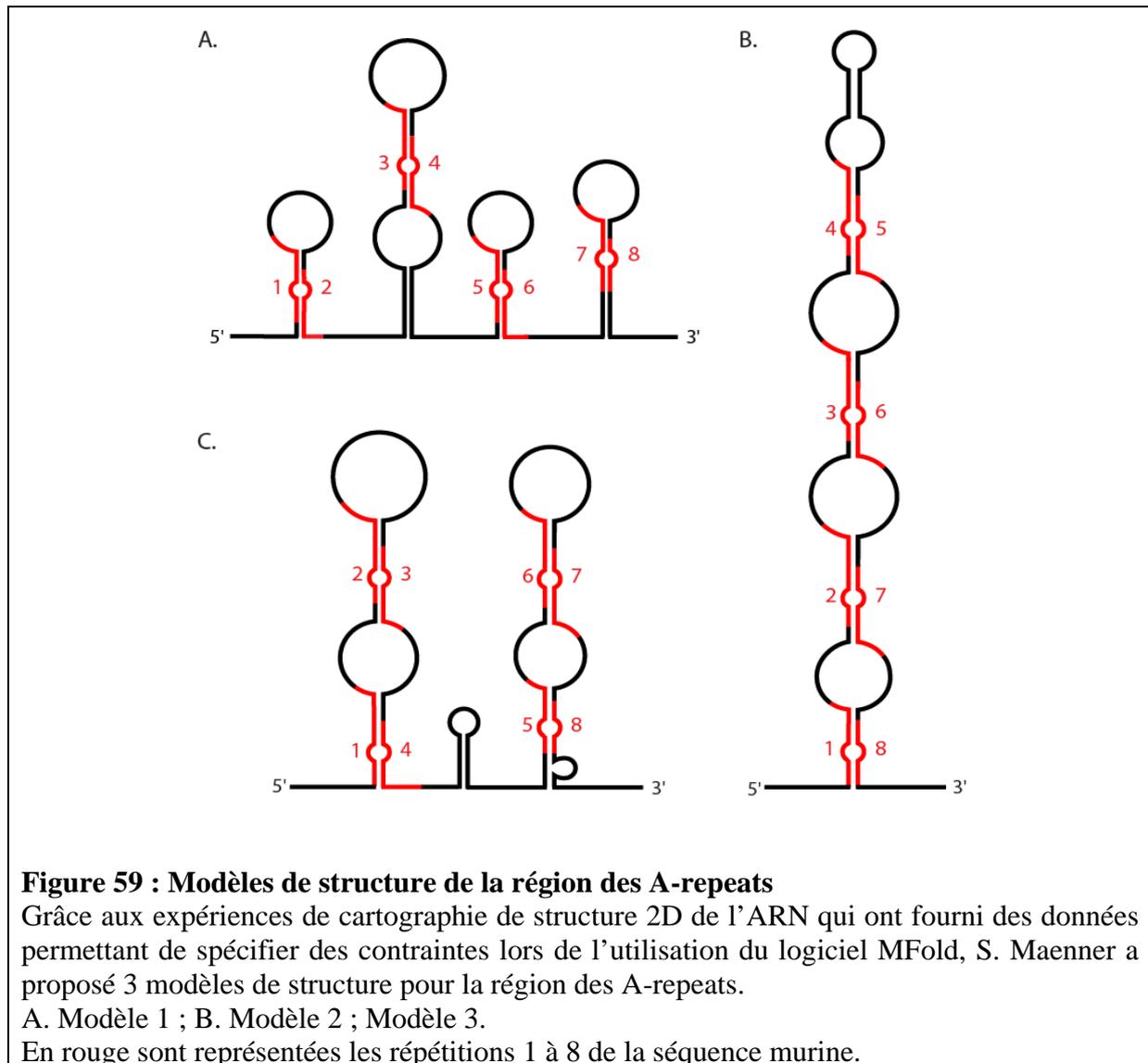
**Figure 58 : Représentation de la structure d'un élément répété de la région des A-repeats selon A. Wutz**

Prédiction de la structure secondaire d'un élément répété de la région des A-repeats

Cependant, S. Maenner a réalisé une étude sur la région entière, murine et humaine. Par l'emploi de sondes chimiques et enzymatiques qui lui ont permis de déterminer les régions en simple brin et celles en double brin. Les résultats observés sur chaque élément répété ne permettaient pas de retrouver le modèle proposé par Wutz et ses collègues. En effet, des nucléotides appariés dans ce modèle étaient obtenus en simple brin dans ces expériences, et inversement.

S. Maenner a ensuite utilisé le logiciel MFold en utilisant les données expérimentales précédentes comme contraintes. Il en est ressorti plusieurs modèles de structure possibles présentant une caractéristique commune, les éléments répétés y étaient associés deux à deux, mais selon différentes combinaisons. Plus précisément, 3 modèles en accord avec les données expérimentales et les données de stabilité thermodynamique fournies par MFold ont été retenus. Ils sont décrits dans l'article qui sera présenté dans le paragraphe II des Résultats (Maenner *et al.*, 2010) et dans la **Figure 59**. Le modèle 3 semblait tout de même le plus

favorable puisque la séquence correspondant aux 4 premières répétitions était prédite comme la structure 2D la plus stable, même sans ajouter de contraintes au logiciel.



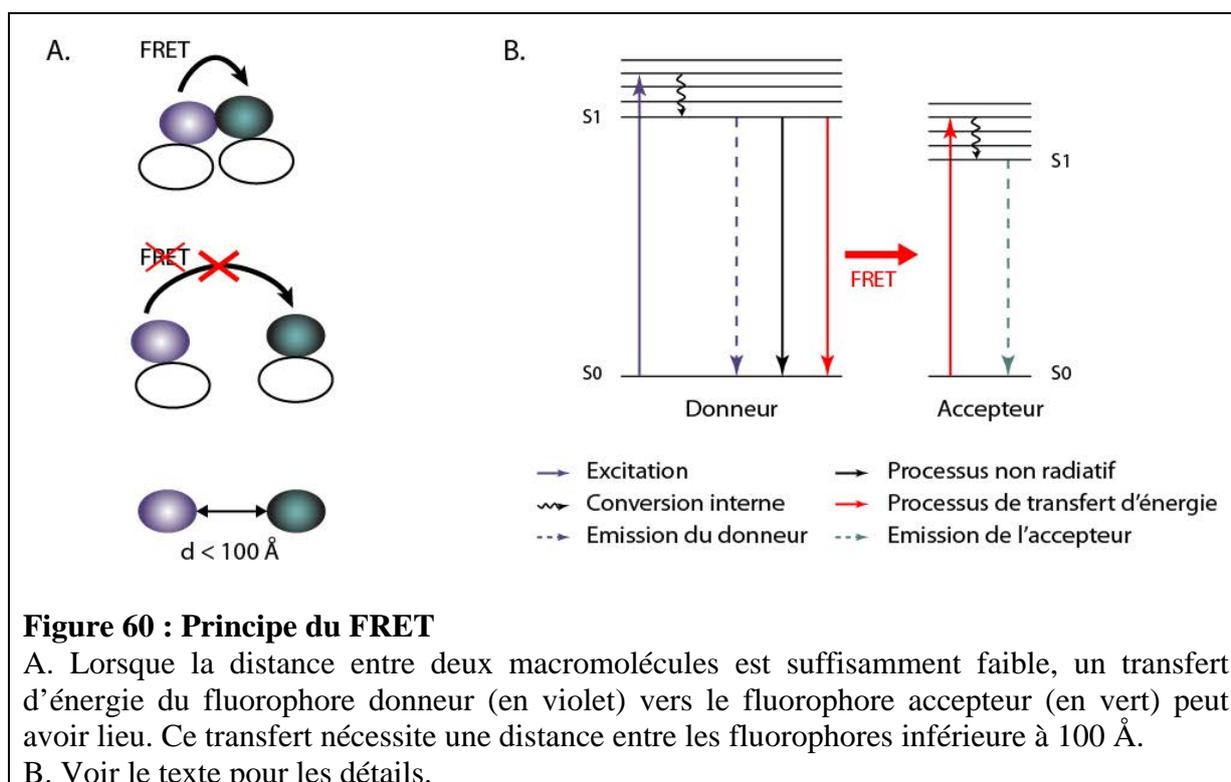
Des données phylogénétiques supplémentaires, par alignement de la séquence du gène *Xist* des mammifères pour lesquels elle était connue, ont permis de montrer la conservation de structuration en tiges-boucles des quatre premières répétitions comme dans le modèle 3. L'ensemble de ces données suggérait donc que la région des A-repeats adopte la structure 3 en solution. Cependant, il était nécessaire d'en apporter la preuve formelle. C'est pourquoi nous avons mis en place une technique d'étude novatrice pour la structure 2D des ARN. Elle repose sur le transfert d'énergie entre deux fluorophores (FRET).

Les expériences de FRET ont été mises au point et réalisées en collaboration avec S. Maenner et M. Blaud (au cours de son stage post-doctoral).

## I.2 Présentation de la technique de FRET

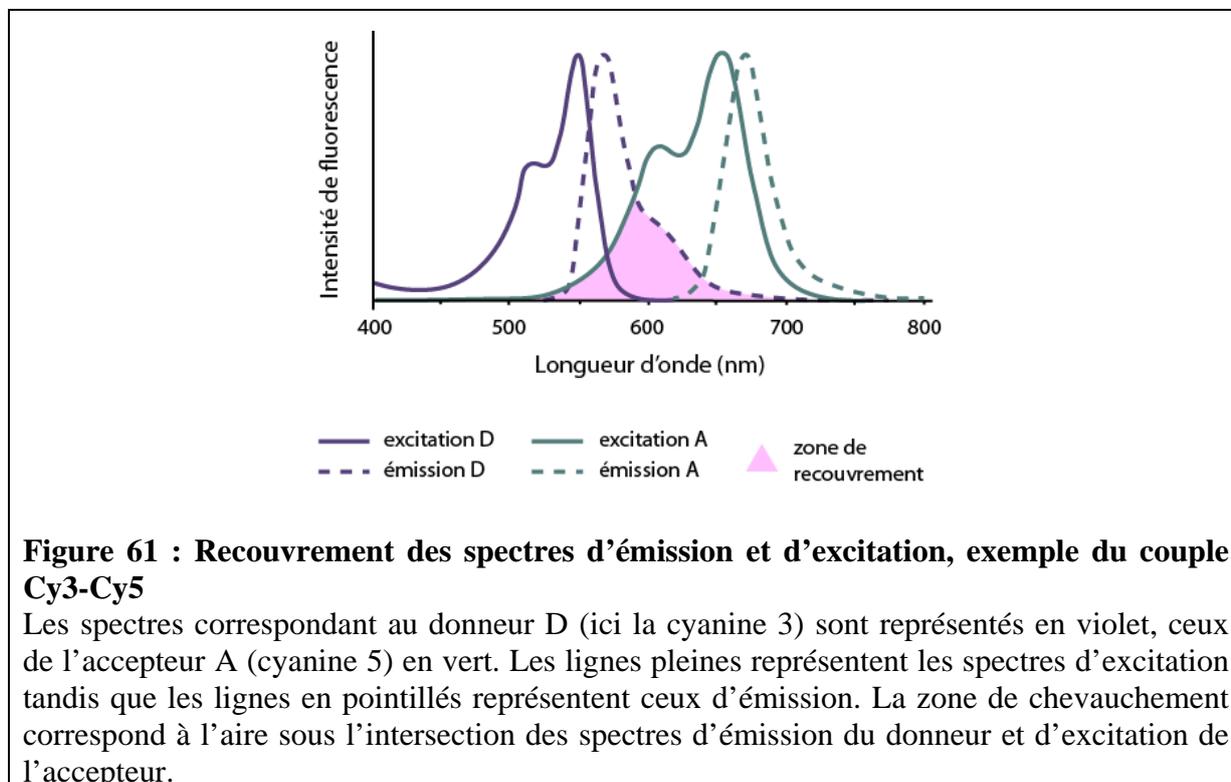
Le FRET (Förster Resonance Energy Transfer) est un processus physique dépendant de la distance entre deux fluorophores, par lequel l'énergie est transférée de manière non radiative d'un fluorophore moléculaire excité (le donneur, D) vers un autre fluorophore (l'accepteur, A). Le FRET peut être un outil de mesure très précis de la proximité entre deux molécules ou entre 2 fonctions au sein d'une même molécule, pour des distances de l'ordre de la dizaine d'angström (de 10 à 100 Å) (**Figure 60.A**).

La première étape implique l'absorption d'énergie par le fluorophore donneur, ce qui a pour conséquence de le faire passer d'un état fondamental  $S_{0D}$  à un état singulet excité  $S_{1D}$ . Plusieurs états excités sont disponibles pour le donneur. Cependant, la relaxation vers  $S_{1D}$  par conversion interne est rapide, ce qui assure que la majorité des émissions s'effectue à partir de cet état. Plusieurs destins sont alors possibles pour l'émission du donneur, incluant l'émission spontanée et le processus non radiatif. Si un fluorophore adapté est présent dans son environnement proche, alors un transfert d'énergie non radiative entre le donneur et l'accepteur peut avoir lieu (**Figure 60.B**).

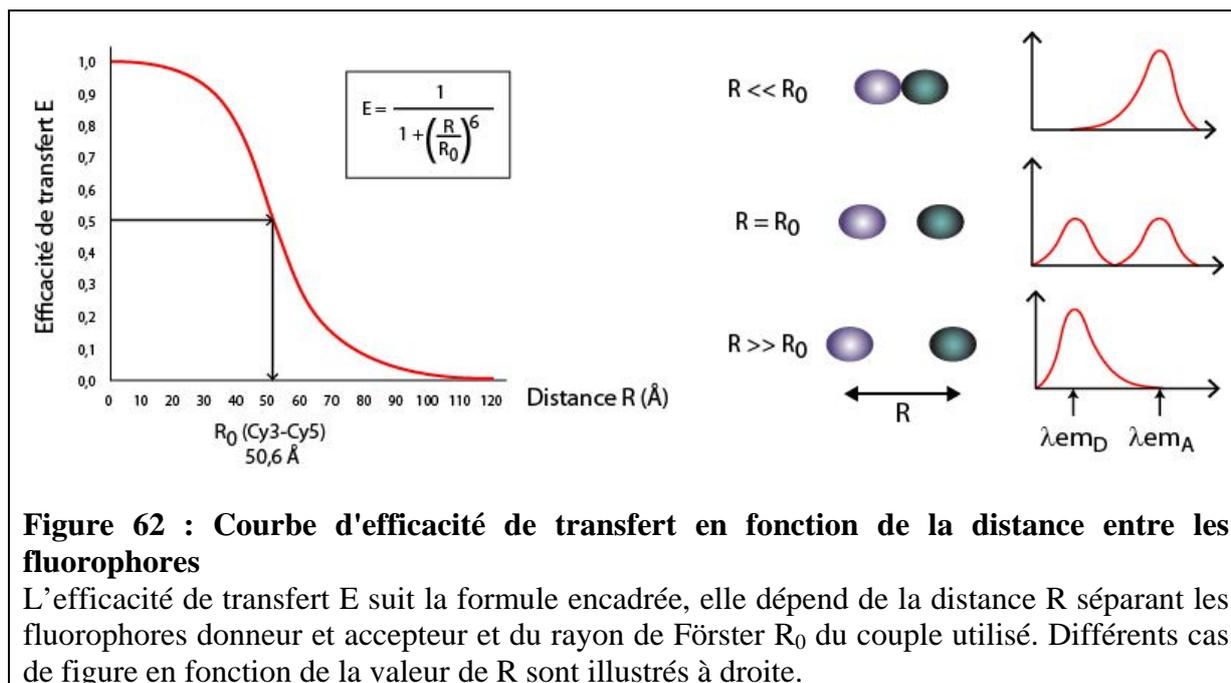


Ce transfert implique une résonance entre les transitions électroniques singulet-singulet des deux fluorophores. Ainsi, l'efficacité du FRET et la gamme de distances dans laquelle il peut être observé sont déterminées par les propriétés spectrales du couple donneur-accepteur. Les

molécules donneur et accepteur sont sélectionnées sur la base de leurs caractéristiques spectrales d'absorption et d'émission. Pour un transfert d'énergie de résonance optimal, le spectre d'émission du donneur doit recouvrir considérablement le spectre d'excitation de l'accepteur (**Figure 61**).



Si un système contient deux espèces de fluorophores dont le spectre d'émission du premier (D) recouvre celui d'excitation du second (A), alors l'énergie absorbée par D peut être transférée à A sur une distance qui dépend de l'intégrale de chevauchement des deux spectres. Ce phénomène peut être utilisé dans une multitude de cas afin de déterminer la distance entre deux fluorophores. D'après la théorie de Förster, l'intensité du phénomène de FRET suit la formule suivante :  $E = 1/[1+(R/R_0)^6]$ , où E est l'efficacité de transfert (%), R la distance entre D et A et  $R_0$  le rayon de Förster. Ce dernier correspond à la distance pour laquelle l'efficacité de transfert est de 50%. Cette distance dépend de la nature du couple de fluorophores utilisé. Par exemple, le couple Cy3-Cy5 que nous avons adopté pour l'étude a un  $R_0$  de 50,6 Å (**Figure 62**).



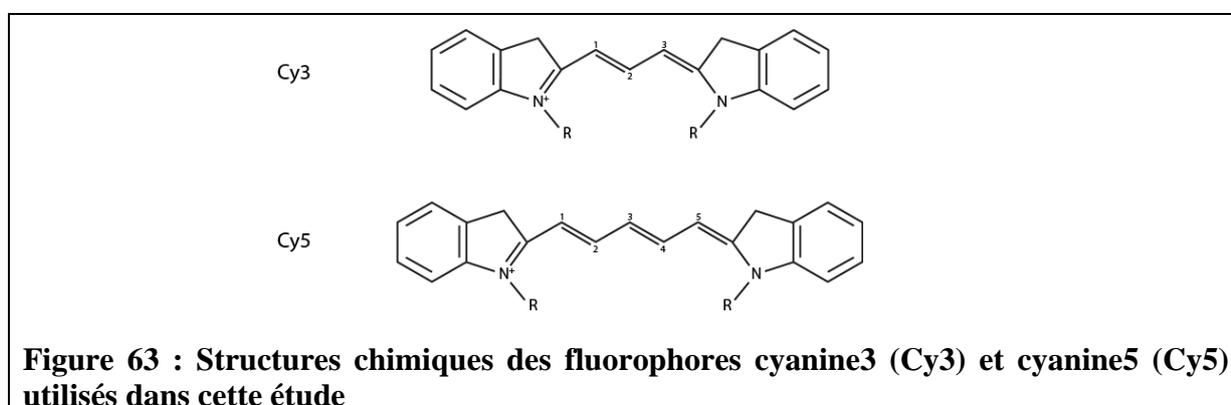
**Figure 62 : Courbe d'efficacité de transfert en fonction de la distance entre les fluorophores**

L'efficacité de transfert E suit la formule encadrée, elle dépend de la distance R séparant les fluorophores donneur et accepteur et du rayon de Förster  $R_0$  du couple utilisé. Différents cas de figure en fonction de la valeur de R sont illustrés à droite.

### I.3 Mise au point du protocole expérimental

#### I.3.1 Choix des fluorophores

Nous avons choisi le couple de fluorophores cyanine3-cyanine5. Il s'agit de molécules constituées de cycles aromatiques reliés par une chaîne de 3 (Cy3) ou 5 (Cy5) atomes de carbone (**Figure 63**). Ce couple est actuellement l'un des plus utilisés car le spectre d'émission de Cy3 (le donneur) recouvre largement celui de Cy5 (l'accepteur) (cf. **Figure 61**). De plus, ces fluorophores sont plus stables que les autres fluorophores disponibles tels que le couple FITC/Rhodamine. Les longueurs d'onde caractéristiques d'excitation et d'émission du Cy3 sont respectivement de 548 et 565 nm ; celles du Cy5 sont respectivement de 646 et 662 nm.



**Figure 63 : Structures chimiques des fluorophores cyanine3 (Cy3) et cyanine5 (Cy5) utilisés dans cette étude**

En théorie, lors d'une mesure de FRET, l'excitation est effectuée à la longueur d'onde d'excitation du donneur, c'est-à-dire 548 nm dans le cas du groupement Cy3. La mesure de fluorescence est réalisée sur un spectre large, ici de 500 à 750 nm, de manière à observer l'émission de l'accepteur (à 646 nm pour le groupement Cy5) mais aussi du donneur (à 564 nm pour Cy3). Cependant, il apparaît un pic résiduel à la longueur d'onde d'excitation. Les longueurs d'onde de 548 et 564 nm étant proches, ce pic empiète sur celui correspondant au signal d'émission du donneur, ce qui gêne la lecture de ce dernier. Pour éviter ce phénomène, nous avons opté pour une excitation du donneur Cy3 à 515 nm.

### I.3.2 Choix des oligonucléotides

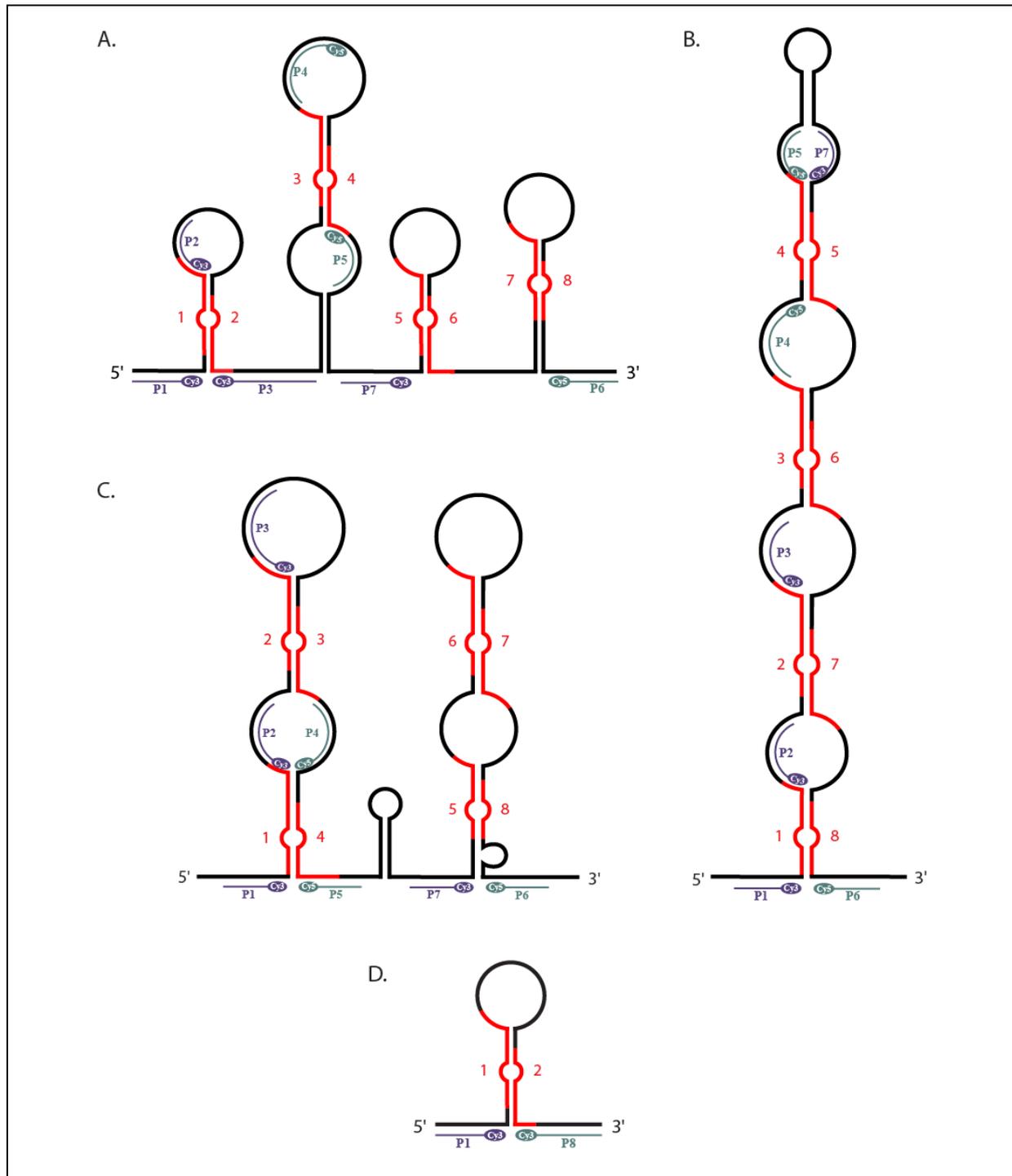
Les oligonucléotides utilisés devaient répondre à plusieurs critères. D'une part, la région des A-repeats est composée d'éléments répétés. La séquence des oligonucléotides ne devait donc pas être complémentaire à une région répétée afin d'assurer sa spécificité d'hybridation. D'autre part, il était essentiel que la région de l'ARN auquel l'oligonucléotide allait se fixer soit en simple brin, sous peine d'altérer la structure 2D. Différents modèles de structure étaient possibles (cf. **Figure 59**) : le modèle 1 dans lequel la région des A-repeats se structure selon 4 tiges-boucles, chacune étant formée par l'appariement de 2 à 2 de 2 éléments répétés successifs ; le modèle 2 structuré en une grande tige-boucle formée par l'appariement 2 à 2 de tous les éléments répétés ; et le modèle 3, dans lequel les tiges-boucles sont formées par l'appariement 2 à 2 de 4 éléments répétés. Dans tous ces modèles, les régions en simple brin sont identiques, seule la manière dont les éléments répétés s'apparient est différente. Ainsi, un oligonucléotide s'hybridant dans une région simple brin selon le modèle 1, s'hybride également dans une région simple brin selon les autres modèles.

Le but étant de valider l'un des 3 modèles, nous devons choisir les oligonucléotides de manière à ce que les résultats de FRET obtenus permettent de les discriminer. Pour cela, nous avons défini 7 oligonucléotides, P1 à P7, répondant aux critères énoncés précédemment (**Figure 64**). Ainsi, l'utilisation des couples P1-P5, P2-P4 et P7-P6 devait nous permettre de différencier les 3 modèles puisque les distances sont différentes pour chacun d'entre eux. En effet, les oligonucléotides P1 et P5 s'hybrident à la base de la première tige-boucle du modèle 3, tandis qu'ils s'hybrident dans 2 tiges-boucles différentes dans le modèle 1 et sont totalement à l'opposé l'un de l'autre dans le modèle 2. Nous devons observer un fort transfert d'énergie si la région se structurait selon le modèle 3, alors qu'il devait être faible dans les deux autres cas possibles.

En plus de ces 7 oligonucléotides, nous avons utilisé l'oligonucléotide P8 pour les mesures sur le témoin. Il s'agit de l'ARN correspondant aux 2 premières répétitions de la région des A-repeats, l'ARN 2R. Cet ARN ne peut se structurer que selon un modèle formé d'une tige-boucle bordée de 2 séquences simple brin (**Figure 64.D**). L'oligonucléotide P8 possède la même séquence que P3 mais le groupement greffé à l'extrémité 3' est dans ce cas le fluorophore Cy5. Utilisé avec l'oligonucléotide P1, ils s'hybrident de part et d'autre de l'hélice formée par l'ARN, les fluorophores se faisant face. L'efficacité de transfert obtenue dans cette situation correspond donc à l'efficacité maximale.

Les efficacités observées par utilisation des autres couples sur la région entière pouvaient être comparées à cette valeur, et nous avons ainsi pu déterminer leur proximité relative. Par exemple, les couples P1-P5 et P7-P6 s'hybrident à la base d'hélices selon le modèle 3. Si ce modèle correspond à la structure adoptée par l'ARN, ils devraient présenter une efficacité de transfert maximale, c'est-à-dire du même ordre que celle obtenue avec le témoin.

La séquence des oligonucléotides utilisés est disponible dans le paragraphe II.1.2 de la partie Matériel et Méthodes.



**Figure 64 : Oligonucléotides utilisés pour évaluer la validité des 3 modèles possibles de la région des A-repeats**

A. Modèle 1.

B. Modèle 2.

C. Modèle 3.

D. Témoign utilisé pour déterminer le taux de transfert entre 2 oligonucléotides s'hybridant à la base d'une structure tige-boucle.

Les répétitions et leur numéro sont indiqués en rouge. Les oligonucléotides portant un groupement Cy3 sont représentés en violet, ceux portant un groupement Cy5 en vert.

### I.3.3 Hybridation ARN-oligonucléotides

Les oligonucléotides sélectionnés devaient être complémentaires à des régions en simple brin, mais les conditions d'hybridation restaient à être déterminées. Plusieurs points du protocole ont ainsi été optimisés.

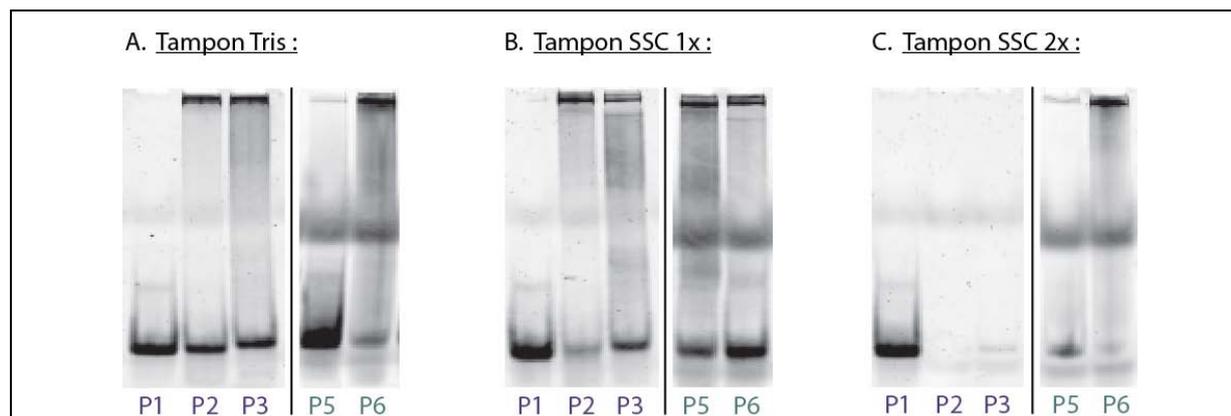
#### a. La concentration en ARN

Les premiers essais ont été réalisés à une concentration en ARN de 0,4  $\mu\text{M}$  (soit 10  $\mu\text{g}$ ) et il s'est avéré que cette concentration permettait d'obtenir une intensité de signal de fluorescence entre 50 et 80 UA, ordre de grandeur idéal pour observer des variations positives comme négatives. Pour des raisons pratiques, la concentration effectivement utilisée a été de 0,38  $\mu\text{M}$ .

#### b. Le tampon et le temps d'hybridation

L'hybridation entre l'ARN et les oligonucléotides a été analysée après électrophorèse en gel acrylamide non dénaturant 7% suivie d'une mesure de fluorescence grâce à un appareil Typhoon (9410, GE Healthcare) (**Figure 65**). Le premier tampon testé était composé de Tris-HCl à 83 mM, de KCl à 150 mM et de  $\text{MgCl}_2$  à 5 mM, pH 7,5. Même si l'hybridation a bien eu lieu avec certains oligonucléotides (par exemple P2, P3 et P6, **Figure 65.A**), d'autres sont restés sous forme libre, comme les oligonucléotides P1 et P5. Ensuite, nous avons testé une hybridation en tampon SSC 1x final, c'est-à-dire du chlorure de sodium à 150 mM et du citrate de sodium à 15 mM, pH 7,0, complété de 3,125 mM de  $\text{MgCl}_2$ . Avec ce tampon, nous avons observé l'hybridation de l'oligonucléotide P5 qui n'avait pas lieu en présence du tampon Tris (**Figure 65.B**). Nous avons augmenté la quantité de sels en incubant le mélange dans du tampon SSC 2x, mais cela a provoqué une diminution de l'hybridation (**Figure 65.C**). Le tampon sélectionné a donc été le tampon à pH 7,0 contenant 150 mM chlorure de sodium, 15 mM citrate de sodium et 3,125 mM de  $\text{MgCl}_2$ .

L'oligonucléotide P1 ne s'hybridant dans aucun cas, nous avons modifié sa séquence en supprimant 2 nucléotides à son extrémité 5' qui se situaient juste avant le début de l'hélice et pouvaient gêner l'hybridation par encombrement stérique dû au groupement Cy3 greffé en 5'. Le nouvel oligonucléotide P1 s'est en effet très bien hybridé. La séquence indiquée dans la partie Matériel et Méthodes correspond au P1 modifié.



**Figure 65 : Tests de tampon d'hybridation des oligonucléotides à l'ARN**

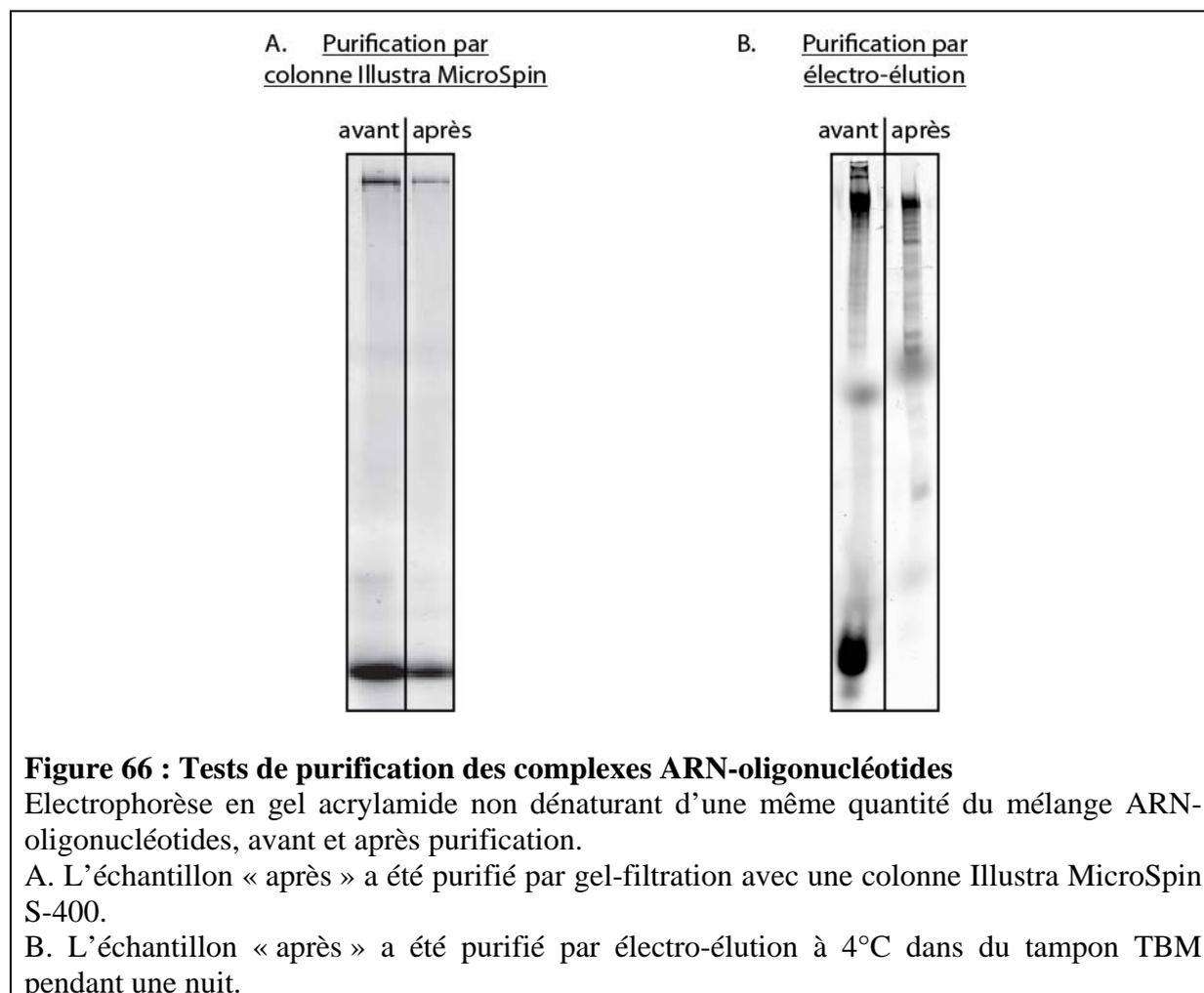
L'ARN à une concentration de  $0,38 \mu\text{M}$  a été incubé avec chaque oligonucléotide (les oligonucléotides P1 à P3 en violet portent un groupement Cy3, les oligonucléotides P5 et P6 en vert portent un groupement Cy5) en conditions stœchiométriques dans les différents tampons d'hybridation indiqués au dessus de chaque gel. Les échantillons ont ensuite été fractionnés en gel acrylamide non dénaturant 7 % et la fluorescence émise a été mesurée au Typhoon aux longueurs d'onde de 515 et 646 nm selon l'oligonucléotide (Cy3 et Cy5 respectivement), afin de visualiser les formes libres et les formes complexées.

Nous avons également remarqué qu'une hybridation sur la nuit entière n'augmentait pas la proportion d'oligonucléotides en complexe avec l'ARN. Nous avons donc choisi de la réaliser sur une durée de 4 heures, à une température de  $4^{\circ}\text{C}$  afin de préserver l'intégrité de l'ARN.

#### c. La concentration en oligonucléotides et la purification des complexes

Nous avons réalisé plusieurs essais avec des ratios ARN:oligonucléotides variant de 1:1 à 1:10 afin d'accroître la quantité d'oligonucléotide fixé. Nous avons remarqué qu'un excès d'oligonucléotides ne permettait pas d'augmenter de manière significative la proportion de complexes. En revanche, la quantité d'oligonucléotide sous forme libre étant importante, nous avons décidé d'ajouter une étape de purification des complexes afin que les oligonucléotides libres n'interfèrent pas avec les mesures d'émission. Différents types de purification ont été testés : (i) une gel-filtration sur billes de sépharose (Illustra MicroSpin Illustra S-400 HR, GE Healthcare). Cette méthode permet de retenir les petites molécules sur la colonne, et donc d'éluer les complexes ARN-oligonucléotides. Nous n'avons malheureusement pas observé de disparition totale des oligonucléotides libres mais seulement une diminution. De plus, la quantité de complexes ARN-oligonucléotides a également diminué (**Figure 66.A**). Donc la proportion d'oligonucléotides [complexés] / [libres] était au final identique à celle de départ.

(ii) une électro-élution en conditions non dénaturantes (tampon TBM) après électrophorèse en gel acrylamide non dénaturant. Les bandes correspondant aux complexes ont été découpées puis placées dans une cuve à électro-élution soumise à un champ électrique pendant toute la nuit à une température de 4°C. Les complexes ARN-oligonucléotides ont ensuite été récupérés et à nouveau déposés sur gel acrylamide non dénaturant. Bien que plus aucun oligonucléotide ne soit sous forme libre comme souhaité, nous avons observé une forte dégradation de l'ARN malgré toutes les précautions prises contre les RNases (**Figure 66.B**).



Nous avons donc opté pour des quantités stœchiométriques d'ARN et d'oligonucléotides pour plusieurs raisons. Tout d'abord, la quantité de complexes formés était similaire à celle avec un excès d'oligonucléotides. De plus, cette solution nous dispensait de l'étape de purification. Et enfin, elle nous permettait de réaliser les mesures d'émission des complexes binaires et ternaires à partir du même échantillon, ce qui était essentiel pour la comparaison des résultats et donc la détermination du taux de transfert d'énergie. Le protocole ainsi mis au point est présenté dans le paragraphe I.3.5.

### I.3.4 Paramètres de mesure

Nous avons testé plusieurs températures lors de la mesure de fluorescence grâce à un cryostat. Celle de 4°C a donné la meilleure reproductibilité du signal, les complexes étant plus stables. Nous avons donc choisi cette température pour les mesures.

De plus, des essais sur les paramètres du spectrofluoromètre nous ont permis de conclure que la bande passante devait être de 3 nm pour l'excitation comme pour l'émission et que le photomultiplicateur, qui sert à amplifier le signal émis afin d'atteindre des valeurs quantifiables sans pour autant saturer l'appareil, devait être réglé à 750 V.

Chaque expérience a été réalisée par la mesure de 10 spectres, et ce au moins 3 fois avec des lots d'ARN purifiés et des aliquots d'oligonucléotides différents.

### I.3.5 Protocole utilisé

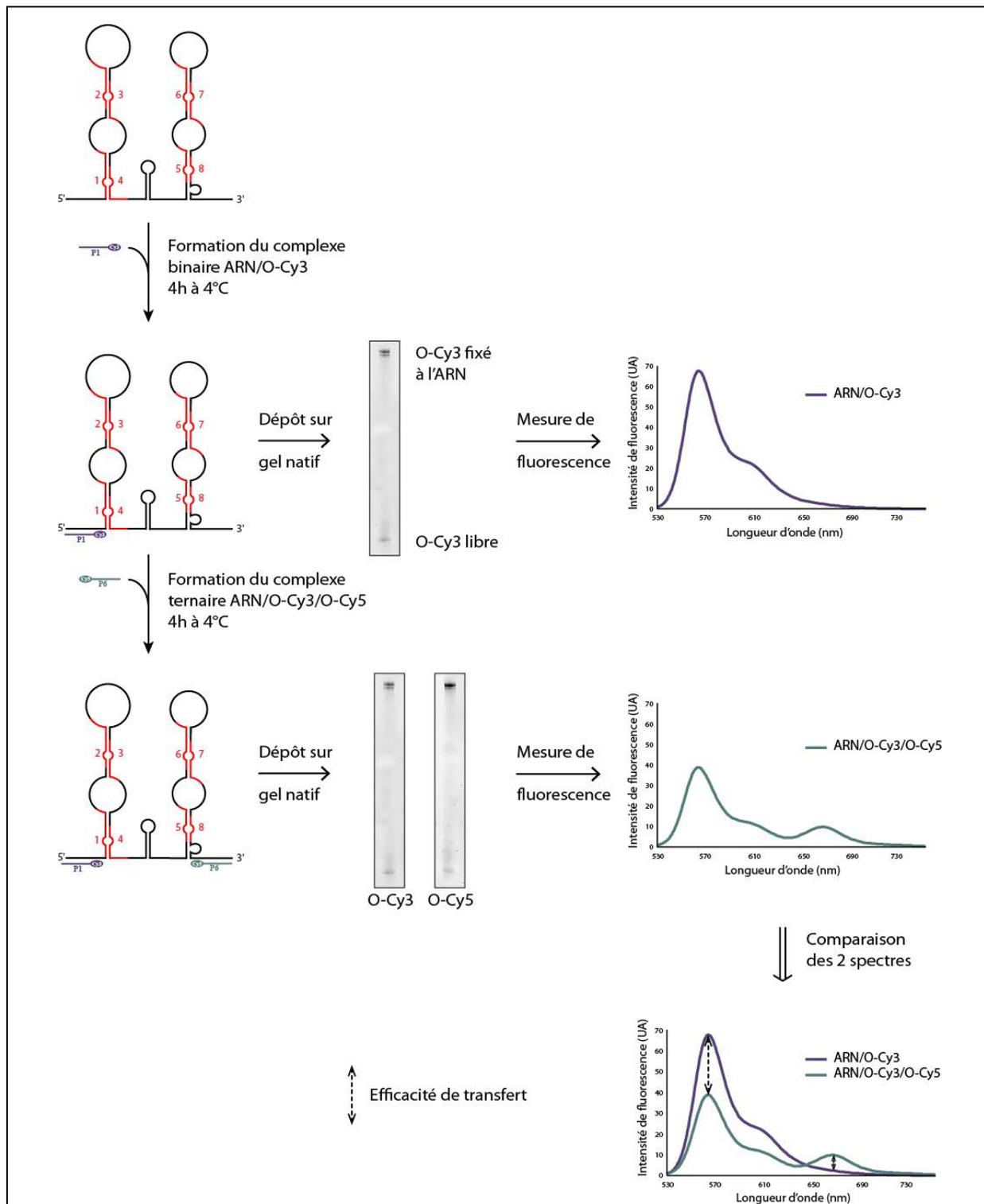
Nous avons réalisé plusieurs mesures de contrôle afin de s'assurer que tous les paramètres étaient corrects. Tout d'abord, nous avons effectué un spectre d'émission sur le tampon seul à une longueur d'onde d'excitation de 515 nm dans le but de vérifier l'absence d'émission de fluorescence. En effet, le spectre s'est avéré suivre la ligne de base, il s'agit donc d'un élément neutre dans la mesure. Nous avons ensuite réalisé un spectre d'émission sur les oligonucléotides libres, couplés à un groupement Cy3 (O-Cy3) ou Cy5 (O-Cy5), à une excitation de 515 nm et 646 nm respectivement. Cela nous a permis de vérifier la longueur d'onde d'émission de chaque fluorophore. Enfin, nous avons procédé aux mêmes mesures mais avec un mélange O-Cy3/O-Cy5, afin de s'assurer que la présence des 2 groupements n'interférait pas dans les longueurs d'onde d'émission de chacun, et surtout qu'il ne se produisait pas de transfert d'énergie entre eux à l'état libre.

Afin d'observer le potentiel effet du transfert d'énergie entre le groupement Cy3 et le groupement Cy5, il fallait tout d'abord déterminer le signal maximal d'émission du Cy3, c'est-à-dire lorsqu'il est seul hybridé à l'ARN. Nous avons ainsi réalisé un protocole en 2 grandes étapes (**Figure 67**) :

(i) Dans un premier temps, l'ARN à une concentration de 0,38  $\mu\text{M}$  est hybridé avec l'oligonucléotide portant le groupement Cy3 seul, dans un ratio 1:1. Après 4 heures d'incubation à 4°C, un aliquot est déposé sur gel non dénaturant afin de vérifier l'hybridation mais aussi de déterminer la part d'oligonucléotide libre par rapport à celle liée à l'ARN. Cette quantification sera utile lors de la comparaison des résultats obtenus lors des différentes expériences. Un spectre d'émission de ce complexe binaire ARN/O-Cy3 est ensuite réalisé à

une longueur d'onde d'excitation de 515 nm. La valeur du pic d'émission à 564 nm représente alors la valeur maximale d'émission du groupement Cy3 lié à l'ARN, en absence de groupement Cy5.

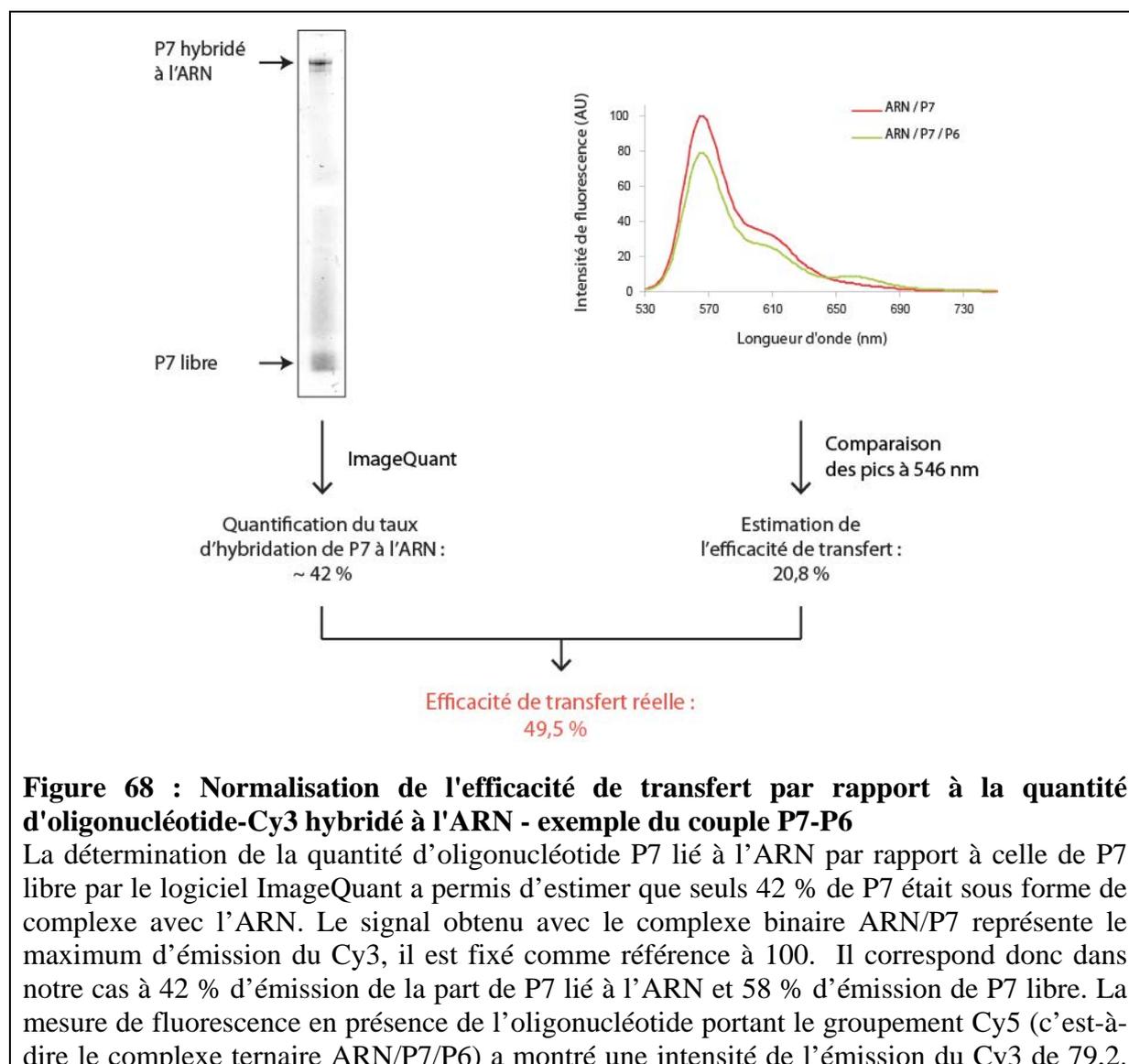
(ii) Dans un deuxième temps, nous ajoutons pendant 4 heures l'oligonucléotide portant le groupement Cy5 au complexe binaire préformé, à une concentration finale de 0,38  $\mu\text{M}$  également. De la même manière, nous avons quantifié la proportion d'oligonucléotide complexé par rapport au libre. Nous avons alors observé que seuls 3 % d'O-Cy5 étaient sous forme libre. Nous n'avons donc pas pris en compte cette faible proportion dans la détermination de l'efficacité de transfert. Nous avons ensuite réalisé le spectre d'émission du complexe ternaire ARN/O-Cy3/O-Cy5, toujours à une longueur d'onde d'excitation de 515 nm. Nous avons au préalable vérifié que l'émission du groupement Cy3 ne diminuait pas entre les 2 mesures de spectres, en suivant un protocole similaire mais avec l'ajout de tampon au lieu de l'oligonucléotide-Cy5 dans la deuxième étape. L'observation d'une diminution de la hauteur du pic à 564 nm (correspondant à la longueur d'onde maximale d'émission du Cy3, le donneur D) et de l'apparition d'un pic à 662 nm (correspondant à la longueur d'onde maximale d'émission du Cy5, l'accepteur A) nous indiquait la présence d'un transfert d'énergie  $D \rightarrow A$ .



**Figure 67 : Protocole suivi pour les mesures de FRET**

L'ARN à une concentration de  $0,38 \mu\text{M}$  est incubé 4 heures à une température de  $4^\circ\text{C}$  avec l'oligonucléotide O-Cy3 dans une stœchiométrie 1:1. Après vérification de la formation des complexes et estimation de la quantité d'oligonucléotide libre et lié sur gel natif d'acrylamide 7 %, le spectre d'émission a été enregistré (500 à 750 nm) à une longueur d'onde d'excitation de 515 nm. Le mélange est ensuite incubé avec l'oligonucléotide O-Cy5 (stœchiométrie 1:1:1) pendant 4 heures à  $4^\circ\text{C}$ . Après vérification de la formation du complexe ternaire sur gel natif, le spectre d'émission a été mesuré et l'efficacité de transfert d'énergie entre le groupement donneur et le groupement accepteur a pu être déterminée.

L'efficacité de transfert a été estimée par comparaison entre les intensités de fluorescence du groupement donneur dans le complexe binaire ARN/O-Cy3 et dans le complexe ternaire ARN/O-Cy3/O-Cy5. Cependant, l'efficacité d'hybridation des O-Cy3 à l'ARN s'est révélée différente selon les oligonucléotides. Il a donc fallu tenir compte de cette différence, et pour cela, nous avons utilisé les gels réalisés dans l'étape (i), correspondant au dépôt d'un aliquot du mélange ARN/O-Cy3 après 4 heures d'incubation. Nous avons donc quantifié le taux de fixation de chaque O-Cy3 sur l'ARN grâce au logiciel ImageQuant. Cette quantification nous a ensuite permis de normaliser chaque efficacité de transfert par rapport à la quantité d'oligonucléotide effectivement lié à l'ARN (**Figure 68**). Nous avons donc pu comparer les efficacités de transfert obtenues avec les différents couples d'oligonucléotides utilisés et ainsi déterminer à quel modèle, parmi les 3 potentiels, elles correspondaient le mieux.



Considérant que l'émission du P7 libre n'a pas diminué, elle représente toujours 58 %. La différence, soit  $79,2 - 58 = 21,2$  %, correspond à l'émission de la part du P7 lié à l'ARN. Par rapport à la valeur initiale de 42 %, cela représente une diminution de 49,5 %. Cette valeur constitue donc l'efficacité réelle du transfert d'énergie entre les oligonucléotides P7 et P6 hybridés à l'ARN.

#### I.4 Résultats obtenus

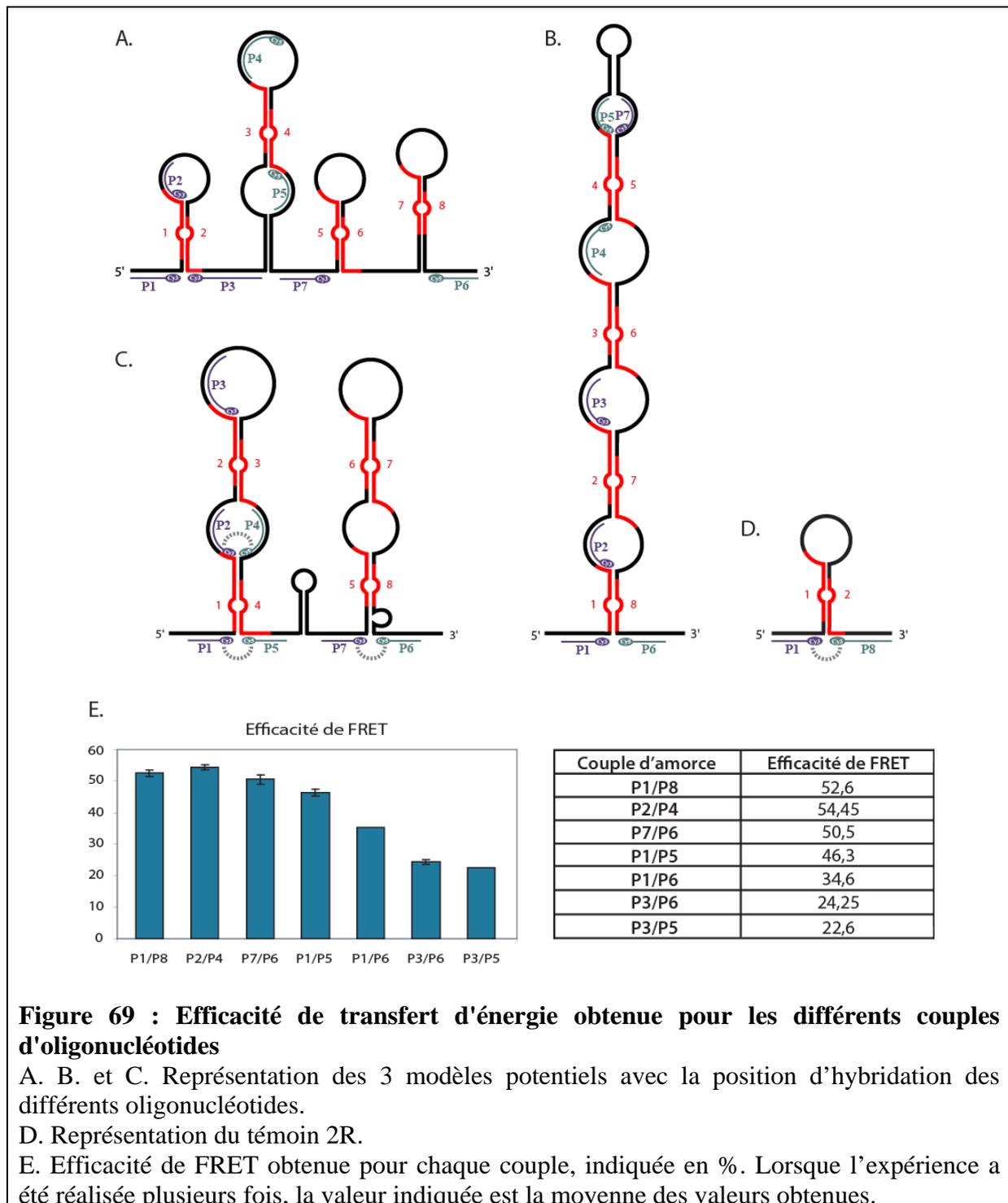
Les résultats obtenus sont détaillés dans l'article faisant l'objet du paragraphe II de cette partie Résultats. La **Figure 69** résume les valeurs obtenues.

L'efficacité de transfert obtenue avec le couple P1/P8 utilisé pour l'ARN témoin 2R a été de 52,6 % en moyenne (**Figure 69.E**, **Figure S7** de l'article). Cette valeur représente l'efficacité attendue pour un couple d'oligonucléotides s'hybridant à la base d'une hélice. Lors de l'utilisation des couples P1/P5, P2/P4 et P7/P6, l'efficacité de FRET observée a été de cet ordre de grandeur (46,3%, 54,5 % et 50,5 % respectivement) (**Figure 69.E**). Ces résultats concordent tout à fait avec le modèle 3 dans lequel ces 3 couples s'hybrident à une extrémité d'hélice (**Figure 69.C**). Ils pourraient également correspondre au modèle 2 si la structure tige-boucle se repliait sur elle-même et que la boucle terminale se retrouvait à proximité de la première hélice. Cependant les résultats suivants vont écarter ce modèle.

En effet, l'efficacité de transfert obtenue entre les fluorophores du couple P1/P6 est plus faible que celle obtenue avec le témoin. Nous pouvons en conclure que ces couples ne s'hybrident pas à la base d'une hélice, ce qui exclut le modèle 2. En revanche, le FRET observé avec les couples P1/P6 et P3/P6 satisfait une fois de plus le modèle 3.

En ce qui concerne le modèle 1, les oligonucléotides du couple P3/P5 devraient être aussi éloignés que ceux du couple P1/P5 ou P2/P4 (**Figure 69.A**). Or le phénomène de FRET observé avec P3/P5 est 2 fois plus faible qu'avec P1/P5 et P2/P4. Une structure tridimensionnelle rapprochant P1 de P5 rapprocherait également P3 de P5. Ce modèle ne convient donc pas au vu de ces résultats.

Ainsi, tous les résultats obtenus vont dans le sens d'une structuration de la région des A-repeats en 2 structures tige-boucle formées par l'appariement 2 à 2 de 4 éléments répétés successifs (**Figure 69.C**).



**Figure 69 : Efficacité de transfert d'énergie obtenue pour les différents couples d'oligonucléotides**

A. B. et C. Représentation des 3 modèles potentiels avec la position d'hybridation des différents oligonucléotides.

D. Représentation du témoin 2R.

E. Efficacité de FRET obtenue pour chaque couple, indiquée en %. Lorsque l'expérience a été réalisée plusieurs fois, la valeur indiquée est la moyenne des valeurs obtenues.

Cette structure est stable d'après les calculs thermodynamiques, cependant il est possible que sa stabilité soit renforcée, notamment par la fixation de protéines nucléaires. De plus, cette structuration particulière est certainement liée au fait que la région des A-repeats joue un rôle important dans l'inactivation du chromosome X. Il était donc logique de poursuivre l'étude de cette région par l'identification des facteurs protéiques auxquels elle est associée. C'est ce que nous verrons dans la partie III, qui succède à l'article.



**II. « 2D structure of the A region of Xist RNA and its implication for PRC2 association »**

**PLoS Biology, 2010**



# 2-D Structure of the A Region of Xist RNA and Its Implication for PRC2 Association

Sylvain Maenner<sup>1</sup>, Magali Blaud<sup>1</sup>, Laetitia Fouillen<sup>2</sup>, Anne Savoye<sup>1</sup>, Virginie Marchand<sup>1\*</sup>, Agnès Dubois<sup>3</sup>, Sarah Sanglier-Cianféron<sup>2</sup>, Alain Van Dorsselaer<sup>2</sup>, Philippe Clerc<sup>3</sup>, Philip Avner<sup>3</sup>, Athanase Visvikis<sup>1</sup>, Christiane Branlant<sup>1\*</sup>

**1** AREMS, Nancy Université, UMR 7214 CNRS-UHP 1, Faculté des Sciences et Techniques, BP 70239, Vandoeuvre-lès-Nancy, France, **2**Laboratoire de Spectrométrie de Masse BioOrganique, Institut Pluridisciplinaire Hubert Curien, Département des Sciences Analytiques, Université de Strasbourg, CNRS UMR 7178, ECPM, Strasbourg, France, **3** Génétique Moléculaire Murine, CNRS2578, Institut Pasteur, Paris, France

## Abstract

In placental mammals, inactivation of one of the X chromosomes in female cells ensures sex chromosome dosage compensation. The 17 kb non-coding Xist RNA is crucial to this process and accumulates on the future inactive X chromosome. The most conserved Xist RNA region, the A region, contains eight or nine repeats separated by U-rich spacers. It is implicated in the recruitment of late inactivated X genes to the silencing compartment and likely in the recruitment of complex PRC2. Little is known about the structure of the A region and more generally about Xist RNA structure. Knowledge of its structure is restricted to an NMR study of a single A repeat element. Our study is the first experimental analysis of the structure of the entire A region in solution. By the use of chemical and enzymatic probes and FRET experiments, using oligonucleotides carrying fluorescent dyes, we resolved problems linked to sequence redundancies and established a 2-D structure for the A region that contains two long stem-loop structures each including four repeats. Interactions formed between repeats and between repeats and spacers stabilize these structures. Conservation of the spacer terminal sequences allows formation of such structures in all sequenced Xist RNAs. By combination of RNP affinity chromatography, immunoprecipitation assays, mass spectrometry, and Western blot analysis, we demonstrate that the A region can associate with components of the PRC2 complex in mouse ES cell nuclear extracts. Whilst a single four-repeat motif is able to associate with components of this complex, recruitment of Suz12 is clearly more efficient when the entire A region is present. Our data with their emphasis on the importance of inter-repeat pairing change fundamentally our conception of the 2-D structure of the A region of Xist RNA and support its possible implication in recruitment of the PRC2 complex.

**Citation:** Maenner S, Blaud M, Fouillen L, Savoye A, Marchand V, et al. (2010) 2-D Structure of the A Region of Xist RNA and Its Implication for PRC2 Association. *PLoS Biol* 8(1): e1000276. doi:10.1371/journal.pbio.1000276

**Academic Editor:** Kathleen Hall, Washington University School of Medicine, United States of America

**Received:** August 3, 2009; **Accepted:** November 25, 2009; **Published:** January 5, 2010

**Copyright:** © 2010 Maenner et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study was supported by the French Centre National de la Recherche Scientifique (CNRS, <http://www.cnrs.fr/>), the French Ministry of "Enseignement Supérieur et la Recherche" (<http://www.enseignementsup-recherche.gouv.fr/>), the French National Agency for Research (ANR, <http://www.agence-nationale-recherche.fr/>) (contract Number ANR 07 BLAN 0047 102), Institut Pasteur, European Union contract from Epigenome NoE, Region Alsace, and AliX. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** CMCT, N-cyclohexyl-N-(2-morpholinoethyl)-carbodiimide metho-p-toluolsulfonate; DMS, dimethylsulfate; Eed, embryonic ectoderm development; ES cell, embryonic stem cell; Ezh2, enhancer of zeste homolog 2; FRET, Förster resonance energy transfer; HIV, human immunodeficiency virus; HOTAIR, HOX antisense intergenic RNA; Lnx3, ligand of numb-protein X 3; NMR, nuclear magnetic resonance; PRC2, polycomb group protein 2; PTB, polypyrimidine tract-binding protein; Rbap46–48, retinoblastoma-binding protein p46–p48; RNP, ribonucleoprotein particle; RRM, RNA recognition motif; Suz12, suppressor of zeste 12 protein homolog; XCI, X chromosome inactivation; Xist, X (inactive)-specific transcript

\* E-mail: [christiane.branlant@maem.uhp-nancy.fr](mailto:christiane.branlant@maem.uhp-nancy.fr)

‡ Current address: Developmental Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

## Introduction

In mammals, the transcriptional silencing of one of the two X chromosomes in female cells (X chromosome inactivation, XCI) ensures sex chromosome dosage compensation [1]. Once acquired early in development, the inactivated state is faithfully inherited through successive cell divisions. XCI initiation is associated with increased Xist RNA transcription. Whilst first retained near its transcription site, Xist RNA then spreads along the entire X chromosome from which it has been transcribed [2–5] whilst, a series of epigenetic marks, which include the repressive histone modifications H3K27me<sub>3</sub>, H3K9me<sub>3</sub>, are recruited to the presumptive inactive X chromosome. Xist RNA is a long non-

coding RNA (17 kb in length in the mouse), which is capped, spliced, and polyadenylated. Little is known about its structure and mechanism of action.

The Xist gene has a complex origin. It includes degenerated pieces of an ancient protein gene *Lnx3* as well as genomic repeat elements derived at least in part from transposon integration events [6,7]. The most conserved Xist RNA regions correspond to repeat elements (denoted A to E in mouse [8]), which are organized as tandem arrays. The A region (positions 292 to 713 in mouse, accession no. gi|37704378|ref|NR\_001463.2| [2], and 350 to 770 in human, accession no. gi|340393|gb|M97168.1| [5]) is the most highly conserved of the repeat regions and is critical for initiation of XCI. The observation that female mouse

## Author Summary

In placental mammal females, Xist RNA is crucial for inactivation of one of the two X chromosomes in order to maintain proper X chromosome dosage. It is known that the conserved A region of Xist RNA, which contains eight or nine repeated elements, plays an essential role in this process, however, little is known about its structure and mechanism of action. By using chemical and enzymatic probes, as well as FRET experiments, we performed the first experimental analysis of the solution structure of the entire Xist A region. Both mouse and human A regions were found to form two long stem-loop structures each containing four repeats. In contrast to previous predictions, interactions take place both between repeats and between repeats and spacers. Affinity-purification of RNA-protein complexes formed by incubation of RNA in mouse ES cell nuclear extract, followed by mass spectrometry and antibody-based analyses of their protein contents, showed that the isolated 4-repeat structures from the A region can recruit components of the PRC2 complex that is needed for X chromosome inactivation. However, association of one component of this complex, Suz12, was more efficient when the entire A region was used.

embryos carrying a mutated Xist<sup>AA</sup> gene inherited from males are selectively lost during embryogenesis underlines the importance of this element [9]. Recent data have shown that an early event in silencing is the formation of a Xist RNA compartment and that the A region whilst not necessary for formation of this compartment is needed for relocation of X linked genes into this territory [10]. Over-expression of a Xist<sup>AA</sup> RNA in transgenic mouse ES cells indicates that the A region whilst not necessary for Xist coating is implicated in the recruitment of the PRC2 complex [11–16]. The PRC2 complex contains the Suz12, Eed, Ezh2, and Rbap46–48 proteins [17,18]. Eed and Suz12 have been proposed to bind nucleic acids [19,20], whereas Rbap46–48 may interact with nucleosome protein components [17]. Lysine 27 trimethylation of histone H3 is catalysed by Ezh2 [12,14] and both Eed and Suz12 are required for this activity [20]. Recently a short 1,600 nucleotide-long RNA which contains the A region at its 5' extremity was suggested to be expressed early in XCI initiation and to bind the PRC2 complex [21].

Since the function of Xist RNA is expected to depend on its 2-D structure, studies aimed at establishing the 2-D structure of the Xist A region have considerable interest. Based on nucleotide sequence of the A region and computer prediction, Wutz and colleagues have proposed that each repeat forms two short stem-loop structures [11]. Recent NMR analysis confirmed that one of these stem-loop structures can be formed in vitro by an RNA molecule bearing a single copy of the mouse repeat A sequences [22]. In Xist RNA, the repeat sequences are, however, separated by long spacer regions (21 to 48 nt long for mouse). Since current models fail to take account of this sequence complexity, an experimental analysis of the entire A region was thought likely to provide valuable information on the structure of the A region. As conventional probing experiments are, however, hindered by the presence of the repeated sequences and long U tracks, we applied a combined approach exploiting both chemical and enzymatic probing of RNA structure in solution and FRET experiments using fluorescent oligonucleotide probes complementary to different parts of the A region.

Using this dual approach, we could show that repeats in the A region interact with each other to form long irregular stem-loop

structures. Such inter-repeat interactions appear to be required for the binding of the various components of the PRC2 complex. We identified the minimal number of repeats necessary for such binding. The implications of our results within the wider context of X-inactivation and of the XCI mechanism(s) underlying silencing are discussed.

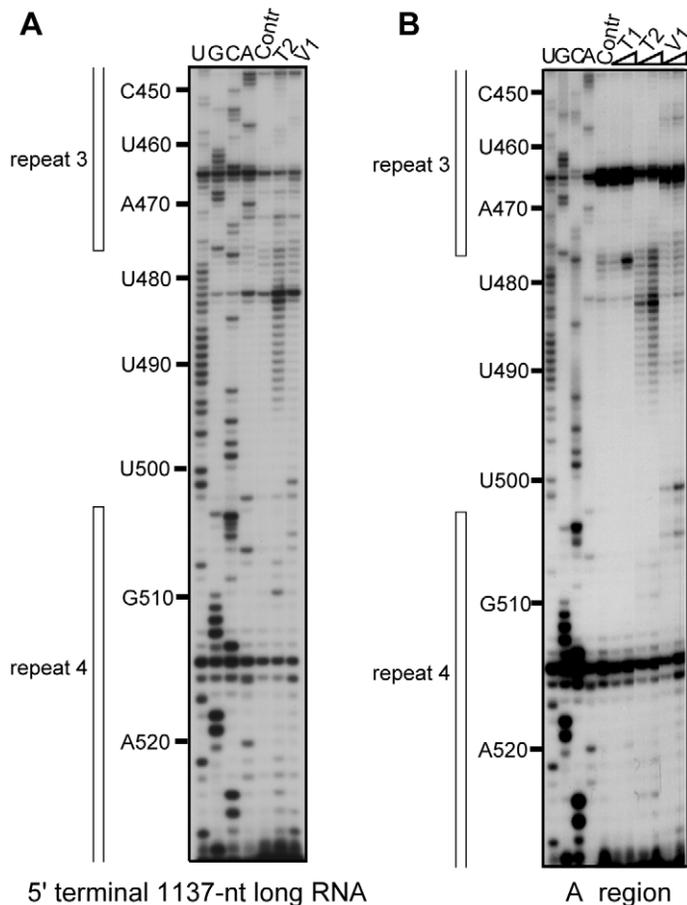
## Results

### Probing of Mouse and Human A Region 2-D Structures

We analysed in parallel both the entire mouse and human A regions, as the sequence divergence of the inter-repeat linking sequence between mouse and human was expected to provide insight into how the spacer regions might influence the A repeat structure. The specific primers for the two A regions are listed in Table S1. To test if the A region interacts with neighbouring Xist RNA sequences, an RNA that contained only the mouse A region (positions 277 to 760 in mouse Xist RNA) and a larger RNA including the sequence extending from positions 1 to 1137 were studied in parallel by limited enzymatic digestion. Very similar digestion profiles (Figure 1A and B) were obtained for the two RNAs when digestions were performed on the T7 RNA transcripts after folding under the conditions outlined in Materials and Methods. We conclude that the A region probably folds on itself without major interaction with other upstream and downstream Xist sequences. Hence, our subsequent analyses of the 2-D structure of the Xist A region were carried out in the absence of flanking sequences (positions 227 to 760 and 330 to 796 for the mouse and human A regions, respectively). Each enzymatic digestion and chemical modification assay was carried out in duplicate using different transcript preparations and each extension analysis repeated 2 or 3 times for each primer. Representative examples of the primer extension analyses are provided in Figure 1 and Figure S1A–S1E.

### M-Fold Assisted Modelling of the Mouse A Region 2-D Structure

The structure proposed by Wutz and colleagues, in which each of the repeats fold into a double stem-loop structure, could not explain the numerous V1 RNase cleavages that we observed with both the mouse and human A regions (Figure 2) [11]. We explored the possibility that each repeat folds into a unique longer stem-loop structure. Such folding was similarly unable to explain V1 RNase cleavages (Figure S2). We conclude that the 2-D structure may involve interactions between repeats and spacers and inter-repeat interactions. There is, however, a multitude of potential ways for duplex formation between repeats (Figures 3–5, models 1–3). Our design of the putative structure was orientated by the detection of six successive strong V1 RNase cleavages in the central poly A sequence (positions 550 to 555), suggesting the involvement of this segment in a helical structure. The strong modification by DMS of a sequence immediately downstream (positions 555 to 561) was an indication for a single-stranded state. One possible explanation for these data was the formation of a central stem-loop structure called SLS2M with a U track on one strand and an A track on the other strand (Figure 5). Formation of this central stem-loop structure was subsequently imposed as a constraint when exploring the possible folding of the mouse A region. This excluded structures in which two successive repeats would interact with each other (Model 1, Figure 3), since in this case, the entire poly A region would interact with the poly U track located upstream of repeat 3, which was not in agreement with the probing data (Figure 3). Another possible structure involved formation of an interaction between the 5' and 3' halves of the A region. This would generate a very long irregular



**Figure 1. Probing of mouse Xist A region RNA structure alone and inside 5'-terminal region.** The two RNAs (A. for 5' terminal 1137-nt long RNA, B. for A region) were in vitro transcribed and renatured as described in Material and Methods, before being subjected to limited digestion with T1, T2, or V1 RNases under the conditions described in Materials and Methods. Extension analyses were performed using oligonucleotide 3866 (Table S1) as the primer. The resulting cDNAs were fractionated by electrophoresis on 7% denaturing polyacrylamide gel. Lanes U, G, C, and A correspond to the sequencing ladder obtained with the same primer. Lanes marked by Contr corresponds to primer extension analysis of undigested RNA transcripts. Nucleotide numbering on the left-hand side of the autoradiogram takes the first residue of mouse Xist RNA as residue 1. The sequences corresponding to repeats 3 and 4 are indicated by vertical bars on the right-hand side of the autoradiograms.  
doi:10.1371/journal.pbio.1000276.g001

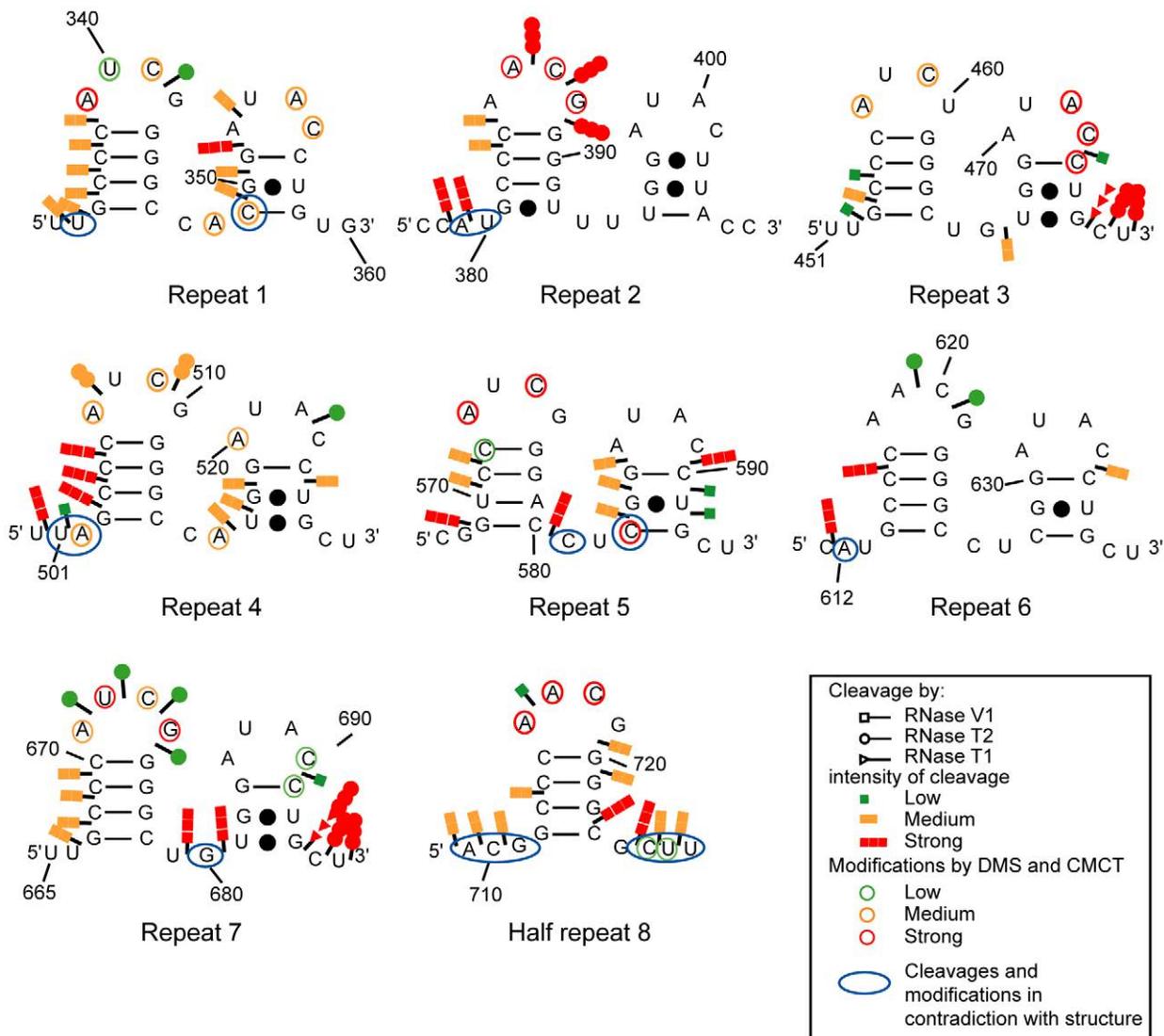
stem-loop structure with an A rich terminal loop (Model 2, Figure 4). An alternative structure involved the folding on themselves of the 5' and 3' parts of the A region, with SLS2M in between (Model 3, Figure 5). Several other alternative pairings of the repeats were also explored—none fitted the chemical and enzymatic data perfectly.

The notion of independent folding of the 5' part of the A region (positions 318 to 521 in mouse RNA) was supported by M-fold analysis of this segment, which identified a highly stable long irregular stem-loop structure, SLS1M ( $\Delta G = -41.96$  kcal/mol at 0°C in 3 M NaCl), in which repeat 1 interacts with repeat 4 and repeat 2 interacts with repeat 3. It is the most thermodynamically stable structure proposed for this 5' segment and was predicted irrespective of whether the experimental data were introduced as a constraint in the M-fold search. In SLS1M, each repeat interacts both with another repeat and with a spacer segment, increasing the stability of the overall structure. Similarly, M-fold analysis of the 3' part of the mouse A region suggested that repeat 5 interacts with repeat 8 and a spacer region and repeat 6 with repeat 7 and a spacer region. The resulting SLS3M structure was predicted as the most favourable structure by M-fold when the experimental data were incorporated as a constraint. The overall predicted three stem-loop structure (Model 3) has a low calculated free energy

( $-77.76$  kcal/mol) and has a better fit to the experimental data than Models 1 and 2 (Figures 3 and 4), suggesting that, in solution, Model 3 is the most likely structure among the numerous possibilities.

#### The Possibility to Form Structure 3 Is Phylogenetically Conserved

If a structure has biological relevance, it is generally conserved throughout evolution. Therefore, we tested whether the most favourable structures identified for the mouse A region were relevant to the human A region in solution. The sequence of the human A region differs from that of mouse by the presence of an additional repeat 5 and the absence of a long central polyA region. Experimental data (Figures 6 and 7 and Figure S3A–E) suggested that the central repeat 5 forms a central stem-loop structure (SLS2H). Based on this, structures similar to mouse Models 2 and 3 could be proposed for the human A region which involve either a long irregular stem-loop structure including all the repeats (Model 2) or a three stem-loop structure (Model 3) with repeats 1 to 4 forming a first stem-loop structure (SLS1H), repeat 5 folded alone in a second stem-loop (SLS2H) and repeats 6 to 9 involved in a third stem-loop structure (SLS3H). As for the mouse A region:



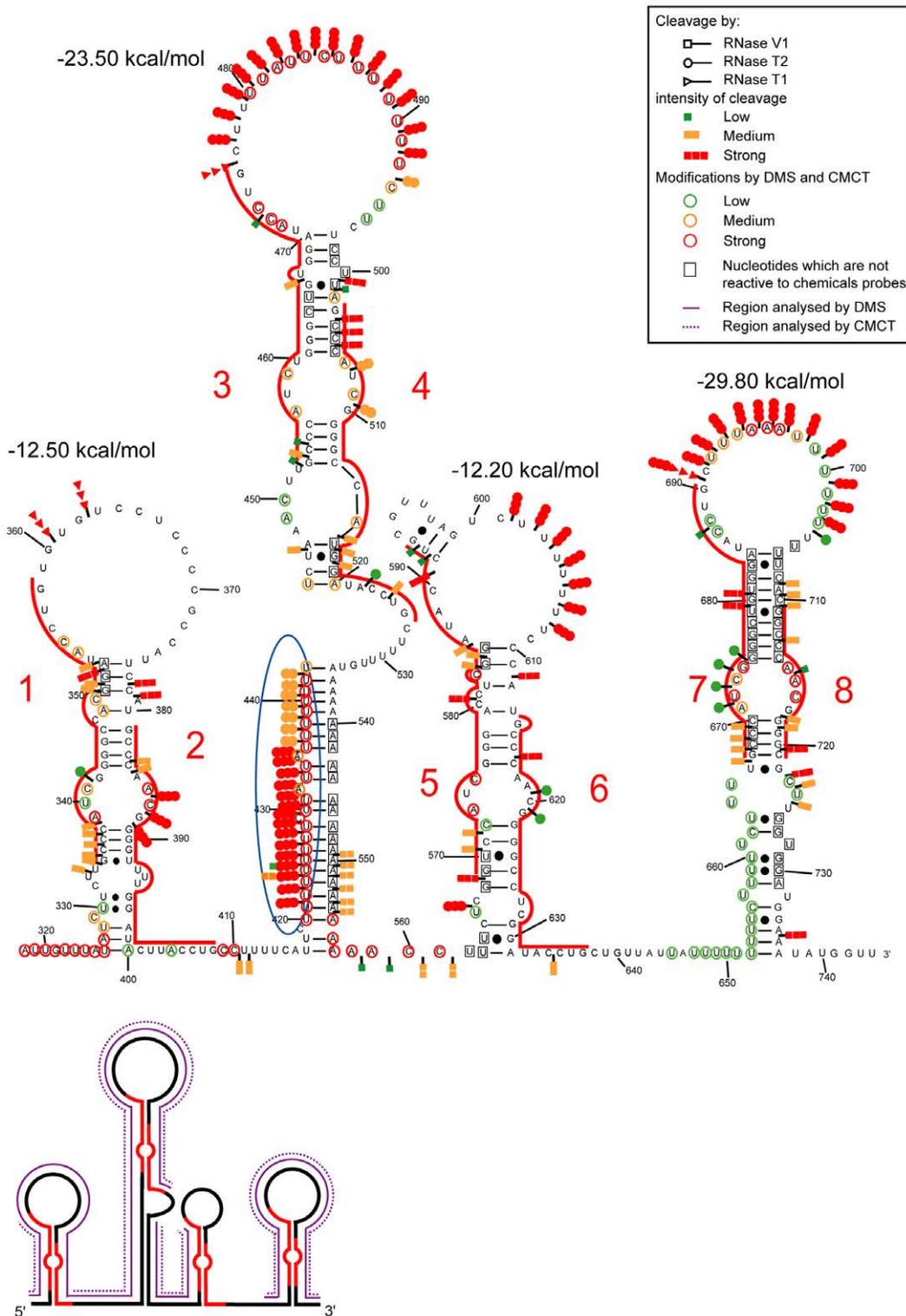
**Figure 2. Representation of experimental data on the previously proposed 2-D structure of the Xist A region.** Each of the seven repeats as well as the eighth half repeat from the mouse A region was folded according to the previously predicted two stem-loop structure [11]. T1, T2, and V1 RNase cleavages were represented by arrows surmounted by circles, triangles, and squares, respectively. Nucleotides modified by DMS or CMCT are circled. Colours of circles and arrows indicate the yields of modifications and cleavages—red, yellow, and green for strong, medium, and low modification or cleavage, respectively. The V1 RNase cleavages and chemical modifications that cannot be explained by the two stem-loop structure models are encircled by blue lanes.  
doi:10.1371/journal.pbio.1000276.g002

(i) A 2-D structure in which each repeat interacts with its immediate downstream repeat (repeats 1, 3, 6, and 8 with repeats 2, 4, 7, and 9, respectively, Model 1) was not supported by the probing data (segment 688 to 696) (Figure S4); (ii) M-fold analysis of the 5' portion of the human A region (positions 370 to 530) either with or without the experimental data as a constraint identified SLS1H as the most stable structure ( $\Delta G = -42.70$  kcal/mol) (Figure 7); (iii) SLS3H was retained as the most stable structure for the 3' part of the A region, when the experimental data were added as a constraint to an M-fold search; and (iv) the 3 stem-loop structure corresponding to Model 3 ( $\Delta G = -86.6$  kcal/mol) had the best fit with probing data compared to the other 2-D models. Further support for Model 3 was provided by our observation of identical patterns of enzymatic cleavage for the entire human A region and for the isolated SLS1H portion (Figure 6).

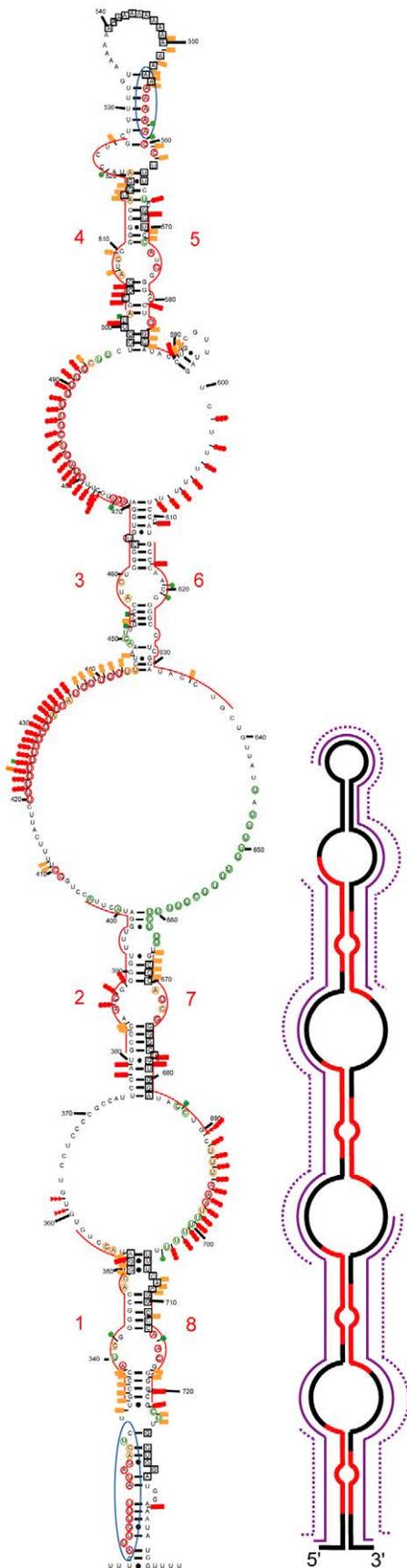
The maintenance of interactions between both spacers and repeats during mammalian evolution of the A region implies that the nucleotide sequences involved in these interactions were either conserved or subjected to compensatory base changes. This was confirmed by the alignment of the mouse, human, orangutan, baboon, lemur, dog, rabbit, cow, and elephant A region sequences (Figure S5). Nucleotide sequence conservation extends out beyond the repeats themselves for the majority of the repeats, allowing formation of the SLS1 and SLS3 structures in all sequenced Xist RNAs (Figure S6).

#### FRET Experiments Bring Additional Data in Favour of Model 3

Three oligonucleotide pairs (P1–P5, P2–P4, and P6–P7) were retained in order to test Model 3 by FRET experiments (Figure 8). This Model predicts that the P1–P5 and P2–P4 pairs of



**Figure 3. Representation of experimental data on the possible 2-D structure 1 of the Xist A region.** In model 1, the various stem-loop structures all involve two successive repeats. The repeats are indicated by red lines and are numbered from 1 to 8. Representations of chemical and enzymatic data are as in Figure 2. In each of the three panels, segments in which DMS and/or CMCT modifications were identified are indicated on a schematic drawing of the 2-D structure (full and dot lines for DMS and CMCT, respectively). In segments analyzed by the two chemical reagents, unmodified nucleotides are squared. However, one should take into consideration the fact that G residues are poorly modified by CMCT in the mild conditions that we used. The free energies of each stem-loop structures at 0°C and in 3 M NaCl were calculated with the M-fold software. doi:10.1371/journal.pbio.1000276.g003



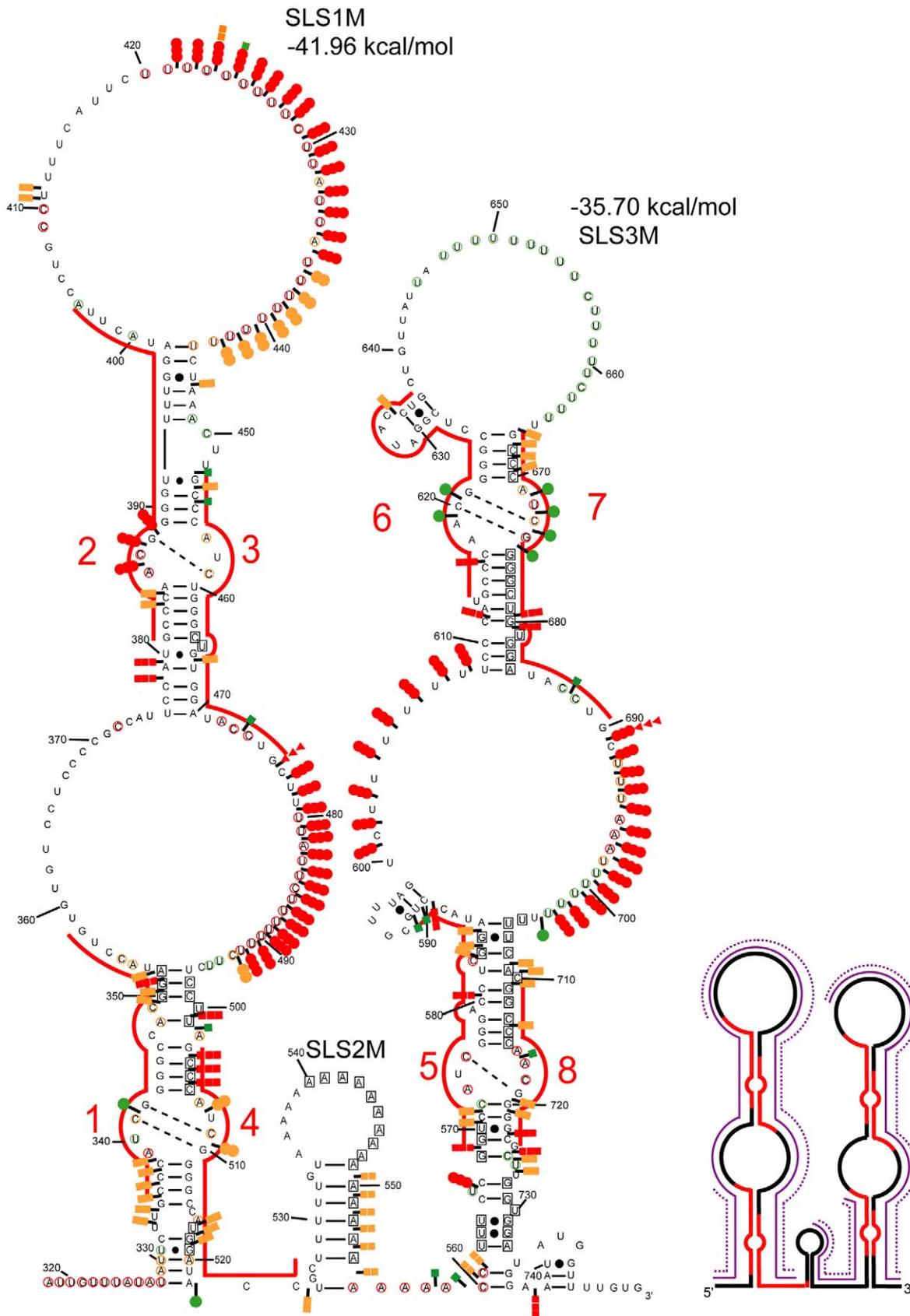
**Figure 4. Representation of experimental data on the possible 2-D structure 2 of the Xist A region.** In model 2, repeats in the 5' half of the A region interact with repeats in the 3' half of this region. Representation of enzymatic cleavages and chemical modifications are as in Figures 2 and 3.

doi:10.1371/journal.pbio.1000276.g004

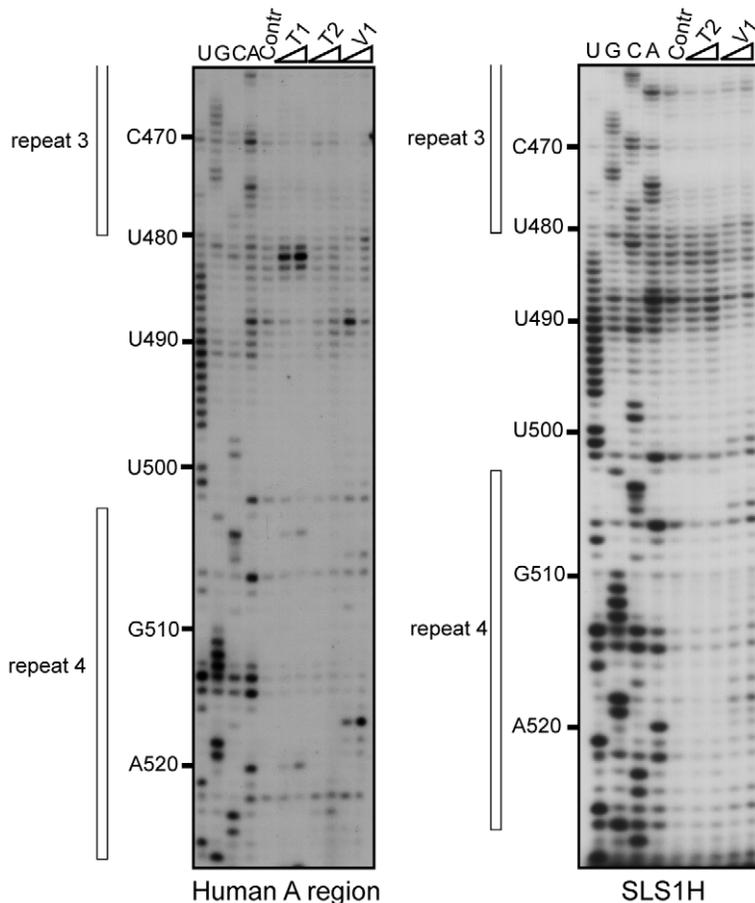
oligonucleotides interact with the single-stranded segments which border the helix formed by repeats 1 and 4, whilst the P6–P7 pair of oligonucleotides should interact with the single-stranded segments bordering the helix formed by repeats 5 and 8. A marked FRET effect would therefore be expected for these three oligonucleotide pairs if the A region was folded as in Model 3. The distance between the fluorophores of these three pairs of oligonucleotides would, on the other hand, be expected to be much larger if region A was folded as in structures 1 or 2. Whilst tertiary structural interactions might decrease the distances, a lower level of FRET would still be expected to be observed for the three pairs of oligonucleotides if the A region was folded according to structures 1 or 2 (Figure 8). The P1 and P6 oligonucleotides bind to two single-stranded segments which flank the helix formed by repeats 1 and 8 in structure 2. A strong FRET effect for P1 and P6 would therefore be expected if the A region was folded according to structure 2. Upon binding to the A region, oligonucleotide P7 partially disrupts the base-pair interactions formed by the central poly A stretch. However, as similar levels of destabilization are expected for the three possible structures, binding of this oligonucleotide was not expected to favour one structure more than the two other ones. The same is true for oligonucleotide P5 that binds to the partner U stretch of the poly A sequence. To monitor the level of FRET obtained for oligonucleotides bordering a helix, we used the short R1–2 transcript containing repeats 1 and 2 and their bordering sequences, which adopt a single unique 2-D structure and the P1–P3' oligonucleotide pair (Figure S7). Other controls exploited the P3–P6 and P3–P5 pairs, which were not expected to be in close proximity in any of the three proposed structures (Figure 8). The oligonucleotide pairs used are shown in Figure 8, along with examples of typical fluorescence intensity spectra recorded in FRET experiments for the P2–P4 and P3–P6 pairs (Figure 8D). High FRET signals in the range of 50% were obtained for the P1/P5, P2/P4, and P7/P6 oligonucleotide pairs, whilst lower FRET signals were observed for the P1–P6 (35%) and especially the P3–P6 and P3–P5 oligonucleotide pairs (25% and 22%, respectively). This is compatible with a large part of the molecules being folded in solution into structure 3. Folding of a large number of molecules into structure 2 would have led to a strong FRET signal for the P1–P6 pair and lower signals for the five other pairs, which was not observed. The strong FRET effects obtained for the P1/P5, P2/P4, and P7/P6 oligonucleotide pairs argues strongly against folding according to structure 1. Based on our FRET data, we conclude that folding predominantly occurs according to Model 3.

### Recruitment of the PRC2 Complex by the A Region

Previous studies have shown that the PRC2 complex interacts with the Xi [12,14,15,23] and the A region has been proposed to recruit the PRC2 complex through the Ezh2 subunit, which would act as an RNA-binding subunit [21]. We wished to explore further the binding of the PRC2 complex to the A region in the light of our structural data. In particular, we were interested in determining how many A region repeats were required to bind the individual Eed, Ezh2, RbAp46, RbAp48, and Suz12 components of the PRC2 complex.



**Figure 5. Representation of experimental data on the possible 2-D structure 3 of the Xist A region.** In model 3, two stem-loop structures containing four repeats are separated by a small stem-loop corresponding to poly A and poly U sequences. Representation of enzymatic cleavages and chemical modifications are as in Figures 2 and 3. doi:10.1371/journal.pbio.1000276.g005



**Figure 6. Probing of the 2-D structures of the entire and 5' half of human A region.** Legend as in Figure 1, except that the transcripts correspond to the entire human A repeat region (human A RNA) and the 5' half of the human A region (SLS1H RNA), respectively.  
doi:10.1371/journal.pbio.1000276.g006

We initiated a proteomic approach based on affinity chromatography purification of complexes formed upon incubation of *in vitro* transcribed Xist A region RNAs with nuclear extracts, followed by protein identification by mass spectrometry and Western blot analysis. Mouse ES cells are a widely exploited model for the study of XCI initiation, and we reasoned that as Xist RNA acts as an initiator of XCI, proteins which have to interact with this RNA to ensure early Xist functions should already be present in the nuclear extract of ES female mouse cells prior to differentiation. We used a control RNA containing only the three MS2 protein binding sites and tested four RNAs containing different segments of the mouse Xist A region flanked by the three MS2 binding sites at their 3' end (Figure 9A). These RNAs denoted as 1R/MS2, 2R/MS2, 4R/MS2, Aregion/MS2, and HIV/MS2 contained, respectively, repeat 1 without any neighbouring sequence, repeats 3 and 4 and their bordering spacers (positions 401 to 552 in mouse Xist RNA), the SLS1M stem-loop structure, the entire A region, and a fragment of HIV-1 RNA (positions 5338 to 5514 in the BRU RNA) used as a negative control.

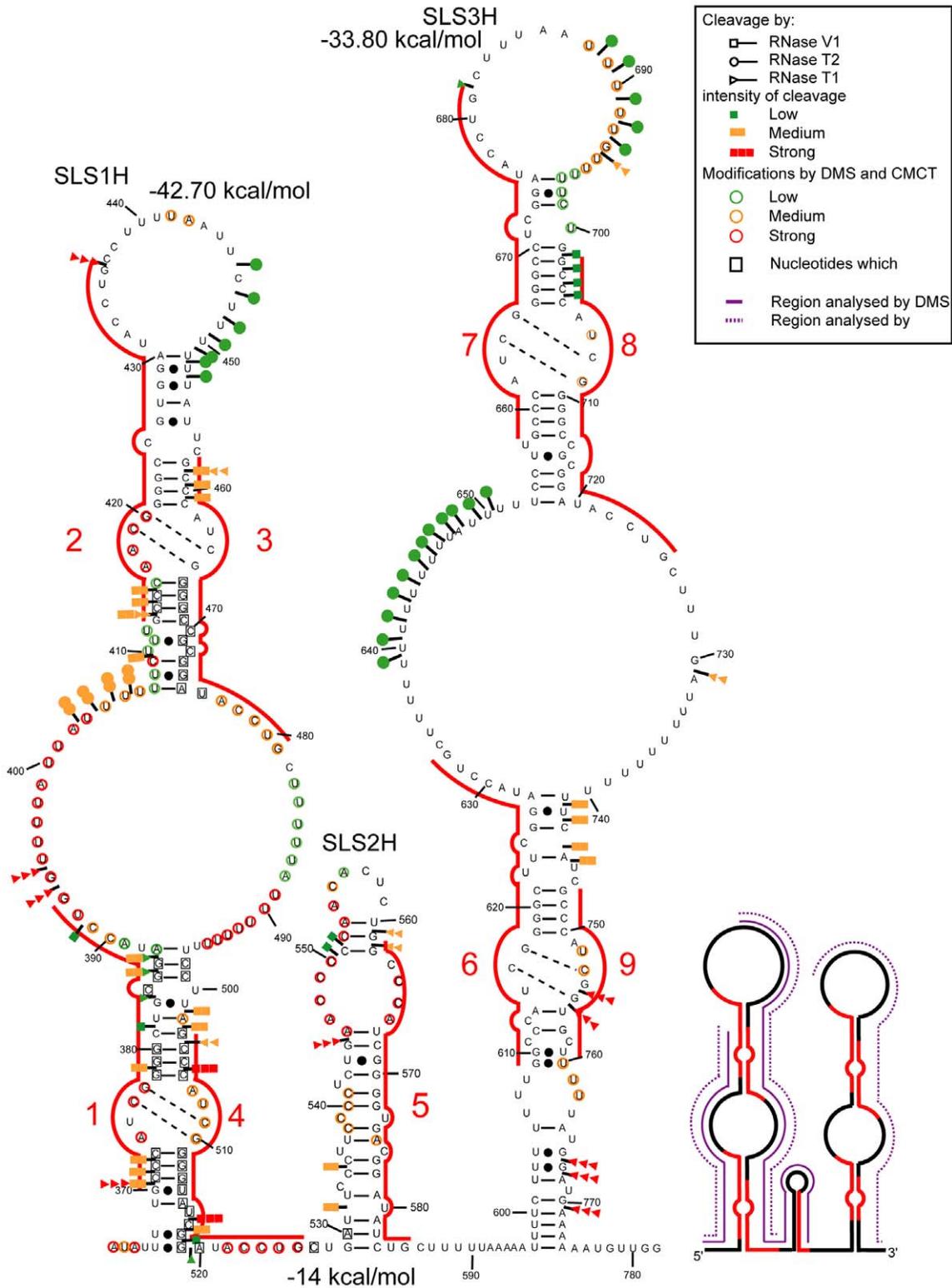
In order to get an idea of the proteins capable of associating with the entire A region, the proteins bound to purified complexes formed on the Aregion/MS2 RNA were analysed by mass spectrometry. Among numerous proteins detected were protein PTB and components of the PRC2 complex (Ezh2, RbAp46, RbAp48, and Suz12) (Figure S8). We then evaluated by Western

blot experiments the relative amounts of each of the PRC2 components in RNPs formed by the various RNAs tested. Whilst Eed, Ezh2, and PTB were detected in complexes formed on RNAs containing two or more repeats, binding of RbAp46 and RbAp48 was detected only when using RNAs with at least four repeats and Suz12 when the entire A region was used (Figure 9C). The control HIV-1 RNA bound none of these proteins (Figure 9B and 9C).

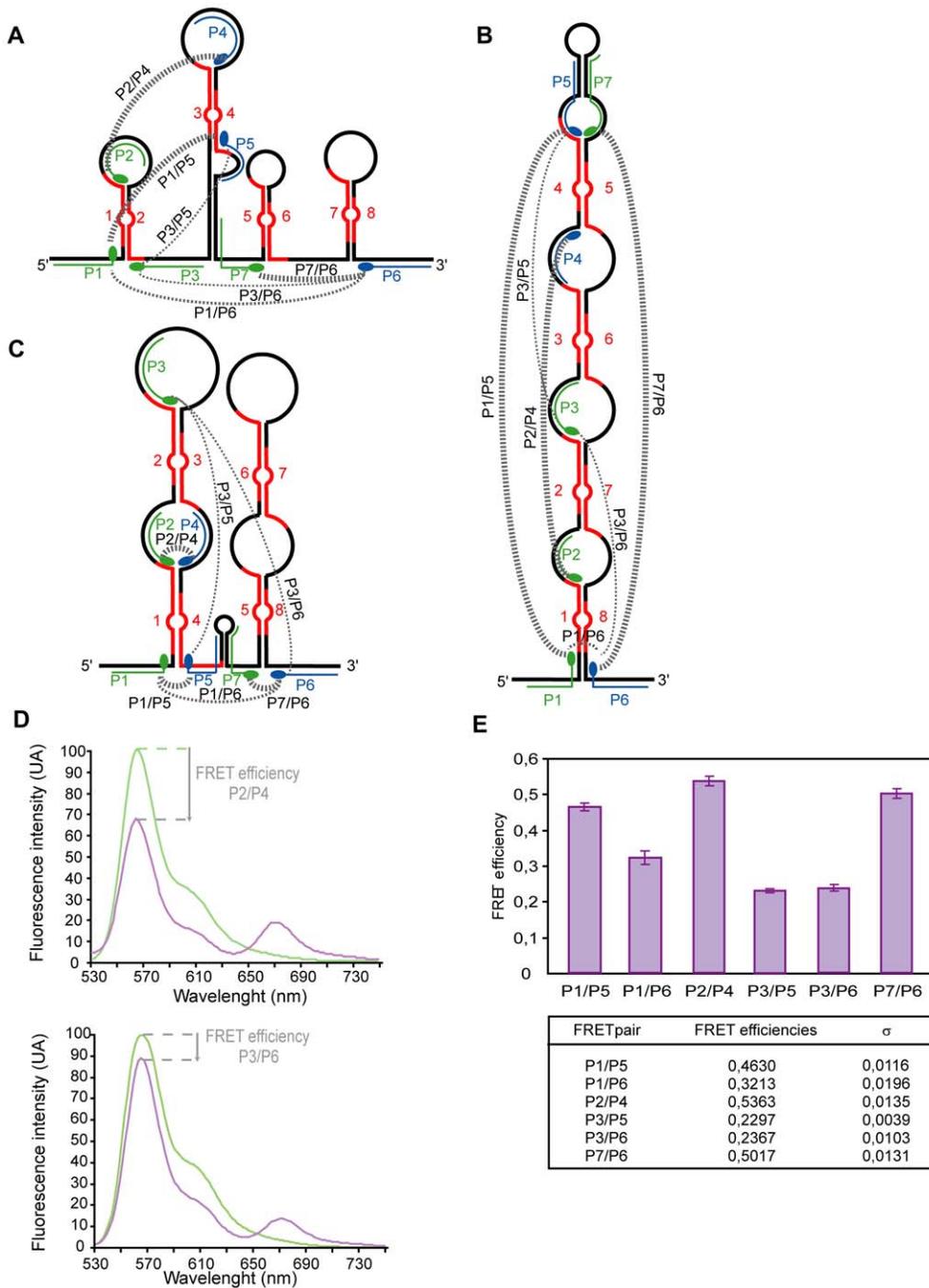
To further explore these data, we performed a series of experiments in which fragments of the A region were transcribed *in vitro* as radio-labelled RNA without MS2 fusion and these RNAs were incubated with mouse ES nuclear extracts. Three distinct RNAs (the complete A region, 4R, and 2R RNAs; Figure 9D and 9E) were used for these experiments.

In confirmation of the possible interaction of Eed with an RNA containing only two repeats, trace amounts of 2R RNA were retained on the beads when an anti-Eed antibody was used. Only complexes containing the 4R RNA or the entire A region were retained when anti-Suz12, anti-Ezh2, anti-RbAp46, and anti-RbAp48 antibodies were used. These observations confirmed the importance of the corresponding regions for association of these proteins. Clearly, however, higher amounts of the entire A region compared to 4R RNA were bound when the anti-Suz12 and Ezh2 antibodies were used.

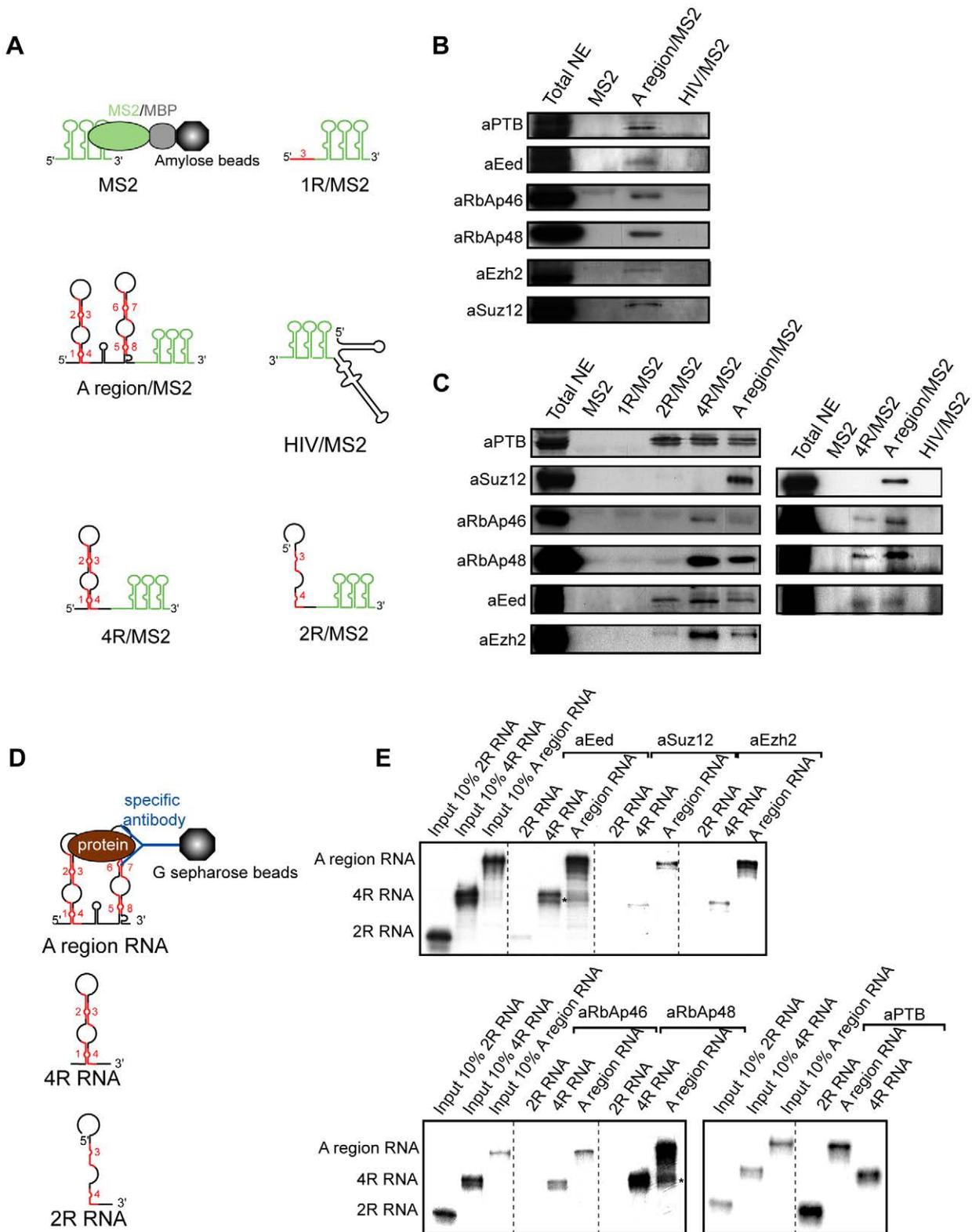
We conclude that whilst some segments of the A region allow the binding of particular PRC2 components, the entire A region is required for efficient association of the entire complex.



**Figure 7. Representation of experimental data on possible structure 3 of human Xist A region.** Repeat 5 is located in a short central stem-loop structure flanked by two larger stem-loop structures that each contain four repeats. Representation of the enzymatic cleavages and chemical modifications are as in Figures 2 and 3. Indication of segments in which DMS and/or CMCT modifications were identified is represented as in Figure 3. The free energies of the stem-loop structures were calculated with the M-fold software.  
doi:10.1371/journal.pbio.1000276.g007



**Figure 8. Steady-state fluorescence studies provide additional support to the A region Model 3.** In (A, B, and C), the binding sites of oligonucleotides P1 to P6 used in the FRET experiments are shown for the three 2-D structures of the A region corresponding to Models 1, 2, and 3. The identity of the chromophore present in each oligonucleotide (donor Cy3 or acceptor Cy5) is indicated in green and blue, respectively. Cy3- and Cy5-labeled oligonucleotides were purchased from Eurogentec. As illustrated in (D), the emission fluorescence spectra from 530 to 745 nm of the donor oligonucleotide bound alone to the RNA (green curve) were collected, as well as the emission spectra obtained in the presence of the donor and acceptor oligonucleotides (violet curve). No energy transfer between Cy3- and Cy5-labeled oligonucleotides was detected in solution. The FRET efficiency for each pair of oligonucleotides was defined as the decrease in fluorescence of the donor at 564 nm in the presence of the acceptor. Two representative examples of FRET assays (oligonucleotide pairs P2/P4 and P3/P6) are shown in (D). The FRET efficiencies measured for the six pairs of oligonucleotides are provided in (E) (mean values of three independent experiments). Standard deviations ( $\sigma$ ) are shown. The relative efficiencies of the FRET obtained for each oligonucleotide pairs are schematically represented in (A, B, and C) by lines joining the oligonucleotides. The thickness of the lines reflects the efficiency of the FRET effect.  
doi:10.1371/journal.pbio.1000276.g008



**Figure 9. The PRC2 complex assembles on fragments of the A region containing at least four repeats.** (A) Representation of the fusion RNAs used for formation of RNP complexes with ES cell nuclear extracts. Retention of RNA containing three MS2 coat protein binding sites (MS2) on amylose beads was mediated by the MS2/MBP fusion protein [31]. Analysis of the protein content of the RNP complexes formed on the A region/MS2 RNA was achieved by mass spectrometry (Figure S8). (B and C) Western blot assays using antibodies specific for the PTB, Suz12, RbAp46, RbAp48, Eed, and Ezh2 protein were used to evaluate the relative amounts of these proteins in the purified complexes formed with the various RNAs shown in (A). Antibodies were purchased from Santa Cruz (Sc), Abcam (ab), and Calbiochem (cb): anti-Ezh2 (anti-ENX-1 H-80, sc-25383), anti-RbAp46 (ab3535), anti-RbAp48 (ab488), anti-Eed (ab4469), anti-Suz12 (ab12073), and anti-PTB (cb NA63). (D and E) Test of the association of radiolabelled fragments of the A

region with components of the PRC2 complex present in nuclear extracts of ES cells. The three RNAs tested are represented in (D). RNAs bound to the G sepharose beads were fractionated by electrophoresis on 7% denaturing gels. Autoradiograms of the gels are shown in (E). Input corresponds to 10% of the material incubated with the beads.  
doi:10.1371/journal.pbio.1000276.g009

## Discussion

The A region of Xist RNA plays an essential role in the X inactivation process. Here, we show that *in vitro* the repeated elements in the A region of both mouse and human Xist RNAs interact together and with the intervening spacer regions. This leads to the formation of peculiar long irregular stem-loop structures containing four repeats and long U rich terminal and internal loops. Our proteomic analysis suggests that these four-repeat structures may correspond to functional modules initiating the assembly of the PRC2 complex.

### A New Conception of the 2-D Structure of the A Region

Until now, both computer [11] and experimental analyses [22] of the possible 2-D structure of the A region of Xist RNA have privileged the individual repeat as the unit of folding. However, the presence of long intervening spacer sequences between the repeats suggests that these spacer sequences may participate in 2-D structure formation, and points to the potential inadequacy of previous models. Our detailed chemical and enzymatic probing of the A region structure in solution involving the design of specific primers for reverse transcriptase extension analysis enabled us to identify for the first time the double-stranded and single-stranded segments making up the A region structure in solution. The data obtained clearly demonstrate that the repeats do not fold on themselves but rather fold one with the other.

Chemical and enzymatic probing of an RNA structure in solution often allows the building of a unique 2-D structure model in agreement with the experimental data. Studies on the A region were, however, complicated by the high degree of sequence redundancy. Use of a recently proposed biophysical approach, based on FRET assays [24], helped overcome these difficulties by providing information on the relative distances between the sequences flanking the various repeats. To our knowledge, up to now, this approach has only been used to confirm the 2-D structure model of telomerase RNA [24]. This method, which involves the utilization of oligonucleotides carrying donor and acceptor fluorescent dyes complementary to single-stranded segments in the studied RNA, proved particularly well suited to the study of the A region, since our probing data identified several long single-stranded segments which were able to bind the oligonucleotide probes. Among the possible 2-D structures for the A region, only one, structure 3, showed perfect agreement with the FRET data. Structure 3, which contains two long irregular stem-loop structures, each involving four repeats (four-repeat structure), also shows the best agreement with the chemical and enzymatic data. The two long stem-loop structures are separated by a short stem-loop structure, corresponding to a divergent region between mouse and human Xist RNAs. One repeat in this segment is common to all the sequenced Xist RNAs (Figure S5), except mouse RNA. In the latter, it is replaced by a poly A sequence forming a short stem loop with a poly U sequence. Interestingly, nucleotide sequence conservation in the A region extends to the spacer extremities, which contribute to the possibility of forming the four-repeat structures (Figure S5).

Although the presence of large internal and terminal loops decreases the stability of stem-loop structures, the predicted free energies of the two four-repeat structures in both mouse and human RNAs have strongly negative values (between  $-33$

and  $-45$  kcal/mol), explaining why they are proposed by the M-fold software. In addition, these four-repeat structures may be stabilized *in cellulo* by RNA-protein interactions. Interestingly, protein PTB, which contains 4 RNA recognition motifs (RRMs), which are each able to interact with UCUU(C), UUCUCU, or CUCUCU sequences, showed high affinity for the A region in nuclear extract binding experiments (Figure 9) [25]. As UCUU motifs are present on each side of the large internal loops in the four-repeat structures and in the terminal loop, one might imagine that interactions of the RRM of a single PTB molecule with these various segments may stabilize the four-repeat structure as suggested by previously proposed models for protein PTB-RNA interaction [26].

In spite that model 3 has the best fit with all the data compared to other models, this does not exclude the possibility of some local dynamic in small areas of the SLS1 and SLS3 structures. More precisely, the instability of a few base-pair interactions can explain the presence of both V1 RNase cleavages and chemical modifications in a limited number of very small segments of the A repeat region.

### Possible Functional Implication of the Four-Repeat Structure

Our adaptation of the affinity purification chromatography, originally developed for purifying spliceosome complexes [27], to complexes formed upon incubation of different fragments of the A region with nuclear extracts prepared from undifferentiated mouse ES cells, coupled with mass spectrometry and Western blot analyses, was powerful. Together with immunoselection assays performed on assembled RNP complexes, it revealed the capability of four components of the PRC2 complex to associate with an RNA corresponding to one of the four-repeat structures formed by the A region. However, our observation that the entire A region is needed for efficient association of the Suz12 protein suggests a putative additional functional role for the entire A region in either binding Suz12 or in stabilising the binding of Suz12 to the four-repeat structure. This is too early to give a convincing molecular explanation of this observation. Further experiments are needed to understand why Suz12 displays different association properties compared to other members of the PRC2 complex.

Whilst UV cross-linking of the RNP complexes formed with ES nuclear extract using the entire A region has confirmed the direct binding of PTB to this RNA region, direct binding of components of the PRC2 complex was not detected (unpublished data). Neither Ezh2 nor Eed, which were previously proposed to be recruited by Xist in an A region dependent manner [21], were cross-linked in significant amounts, suggesting that their association with the A region is mediated via association with other nuclear proteins. Therefore, the peculiar SLS1 and SLS3 structures in the A region may be needed to recruit nuclear proteins which have an affinity for components of the PRC2 complex or to reinforce the RNA affinity for these components. Mass spectrometry analysis of RNP complexes formed with the entire A region showed that, in addition to components of the PRC2 complex, a large number of other nuclear proteins can associate with this RNA region. In further studies, it will be important to identify which of these proteins are required for

PRC2 association and which ones bind directly to the A region structure.

Our finding that Suz12 requires the entire A region, or more simply more than four repeats for efficient association with the RNA, is in good agreement with the observation of Wutz and colleagues (2002) that the presence of at least 5.5 repeats is needed to initiate inactivation [11]. Additional support for the functional significance of the four-repeat model comes from a reworking of data obtained by Wutz and colleagues, who tested the effect of a series of mutations within the A region on XCI initiation [11]. Our structural studies show that all the variants (XR, XSR, XCR) classed by Wutz et al. as active are able to form the four-repeat structure, whereas the two inactive variants (XS1 and XNX) cannot (Figure S9).

Although several data argue in favour of a major role of the four-repeat structure in A repeat activity, we cannot exclude a possible role of alternative structures, for instance in modulating A repeat activity.

### Possible Implication of PRC2 and Other Nuclear Protein Association with the A Region

Although it is clear that the A region is essential for the X inactivation process, the precise role and mechanisms involved in the action of the A region remain unclear. Its deletion was shown to block silencing but not the coating of the X chromosome by Xist [11], an observation in agreement with a possible role of the A region in PRC2 recruitment. PRC2 is needed for apposition of some, but not all, of the epigenetic marks which are specific features of silenced chromatin in general and the inactive X in particular (methylation of histone H3 at position 27) [20]. The association of PRC2 with long ncRNAs before transfer of the PRC2 complex to chromatin may be a general mechanism for chromatin silencing processes that depend on long ncRNAs. Both the HOTAIR and Kcnq1ot1 long ncRNAs, which are involved in gene silencing, were recently found to bind the PRC2 complex [19,28].

Recruitment of PRC2 is a relatively early event in X inactivation [14] in agreement with a possible early association of this complex with Xist RNA prior to extensive Xist coating of chromatin. One could imagine that PRC2 is associated with the chromatin upon Xist coating through its interaction with proteins bound to Xist RNA. Alternatively coating of the Xist RNP may facilitate PRC2 transfer to chromatin by interaction of some of the RNP components with proteins of the chromatin structure. Lee and colleagues recently reported the existence of the 1600 nucleotide long RepA RNA carrying the A region at its 5' extremity, which may be expressed prior to expression of the entire Xist RNA and has been reported to recruit the PRC2 complex in a very early step of XCI [21]. Independent confirmation of these findings will be of major importance to the field.

Screening of the numerous other proteins that we have found to be capable of association with the entire A region by mass spectrometry for their eventual specific involvement in the recruitment of genes to the X inactivation domain [10] or other early events characterising the onset of X initiation and silencing will be of potential major importance to our understanding of X inactivation.

## Materials and Methods

### RNA Preparation

DNA fragments coding for the entire A regions of mouse and human Xist RNAs and their subfragments were PCR amplified

using mouse or HeLa cell genomic DNA, and cloned into plasmid pUC18 under the control of a T7 promoter. RNAs were generated by run-off transcription with T7 RNA polymerase as previously described [29]. DNA templates were digested with RNase-free DNase I and RNA transcripts were purified on denaturing 3% to 8% polyacrylamide gels.

### Enzymatic and Chemical Probing of RNA Secondary Structure

RNA 2-D structures in solution were probed as follows [29]: 200 ng of transcripts dissolved at a 80 nM concentration in buffer D (20 mM Hepes-KOH, pH 7.9, 100 mM KCl, 0.2 mM EDTA pH 8.0, 0.5 mM DTT, 0.5 mM PMSF, 20% (vol/vol) glycerol) were renatured by 10 min heating at 65°C, followed by slow cooling at room temperature with the addition of 1 µl of 62.5 mM MgCl<sub>2</sub> to a final concentration of 3.25 mM MgCl<sub>2</sub>. After a 10 min preincubation at room temperature, RNase T1 (0.02 or 0.0375 U/µl) or T2 (0.025 or 0.0375 U/µl) was added under conditions such that it cleaved single-stranded segments. V1 RNase (2.5×10<sup>-3</sup> or 5×10<sup>-3</sup> U/µl) was used to cleave double-stranded and stacked residues. DMS (1 µl of a 1/4 or 1/8 (V/V) DMS/EtOH solution) was employed to modify single-stranded A and C residues and CMCT (4 or 5 µl of a 180 mg/ml solution) to modify single-stranded U and to a lower extent G residues. Reactions were stopped as described in [29]. Cleavage and modification positions were identified by primer extension [29]. Stable secondary structures having the best fit with experimental data were identified with the Mfold software, version 8.1 [30]. Probing data were introduced as a constraint in the search.

### Steady-State Fluorescence Measurements

Fluorescence spectra were recorded at 4°C, with an excitation wavelength of 515 nm and scanning from 500 to 750 nm (excitation and emission bandwidth of 3 nm). The procedure used was derived from [24]. The RNA and Cy3-oligonucleotide were mixed at a 1:1 molar ratio in 160 µl of 150 mM NaCl, 3.25 mM MgCl<sub>2</sub>, and 15 mM Na citrate (pH 7.0) to a final concentration (0.38 µM) superior to the K<sub>d</sub>, incubated at 85°C for 5 min, and slowly cooled at room temperature for 15 min. After 4 h of incubation at 4°C, the yield of oligonucleotide association was determined by electrophoresis in a non-denaturing gel. Fluorescence in the gel was measured with a Typhoon (9410) Healthcare scanner. When a satisfying yield of association was detected, the emission of the Cy3-labeled complex was measured on a flux spectrofluorometer (SAFAS). Ten spectra were averaged. Then, the Cy5-labeled oligonucleotide was added at a 1:1 molar ratio, and incubation carried out at 4°C for 4 h. Ten spectra were recorded and the Fluorescence Resonance Energy Transfer (FRET) for the Cy3–Cy5 pair was calculated taking into account the bound/unbound ratio of Cy3-oligonucleotide. Each FRET experiment was repeated three times using different batches of transcripts.

### Purification of RNP by MS2 Selection Affinity

The entire mouse A region and several fragments were cloned 3' to a T7 promoter and 5' to the MS2 tag present in plasmid pAdML3 [27,31].

Nuclear extracts from undifferentiated female ES cells (LF2) were prepared according to [32], and dialyzed against buffer D. One hundred pmol of MS2-tagged RNAs were denatured, renatured, as described above, and incubated with a 5-fold molar excess of purified MS2-MBP fusion protein [31] at 4°C for 15 min. The RNA-MS2-MBP complexes formed were incubated

with amylose beads (40  $\mu$ l, GE Healthcare) equilibrated in buffer D for 2 h at 4°C. After three washes with 500  $\mu$ l of buffer D, 1 mg of nuclear extract in 150  $\mu$ l of buffer D containing 5  $\mu$ M of yeast tRNAs was added. After 15 min of incubation at 4°C with constant agitation, three successive washes were performed in Buffer D and RNP complexes eluted by incubation with 80  $\mu$ l of Buffer D containing 10 mM maltose (30 min at 4°C). Half of the eluted RNP complex formed with the entire A region was fractionated by 10% SDS-PAGE for mass spectrometry analyses. For all the purified RNP complexes, 10% of the eluted material was used for Western blot analysis performed according to [33].

### Mass Spectrometry Analysis

Each lane of the SDS-PAGE was cut into 2 mm sections, and proteins submitted to in-gel trypsin digestion. Analysis of extracted peptides was performed using nano-LC-MS-MS on a CapLC capillary LC system coupled to a QTOF2 mass spectrometer (Waters) according to standard protocols (Figure S8). The MS/MS data were analyzed using the MASCOT 2.2.0. algorithm (Matrix Science) for search against an in-house generated protein database composed of protein sequences of *Rattus* and *Mus* downloaded from UniprotKB <http://beta.uniprot.org/> (August 07, 2008) and protein sequences of known contaminant proteins such as porcine trypsin and human keratins concatenated with reversed copies of all sequences. Spectra were searched with a mass tolerance of 0.3 Da for MS and MS/MS data, allowing a maximum of 1 missed cleavage with trypsin and with carbamidomethylation of cysteines, oxidation of methionines, and N-acetyl protein specified as variable modifications. Protein identifications were validated when one peptide had a Mascot ion score above 35. Evaluations were performed using the peptide validation software Scaffold (proteome Software).

### Immunoselection of RNP

RNA transcripts were dephosphorylated, 5'-end labelled with [ $\gamma$ -<sup>32</sup>P]ATP (3,000 Ci/mmol), purified and quantified according to [34]. About 70 pmol of the RNA were denatured, renatured as described above, and incubated with 30  $\mu$ g of nuclear extract for 30 min at room temperature with constant agitation. About 40  $\mu$ l of Protein G-sepharose beads suspension blocked with BSA (2  $\mu$ g) and coated for 2 h at 4°C with 10  $\mu$ l of each antibodies were incubated with the RNP complexes for 2 h at 4°C in 300  $\mu$ l of immunoselection buffer (150 mM NaCl, 10 mM Tris-HCl, pH 8.0, NP40 0.1%). Beads were washed three times for 10 min at 4°C with 750  $\mu$ l of immunoselection buffer containing 0.5% NP40. RNAs were phenol extracted, ethanol precipitated, fractionated on 7% polyacrylamide gel, and analysed by autoradiography.

### Supporting Information

**Figure S1 Identification of enzymatic cleavage and chemical modification of the mouse A region by primer extension.** The A region of mouse Xist RNA was in vitro transcribed and renatured as described in Material and Methods, before being submitted to limited digestion with the T1, T2, or V1 RNases under the conditions described in Materials and Methods. Primer extension analyses were performed using oligonucleotides 3760 (A), 3973 (B), 3866 (C), 3758 (D), 3971 (E), or 3757 (F) (Table S1) as primers. The resulting cDNAs were fractionated by electrophoresis on 7% denaturing polyacrylamide gels. Lanes U, G, C, and A correspond to the sequencing ladder obtained with the corresponding primers. Lanes marked by Contr correspond to primer extension analysis of undigested RNA. Nucleotide numbering on the left side of the autoradiograms is calculated

taking the first residue of mouse Xist RNA as residue 1. The position of the repeats is indicated by vertical bars on the right-hand side of the autoradiograms. Two different analyses of CMCT modifications by primer extension with oligonucleotide 3757 are illustrated in (F). The autoradiogram on the right side of the panel was exposed for a longer time.

Found at: doi:10.1371/journal.pbio.1000276.s001 (10.09 MB TIF)

**Figure S2 Representation of experimental data on a 2-D structure in which repeats form individual stem-loop structures.** Each of the seven repeats as well as the eighth half repeat in the mouse A region were folded into a unique stem-loop structure with an internal loop. T1, T2, and V1 RNase cleavages are represented by arrows surmounted by circles, triangles, and squares, respectively. Nucleotides modified by DMS or CMCT are circled. The colours of circles and arrows indicate the modification and cleavage yields, with red, yellow, and green corresponding, respectively, to strong, medium, and low modification or cleavage. The V1 RNase cleavages and chemical modifications that cannot be explained by this secondary structure model are circled in blue. Found at: doi:10.1371/journal.pbio.1000276.s002 (0.73 MB TIF)

**Figure S3 Identification of enzymatic cleavage and chemical modification in human A region by primer extension analysis.** The A region of human XIST RNA was treated as described for the mouse A region in the legend to Figure 1 of supporting data, except that the primers used for extension analyses were oligonucleotides 4563 (A), 4564 (B), 4622 (C), 4565 (D), and 4242 (E) (Table S1). Found at: doi:10.1371/journal.pbio.1000276.s003 (7.04 MB TIF)

**Figure S4 Representation of experimental data on the possible structure 1 of human Xist A region.** In this model, stem-loop structures involve two successive repeats. The repeats are indicated by red lines and are numbered from 1 to 9. Representation of chemical and enzymatic data is as in Figures 2 and 3. The free energies of each stem-loop structure at 0°C and in 3 M NaCl were calculated with the M-fold software. Found at: doi:10.1371/journal.pbio.1000276.s004 (0.74 MB TIF)

**Figure S5 Conservation of sequences surrounding the repeats in vertebrate A regions.** Sequence alignment of the mouse, human, Orangutan, baboon, lemur, dog, rabbit, cow, horse, and elephant A regions illustrating the degree of species conservation. Identical nts are indicated in red. Repeats are numbered from 1 to 9 and shown as red rectangles, mouse (gi|37704378|ref|NR\_001463.2|), human (gi|340393|gb|M97168.1|), Orangutan (by <http://www.ensembl.org/index.html>, 292L3-1,185272), baboon (by <http://www.ensembl.org/index.html>, 157F22-1,190936), lemur (by <http://www.ensembl.org/index.html>, 176F24-1,134555), dog [by [http://genome.ucsc.edu/\(canFam2\)assembly/canFam1\\_dna](http://genome.ucsc.edu/(canFam2)assembly/canFam1_dna) range = chrX:60100000–60735000], rabbit (gi|1575009|gb|U50910.1|OCU50910), cow (gi|10181229|gb|AF104906.5|), horse (gi|1575005|gb|U50911.1|), and elephant (BROADE1:scaffold\_119260:3220:3899:–1 ENSEMBL). Found at: doi:10.1371/journal.pbio.1000276.s005 (0.20 MB DOC)

**Figure S6 The possibility to form four-repeats stem-loop (SLS1 and SLS3) structure is conserved in vertebrates.** SLS1 and SLS3 in mouse, dog, human, rabbit, and elephant are folded according to the mouse SLS1 structure (Model 3). The name of each species is indicated below the structure. Sequence variations compared to the mouse A sequence are indicated in green. Found at: doi:10.1371/journal.pbio.1000276.s006 (1.21 MB TIF)

**Figure S7 Control FRET experiment performed with two oligonucleotides bordering one helix.** (A) Schematic presentation of the transcripts used in the control experiment. (B) Fluorescence spectra obtained with the donor P1 oligonucleotide bound to naked 2R/RNA (green curve) and with oligonucleotides P1/P3' bound to the RNA (violet curve). See legend in Figure 8 for details.

Found at: doi:10.1371/journal.pbio.1000276.s007 (0.18 MB TIF)

**Figure S8 Identification of components of the PRC2 complex by mass spectrometry.** Peptides that served for each protein identification are summarized in a table with corresponding MS/MS spectra. (A) Identification of Enhancer of zeste homolog 2 (Ezh2). (B) Identification of Polycomb protein Suz12 (Suz12). (C) Identification of Retinoblastoma binding protein 4 (RbAp46). (D) Identification of Retinoblastoma binding protein 7 (RbAp48). (E) Detailed standard protocols for proteomic analyses. Found at: doi:10.1371/journal.pbio.1000276.s008 (3.01 MB DOC)

**Figure S9 Folding capacities of the synthetic A region sequences whose activity was evaluated in [11].** Sequence XR corresponds to the positive control, XSR: replacement of GCCCAUCGCGGG by CGGGAUCGGCCC; XCR have small U-rich spacer regions. XS1: deletion of the GG dinucleotides in the second element of each repetition, XNX: replacement of GGGCAUCGGGGC by GCGCAUCGGAGC. Silencing properties of Xist RNA containing these synthetic variants A region are indicated in the right-hand side panel of the Table S1.

## References

- Lyon MF (1961) Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* 190: 372–373.
- Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, et al. (1992) The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* 71: 515–526.
- Borsani G, Tonlorenzi R, Simmler MC, Dandolo L, Arnaud D, et al. (1991) Characterization of a murine gene expressed from the inactive X chromosome. *Nature* 351: 325–329.
- Cohen HR, Panning B (2007) XIST RNA exhibits nuclear retention and exhibits reduced association with the export factor TAP/NXF1. *Chromosoma* 116: 373–383.
- Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, et al. (1992) The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71: 527–542.
- Duret L, Chureau C, Samain S, Weissenbach J, Avner P (2006) The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* 312: 1653–1655.
- Shevchenko AI, Zakharova IS, Elisaphenko EA, Kolesnikov NN, Whitehead S, et al. (2007) Genes flanking Xist in mouse and human are separated on the X chromosome in American marsupials. *Chromosome Res* 15: 127–136.
- Brockdorff N (2002) X-chromosome inactivation: closing in on proteins that bind Xist RNA. *Trends Genet* 18: 352–358.
- Hoki Y, Kimura N, Kanbayashi M, Amakawa Y, Ohhata T, et al. (2009) A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. *Development* 136: 139–146.
- Chaumeil J, Le Baccon P, Wutz A, Heard E (2006) A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev* 20: 2223–2237.
- Wutz A, Rasmussen TP, Jaenisch R (2002) Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet* 30: 167–174.
- Silva J, Mak W, Zvetkova I, Appanah R, Nesterova TB, et al. (2003) Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell* 4: 481–495.
- Plath K, Talbot D, Hamer KM, Otte AP, Yang TP, et al. (2004) Developmentally regulated alterations in Polycomb repressive complex 1 proteins on the inactive X chromosome. *J Cell Biol* 167: 1025–1035.
- Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, et al. (2003) Role of histone H3 lysine 27 methylation in X inactivation. *Science* 300: 131–135.
- Mak W, Baxter J, Silva J, Newall AE, Otte AP, et al. (2002) Mitotically stable association of polycomb group proteins eed and enx1 with the inactive x chromosome in trophoblast stem cells. *Curr Biol* 12: 1016–1020.
- Kohlmaier A, Savarese F, Lachner M, Martens J, Jenuwein T, et al. (2004) A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol* 2: E171. doi:10.1371/journal.pbio.0020171.
- Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, et al. (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298: 1039–1043.
- Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D (2002) Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* 16: 2893–2905.
- Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, et al. (2007) Functional demarcation of active and silent chromatin domains in HOX loci by noncoding RNAs. *Cell* 129: 1311–1323.
- Cao R, Zhang Y (2004) The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr Opin Genet Dev* 14: 155–164.
- Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322: 750–756.
- Duszczak MM, Zanier K, Sattler M (2008) A NMR strategy to unambiguously distinguish nucleic acid hairpin and duplex conformations applied to a Xist RNA A-repeat. *Nucleic Acids Res* 36: 7068–7077.
- Kalantry S, Mills KC, Yee D, Otte AP, Panning B, et al. (2006) The Polycomb group protein Eed protects the inactive X-chromosome from differentiation-induced reactivation. *Nat Cell Biol* 8: 195–202.
- Gavory G, Symmons MF, Krishnan Ghosh Y, Klenerman D, Balasubramanian S (2006) Structural analysis of the catalytic core of human telomerase RNA by FRET and molecular modeling. *Biochemistry* 45: 13304–13311.
- Perez I, McAfee JG, Patton JG (1997) Multiple RRM domains contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. *Biochemistry* 36: 11881–11890.
- Auweter SD, Allain FH (2008) Structure-function relationships of the polypyrimidine tract binding protein. *Cell Mol Life Sci* 65: 516–527.
- Deckert J, Hartmuth K, Boehringer D, Behzadnia N, Will CL, et al. (2006) Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol* 26: 5528–5543.
- Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, et al. (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 32: 232–246.
- Mougin A, Gregoire A, Banroques J, Segault V, Fournier R, et al. (1996) Secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency. *Rna* 2: 1079–1093.

Found at: doi:10.1371/journal.pbio.1000276.s009 (0.54 MB TIF)

**Table S1 Oligonucleotides used in this study.** The name, sequence, and utilization of each oligonucleotide are given. Nucleotide positions of A region are numbered according to the genBank accession no. gi|37704378|ref|NR\_001463.2| (Mouse Xist gene) [2] and no. gi|340393|gb|M97168.1| (Human XIST gene) [5]. Restriction sites and fluorescent dyes introduced by the oligonucleotides are indicated.

Found at: doi:10.1371/journal.pbio.1000276.s010 (0.05 MB DOC)

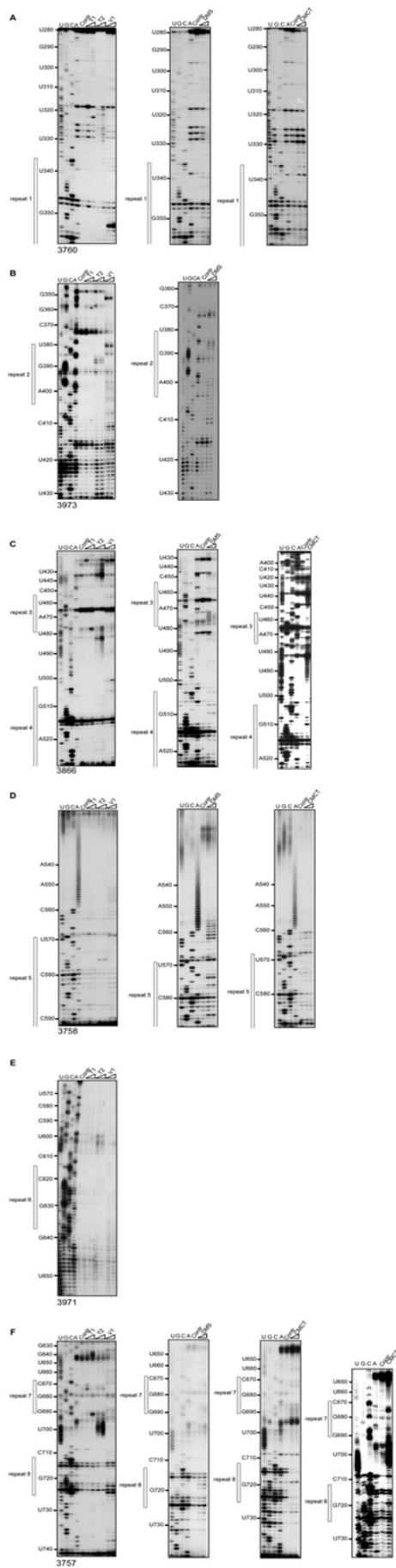
## Acknowledgments

Professor R. Lührmann (Max Planck Institut, Goettingen) is thanked for his generous gift of plasmids pMBP-MS2 and pAdML3 and for help in the use of the MS2-MBP RNP purification approach. Professor I. Motorine and C. Aigueperse (AREMS, Nancy) are acknowledged for their advice in implanting this approach. V. Ségault (AREMS, Nancy) is thanked for advice concerning the immunopurification assays. S Mazeret (IPBS, Toulouse) is thanked for helping us to define the FRET experimental protocol.

## Author Contributions

The author(s) have made the following declarations about their contributions: Conceived and designed the experiments: SM MB VM SSC AVD PC PA AV CB. Performed the experiments: SM MB LF AS AV. Analyzed the data: SM MB LF AS VM SSC PA AV CB. Contributed reagents/materials/analysis tools: AD. Wrote the paper: SM PA AV CB.

30. Jaeger JA, Turner DH, Zuker M (1989) Improved predictions of secondary structures for RNA. *Proc Natl Acad Sci U S A* 86: 7706–7710.
31. Zhou Z, Sim J, Griffith J, Reed R (2002) Purification and electron microscopic visualization of functional human spliceosomes. *Proc Natl Acad Sci U S A* 99: 12203–12207.
32. Dignam JD, Martin PL, Shastry BS, Roeder RG (1983) Eukaryotic gene transcription with purified components. *Methods Enzymol* 101: 582–598.
33. Jacquenet S, Mereau A, Bilodeau PS, Damier L, Stoltzfus CM, et al. (2001) A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. *J Biol Chem* 276: 40464–40475.
34. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning. A laboratory manual*. New York: Cold Spring Harbor Laboratory Press.

**Figure S1**

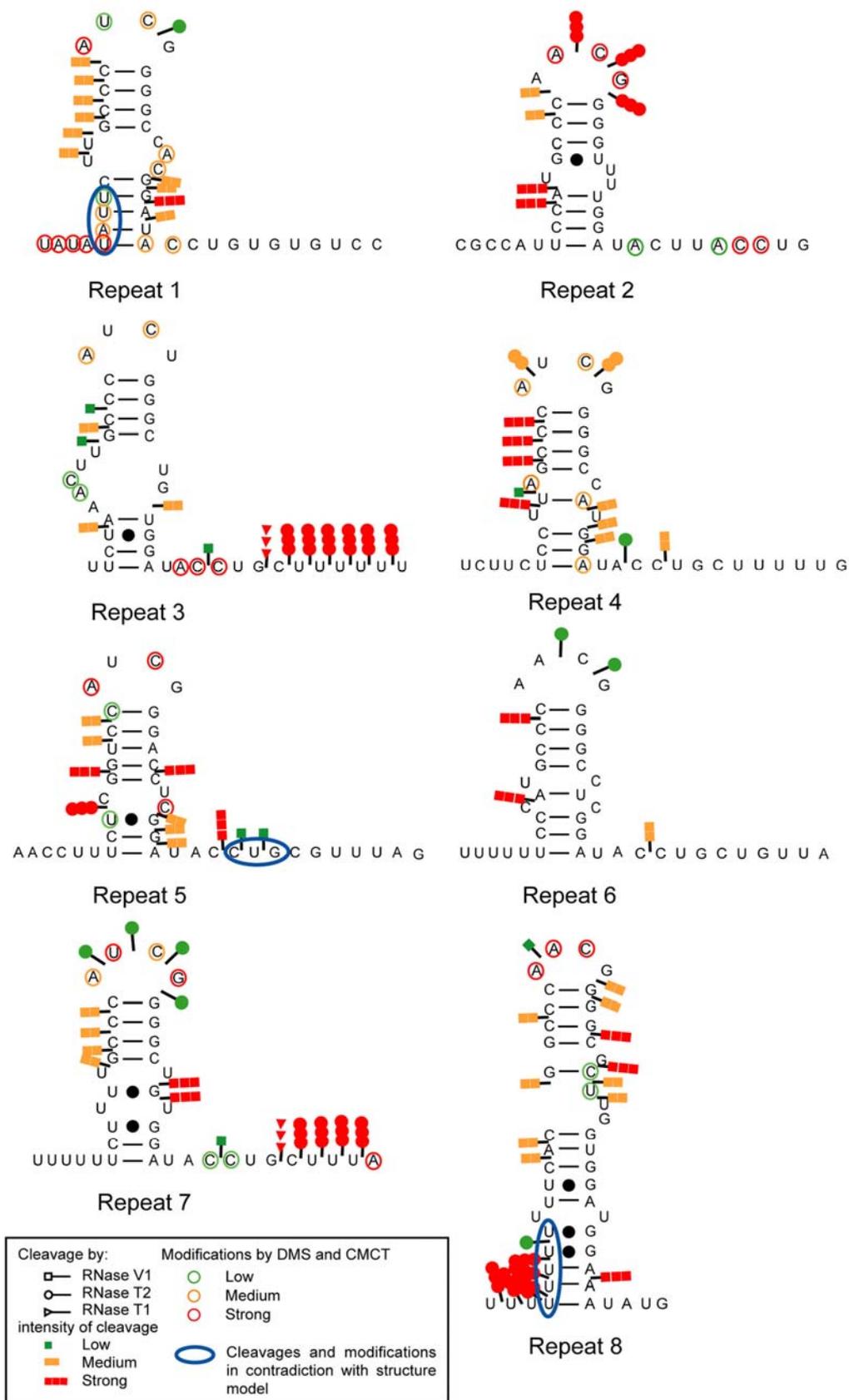


Figure S2

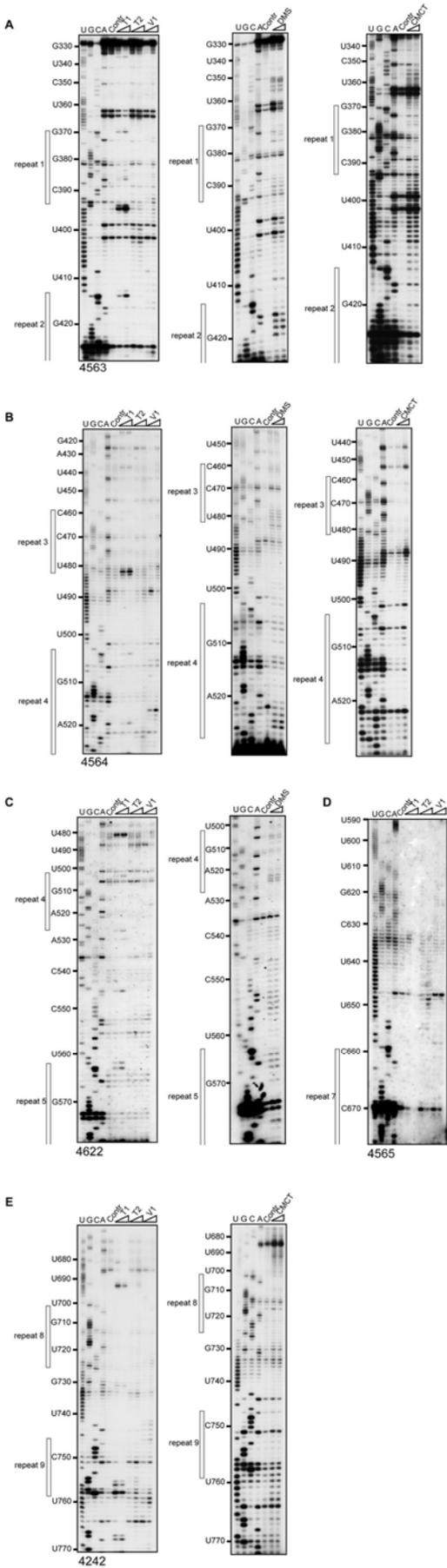
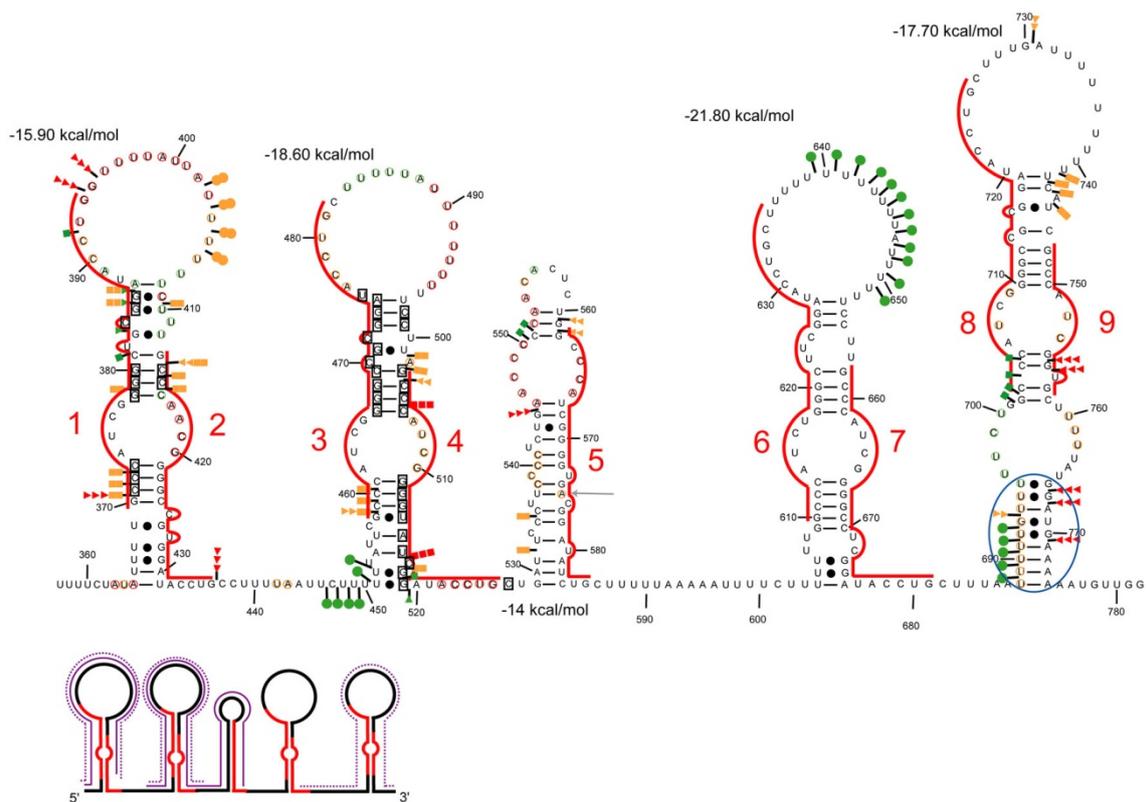


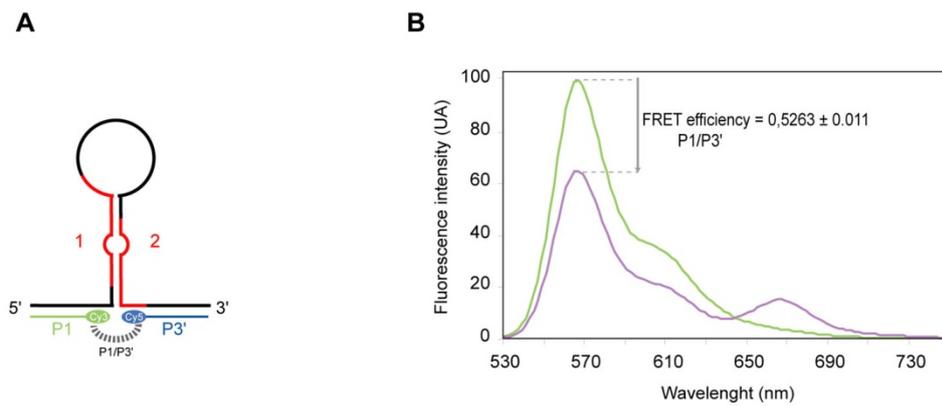
Figure S3

**Figure S4**



**Figure S5**

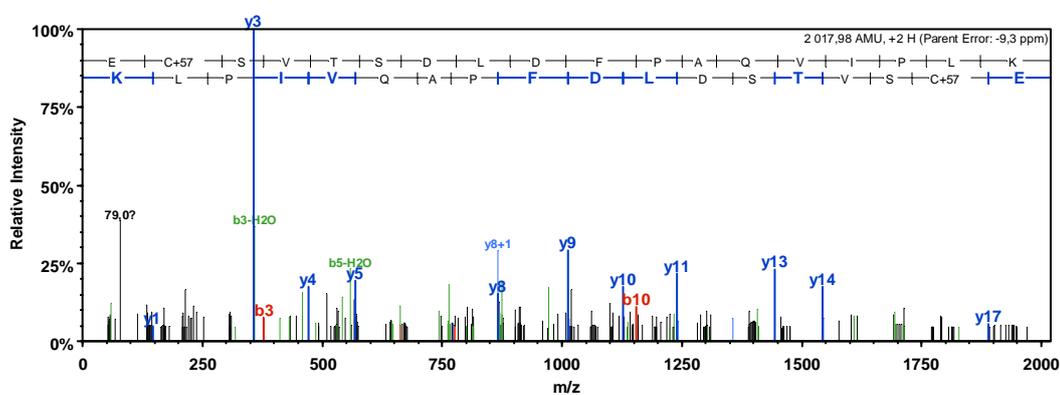
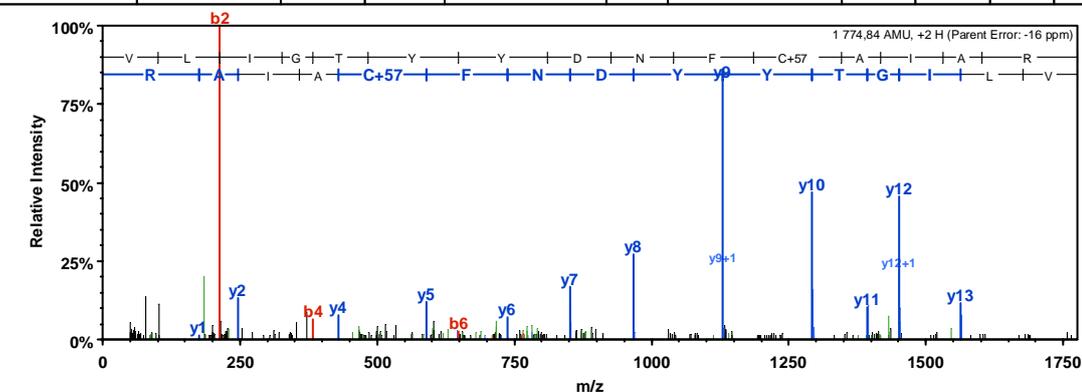




**Figure S7**

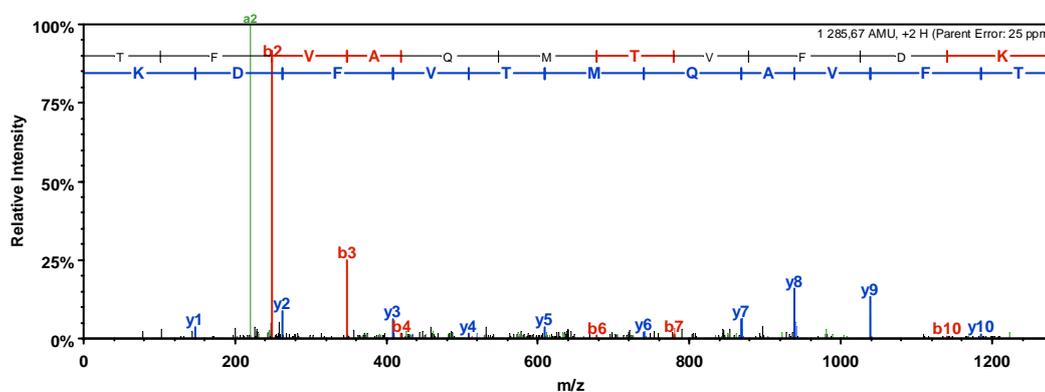
## A Identification of Enhancer of zeste homolog 2 (EZH2),

Protein name	Protein accession numbers	Protein molecular weight (Da)	Sequence coverage	Peptide sequence	Mascot Ion score	Modifications	Observed m/z	Actual peptide mass (AMU)	Spectrum charge	Error on peptide mass (PPM)
Enhancer of zeste homolog 2	Q61188 EZH2_MOUSE	85336	4.4%	VLIGTYDNFCAIAR	49.55	Carbamidomethyl	88.42	1774.84	2	-16
				ECSVTSOLDPFAQVIPLK	36.86	Carbamidomethyl	1010.00	2017.98	2	-9.3



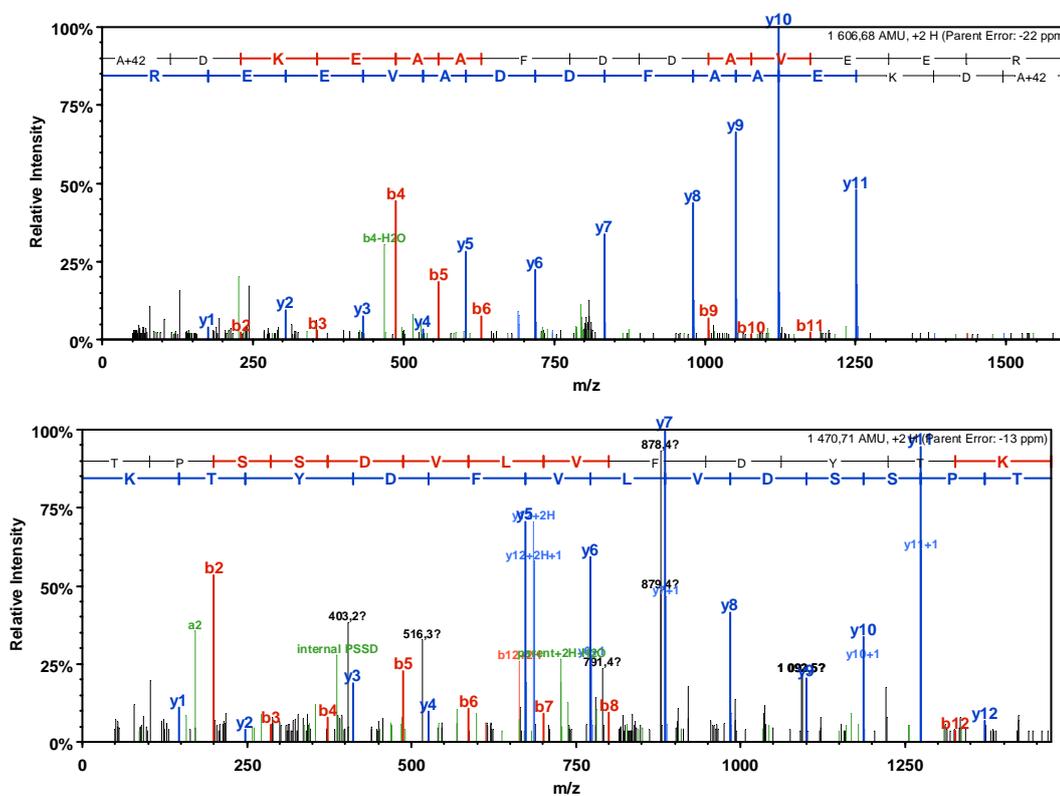
## B Identification of Polycomb protein suz12 (SUZ12),

Protein name	Protein accession numbers	Protein molecular weight (Da)	Sequence coverage	Peptide sequence	Mascot Ion score	Modifications	Observed m/z	Actual peptide mass (AMU)	Spectrum charge	Error on peptide mass (PPM)
Polycomb protein Suz12	Q80U70 SUZ12_MOUSE	83026	1.5%	TFVAQMTVFDK	35.58	/	643.84	1285.67	2	25



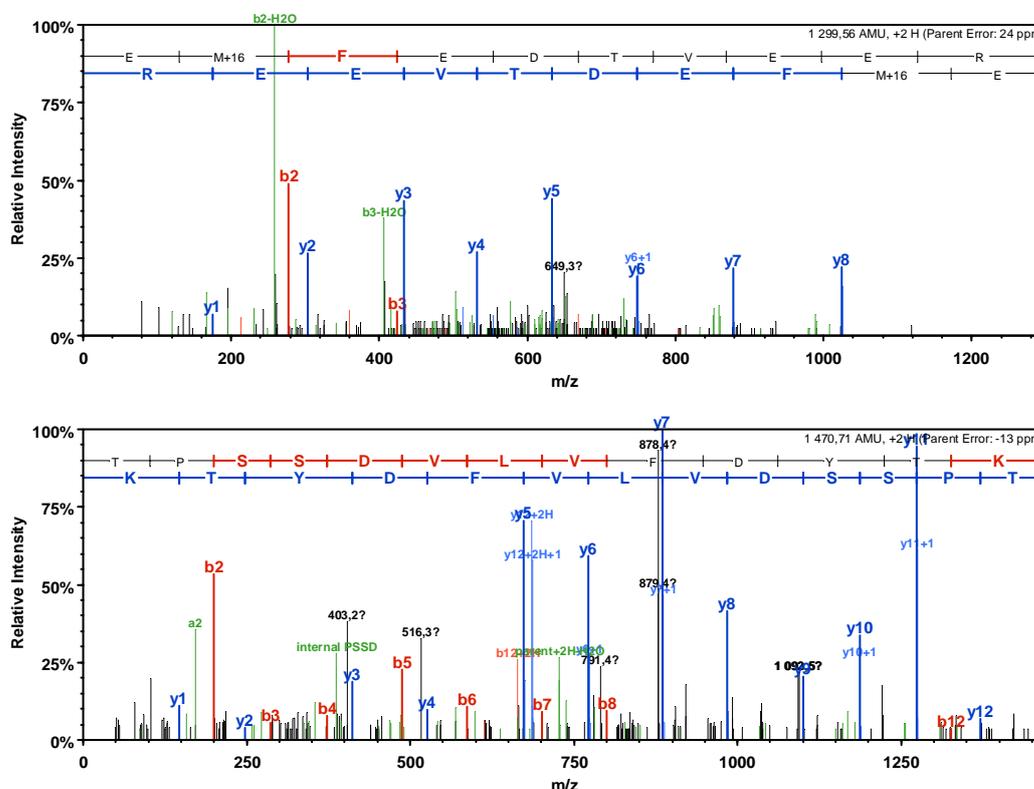
### C Identification of Retinoblastoma binding protein 4 (RbAp46),

Protein name	Protein accession numbers	Protein molecular weight (Da)	Sequence coverage	Peptide sequence	Mascot Ion score	Modifications	Observed m/z	Actual peptide mass (AMU)	Spectrum charge	Error on peptide mass (PPM)
Retinoblastoma binding protein 4	Q60972 RBBP4_MOUSE	47655	6.3%	ADKEAAFDDAVEER	53.34	Acetyl	804.34	1606.68	2	-22
				TPSSDVLVFDYTK	57.68	/	736.36	1470.41	2	-13



## D Identification of Retinoblastoma binding protein 7 (RbAp48).

Protein name	Protein accession numbers	Protein molecular weight (Da)	Sequence coverage	Peptide sequence	Mascot Ion score	Modifications	Observed m/z	Actual peptide mass (AMU)	Spectrum charge	Error on peptide mass (PPM)
Retinoblastoma binding protein 7	Q60973 RBBP7_MOUSE	47790	5.4%	EMFEDTVEER	38.25	Oxidation	650.78	1299.56	2	-24
				TPSSDVLVFDYTK	57.68	/	736.36	1470.41	2	-13



## E Detailed standard protocols for proteomic analyses.

In gel digestion:

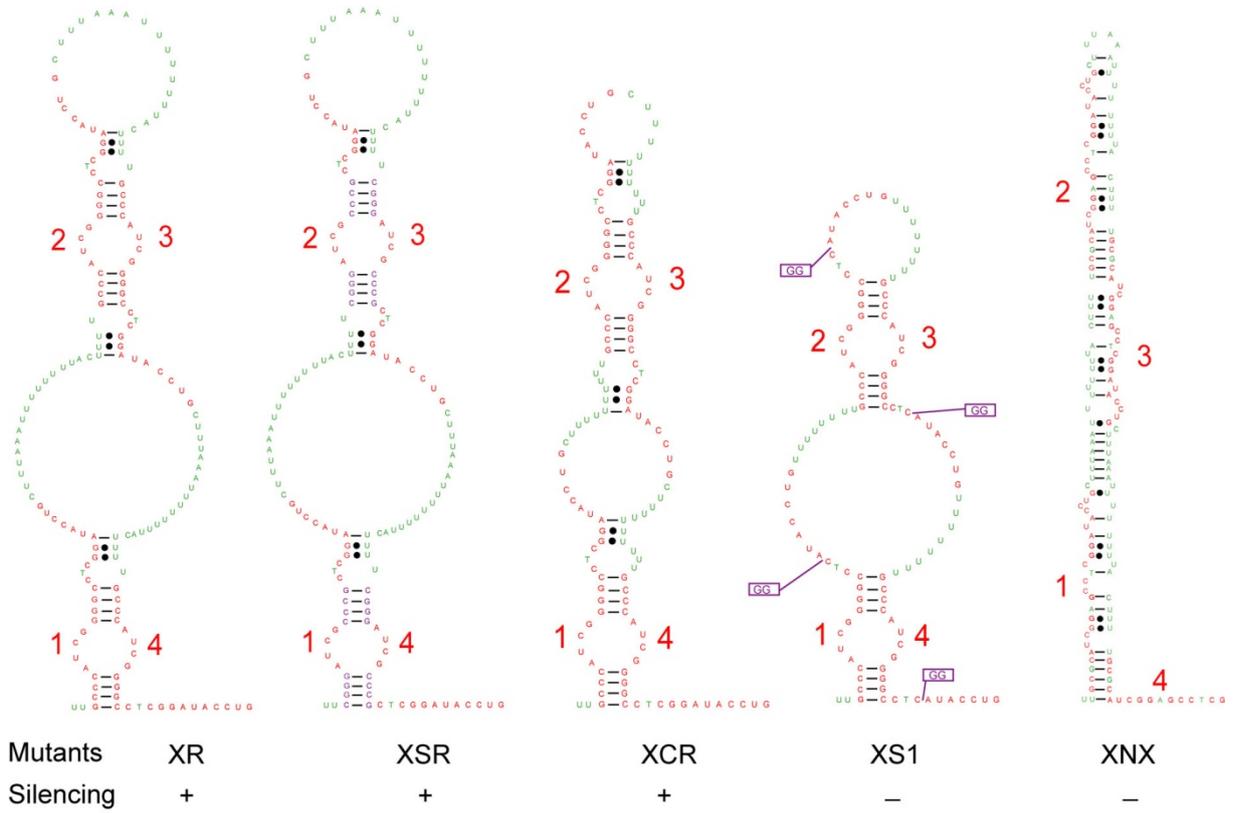
In gel digestion was performed with an automated protein digestion system, MassPrep Station (Waters, Manchester, UK). The gel plugs were washed twice with 50  $\mu$ L of 25 mM ammonium hydrogen carbonate ( $\text{NH}_4\text{HCO}_3$ ) and 50  $\mu$ L of acetonitrile. The cysteine residue were reduced by 50  $\mu$ L of 10 mM dithiothreitol at 57°C and alkylated by 50  $\mu$ L of 55 mM iodoacetamide. After dehydration with acetonitrile, the proteins were cleaved in gel with 20  $\mu$ L of 12.5 ng/ $\mu$ L of modified porcine trypsin (Promega, Madison, WI, USA) in 25 mM  $\text{NH}_4\text{HCO}_3$ . The digestion was performed overnight at room temperature. The generated peptides were extracted with 60% acetonitrile in 1% acid formic.

nanoLC-MS/MS:

From each sample, 6.4  $\mu$ L were loaded on a precolumn, before chromatographic separation on a C18 column (LC Packings C18, 75 mm id, 150 mm length). The gradient was generated by

the CapLC at a flow rate of 200 nL/min. The gradient profile consisted of a linear one from 90% of a water solution acidified by 0.1% HCOOH vol/vol (solution A), to 40% of a solution of CH<sub>3</sub>CN acidified by 0.1% HCOOH vol/vol (solution B) in 30 min, followed by a second gradient ramp to 75% of B in 1 min. Data acquisition was piloted by MassLynx software V4.0 (Waters). Calibration was performed using adducts of 0.1% phosphoric acid (Acros, NJ, USA) with a scan range from m/z 50 to 1800. Automatic switching between MS and MS/MS modes was used. The internal parameters of Q-TOF II were set as follows. The electrospray capillary voltage was set to 3.5 kV, the cone voltage set to 35 V, and the source temperature set to 90°C. The MS survey scan was m/z 300–1500 with a scan time of 1 s and an interscan time of 0.1 s. When the peak intensity rose above a threshold of 15 counts/s, tandem mass spectra were acquired. Normalized collision energies for peptide fragmentation were set using the charge-state recognition files for 1+, 2+, and 3+ peptide ions. The scan range for MS/MS acquisition was from m/z 50 to 2000 with a scan time of 1 s and an interscan time of 0.1 s. Fragmentation was performed using argon as the collision gas and with a collision energy profile optimized for various mass ranges of precursor ions. Data processing was done automatically with the ProteinLynx Process (Micromass) module.

### **Figure S8**

**Figure S9**

Primer name	Sequence (5'-3')	Positions	Characteristics
Amplification for RNA preparation			
O-3755	AAGCTTAAATTTAATACGACTCACTATAGGGCTTGCTCCAGCCATGTT	1-20	HindIII
O-3754	TCTAGACGCGCAACACCGCACACTAA	1117-1137	XbaI
O-3756	AAGCTTAAATTTAATACGACTCACTATAGGGTTGATACTTGTGTGTGTA	277-297	HindIII
O-3757	TCTAGAGCAATAACTCACAAAACCAT	741-760	XbaI
O-4240	AAGCTTAAATTTAATACGACTCACTATAGGGATCAGTTTTTACTCTTC	330-350	HindIII
O-4241	TCTAGAACAACCCACAAAACCAACA	776-796	XbaI
O-4567	TCTAGACCAGAGTGTGGGGGTTTCAG	542-562	XbaI
O-4568	AAGCTTAAATTTAATACGACTCACTATAGGACTCACTATAGGGAACCCCAACACTCTGGCC	545-565	HindIII
2D structure of Mouse A region			
O-3757	TCTAGAGCAATAACTCACAAAACCAT	741-760	XbaI
O-3758	TGTGAAAAAAGACTAAACGC	592-612	
O-3760	ATGGAATGGCGGGGAGGACA	361-381	
O-3866	TTTTTTTTTTTTTTTTTACAAAAG	527-552	
O-3971	AGCCCCGATGGGCAAAGA	661-679	
O-3973	GATGGGCAAGTTTAGA	444-460	
2D structure of Human A region			
O-4242	ACAACCCACAAAACCAACAT	776-796	
O-4563	AAAAAGAATTAAGGCAGG	432-452	
O-4564	GGGTTTCAGAGGGGAAGGGAA	530-551	
O-4565	AAAAATAAAAAAAAAAAG	635-655	
O-4622	AGAAAATTTTTAAAAGCAG	582-602	
FRET experiments			
P1	GAATATATAACAATGAAAGAAAGG	319-332	Cy3
P2	GCGGGGAGGACACACAGGTA	355-372	Cy3
P3	AATGAAAAGGCAGGTAAGTA	399-418	Cy3
P4	GAAGAAAAAAGAATAAAAG	473-496	Cy5
P5	ACAAAAAGCAGGTATCCATG	515-534	Cy5
P6	GCAATAACTCACAAAACCAT	741-761	Cy3
P7	GGTTTTTTTTTTTTTTTTT	542-561	Cy5
P3'	AATGAAAAGGCAGGTAAGTA	399-418	Cy5

**Table S1**



### III. Etude de l'interaction de la région des A-repeats avec des facteurs protéiques

#### III.1 Stratégie de purification et d'identification des complexes ribonucléoprotéiques formés avec la région des A-repeats

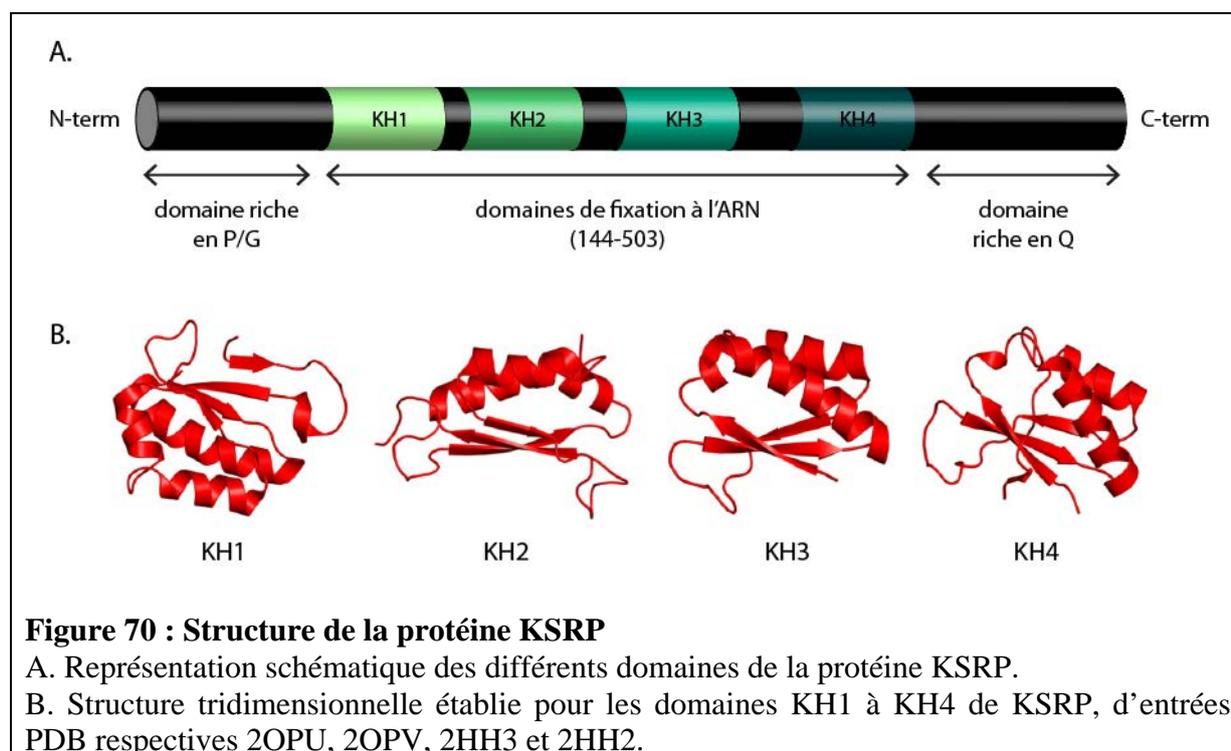
Au laboratoire, pour purifier des complexes ribonucléoprotéiques et identifier les protéines partenaires d'un ARN, nous utilisons une technique de chromatographie d'affinité dite « MS2/MBP ». Elle repose sur l'insertion de manière co-transcriptionnelle de 3 segments de l'ARN du bactériophage MS2 d'*E.coli* à l'extrémité 3' de l'ARN d'intérêt, créant un ARN hybride. Ces motifs se lient avec une forte affinité à la protéine de l'enveloppe de ce bactériophage ( $K_d \approx 38 \text{ pM}$ ). Cette protéine est produite sous forme fusionnée avec la protéine MBP (Maltose Binding Protein), ce qui permet l'immobilisation de l'ARN hybride sur les billes d'amylose auxquelles la protéine de fusion RBP/MS2 est capable de se fixer (cf. **Figure 56** du paragraphe II.2 de la partie Matériel & Méthodes). L'ARN ainsi immobilisé peut alors être incubé avec un échantillon d'extrait nucléaire afin de former des complexes ribonucléoprotéiques. Après plusieurs lavages destinés à éliminer les protéines liées de manière non spécifique, les complexes sont élués en conditions natives par incubation avec une solution de maltose. Les protéines des complexes sont fractionnées par électrophorèse en gel SDS-PAGE. Après coloration, les bandes peuvent alors être découpées et analysées par spectrométrie de masse, ce qui permet d'identifier les partenaires protéiques de l'ARN.

Les analyses de spectrométrie de masse réalisées à partir des expériences de S. Maenner ont permis d'identifier une centaine de partenaires potentiels de la région des A-repeats. Au vu de la quantité de protéines identifiées, il a été nécessaire d'en sélectionner certaines pour une étude plus approfondie de leur interaction avec la région des A-repeats. En tenant compte de leur profil avec des extraits nucléaires de cellules ES de souris femelles préparées à différents stades de différenciation, J0, J2 et J4 (fenêtre de temps lors de laquelle s'initie l'inactivation), mais aussi de l'intensité des bandes de gel et après une analyse approfondie de la littérature, nous avons décidé de nous concentrer sur les protéines KSRP, PTB et Suz12 (qui fait partie du complexe PRC2). Pour ma part, j'ai axé mon travail principalement sur les protéines KSRP et PTB. D'autres protéines intéressantes ont été sélectionnées dans un deuxième temps, nous y reviendrons à la fin de ce chapitre.

#### III.2 Analyse biochimique de l'interaction avec la protéine KSRP

La protéine KSRP, ou KH type-Splicing Regulatory Protein, a la capacité de se fixer aux ARN. En effet, elle possède 4 domaines de fixation à l'ARN, des domaines KH (K-

homology) (**Figure 70**). Elle intervient dans de nombreux mécanismes concernant les ARN, que ce soit dans leur transcription, leur maturation, leur transport ou encore leur dégradation (Davis-Smyth *et al.*, 1996 ; Min *et al.*, 1997 ; Markovtsov *et al.*, 2000). Elle est notamment impliquée dans la dégradation de certains ARN messagers. Elle facilite leur prise en charge par la machinerie de dégradation, l'exosome, en se fixant sur des éléments riches en A/U (ARE, AU-rich element) (Chen *et al.*, 2001 ; Gherzi *et al.*, 2004). Ces séquences, le plus souvent composées du motif AUUUA, se situent en général dans la région 3' UTR des ARNm. Les domaines 3 et 4 de la KSRP constituent la région minimale permettant l'association à haute affinité avec les ARE des ARN ciblés, le domaine KH4 étant essentiel pour la reconnaissance des sites. Le domaine KH3 seul est responsable de l'interaction de KSRP avec la poly(A) ribonucléase et avec l'exosome. Il a également été montré le rôle de la structuration particulière de KH4 dans la spécificité de reconnaissance des ARNm (García-Mayoral *et al.*, 2007). Il est intéressant de noter que la région des A-repeats contient des régions riches en résidus A et U (régions séparant les répétitions) et qu'il a été montré un rôle de l'exosome dans la mise en place l'inactivation du chromosome X (Ciaudo *et al.*, 2006).



**Figure 70 : Structure de la protéine KSRP**

A. Représentation schématique des différents domaines de la protéine KSRP.

B. Structure tridimensionnelle établie pour les domaines KH1 à KH4 de KSRP, d'entrées PDB respectives 2OPU, 2OPV, 2HH3 et 2HH2.

Par ailleurs, cette protéine est impliquée dans le processus d'épissage. En effet, il a été montré que KSRP fait partie d'un complexe multiprotéique se fixant sur un élément intronique activant l'inclusion de l'exon N1 de l'isoforme neuronale de c-src (Min *et al.*, 1997). Le rôle de KSRP est d'induire l'assemblage des 5 autres protéines sur le site de

fixation du complexe. De plus, il a été observé que la concentration de KSRP est plus élevée dans les cellules neuronales par rapport aux non-neuronales. Elle peut donc être considérée comme un régulateur de l'épissage, agissant sur l'inclusion ou non d'intron, en fonction de sa concentration dans certains types cellulaires.

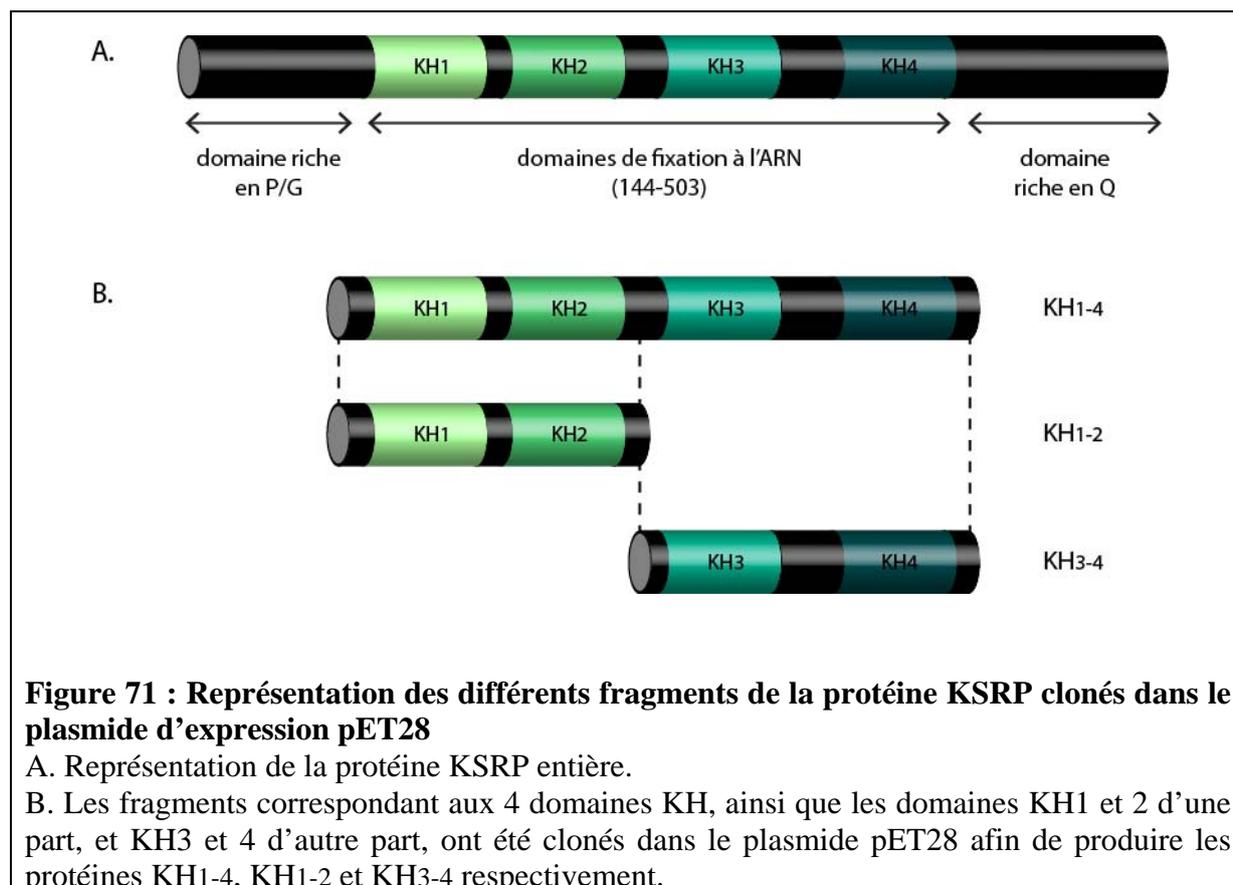
En parallèle, la protéine PTB, que nous avons décrite dans l'introduction pour son rôle dans la régulation de l'épissage, participe également à la régulation de l'épissage de l'exon N1 (Chan and Black, 1997 ; Chou *et al.*, 2000). Elle ne se fixe pas au niveau de la même séquence que KSRP mais elles coopèrent, avec également les protéines hnRNP F ou H notamment, pour la répression de l'épissage de l'exon N1. Or la protéine PTB, comme nous l'avons mentionné dans l'introduction, a été identifiée par notre laboratoire comme un partenaire protéique de la région des A-repeats, se fixant sur les régions riches en C et U que constituent les boucles internes des structures tige-boucle de l'ARN.

Toutes ces données ajoutées au fait que KSRP ait été identifiée parmi les protéines éluées par chromatographie d'affinité sur la région des A-repeats, suggèrent un rôle de cette protéine dans l'inactivation du chromosome X, soit de manière indirecte (par exemple par stabilisation de la structure des A-repeats), soit de manière directe (en agissant sur l'épissage ou la dégradation de Xist). C'est pourquoi nous avons cherché à caractériser l'interaction, potentielle, de KSRP avec la région des A-repeats.

### III.2.1 Clonage et production de la protéine KH1-4

La protéine KSRP est composée de 4 domaines de fixation à l'ARN, dénommés KH1 à KH4. Elle possède également dans sa partie N-terminale une région riche en acides aminés proline et glycine, et dans sa partie C-terminale une région riche en glutamine (**Figure 71.A**). A. Douablin du laboratoire de F. Baklouti (CGMC, Lyon), qui avait effectué un stage dans notre laboratoire, avait tenté de produire la protéine KSRP entière chez *E. coli*. Cependant, elle s'était retrouvée dans les corps d'inclusion. A. Douablin avait donc cloné la région codant les 4 domaines KH de la protéine dans le plasmide pGEX-4T1, permettant la production de la protéine sous forme fusionnée à la GST (pGEX-4T1-KH1-4). Par la suite, S. Maenner a utilisé son clone pour réaliser la construction pET28b-KH1-4 de manière à produire la même région mais fusionnée à une étiquette 6His. L'étiquette 6His ayant une taille beaucoup plus petite que la GST, elle ne devrait pas gêner les études d'interaction. En effet, la protéine GST peut induire une structuration différente de la protéine à laquelle elle est fusionnée, ou encore cacher un domaine d'interaction.

A partir de cette construction, j'ai ensuite cloné le fragment correspondant aux domaines KH1 et 2 et le fragment correspondant aux domaines KH3 et 4 dans le plasmide pET28 également (**Figure 71.B**). Pour réaliser les constructions, le fragment *EcoRI/XhoI* de plasmide pET28b-KH1-4 codant KH1 et KH2 a été cloné dans le plasmide pET28b (pET28b-KH1-2) et le fragment codant KH3 et KH4, digéré par les mêmes enzymes, a été cloné dans le plasmide pET28a (pET28a-KH3-4). Les plasmides ont été choisis afin de respecter le cadre de lecture de chaque protéine recombinante.



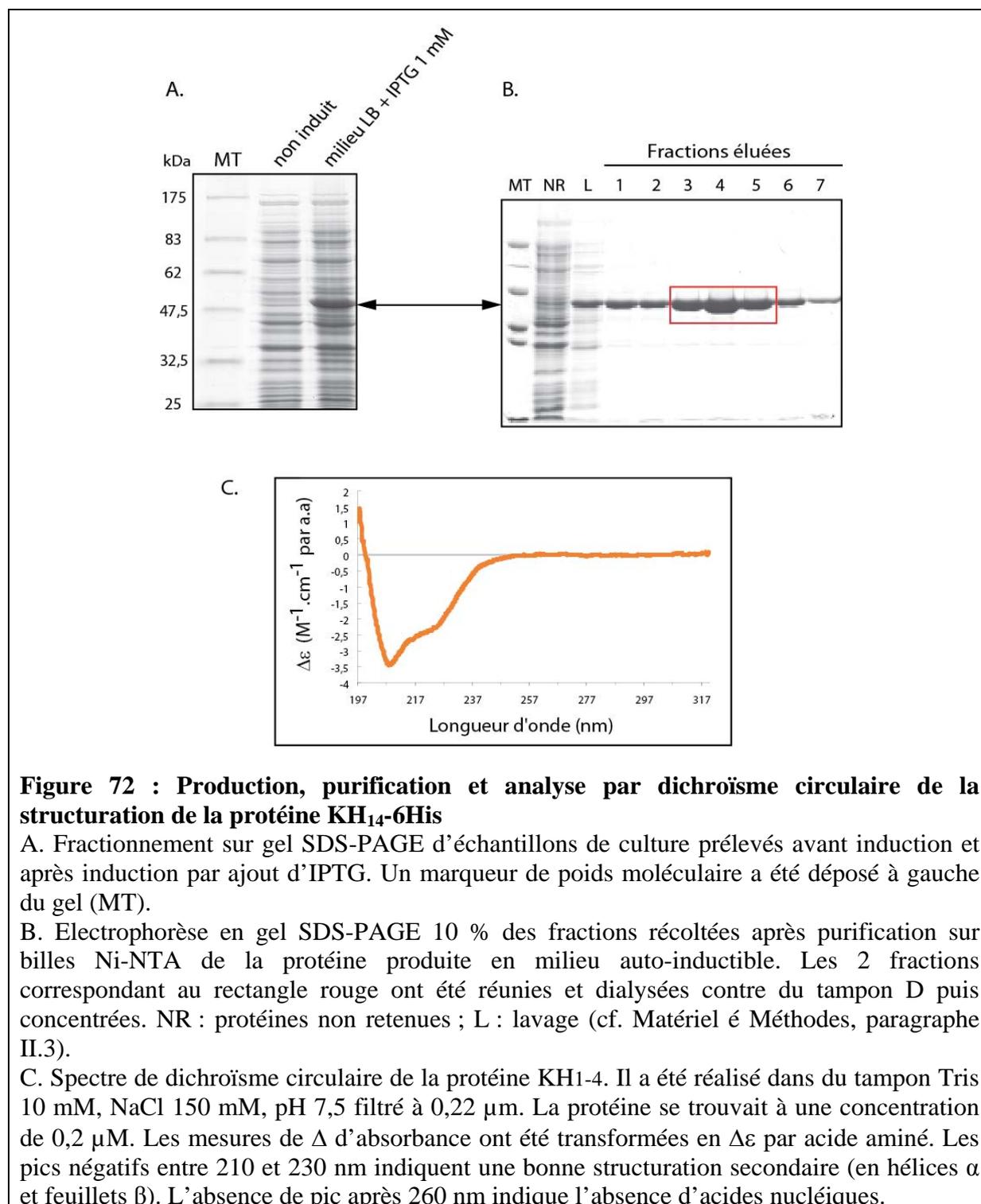
**Figure 71 : Représentation des différents fragments de la protéine KSRP clonés dans le plasmide d'expression pET28**

A. Représentation de la protéine KSRP entière.

B. Les fragments correspondant aux 4 domaines KH, ainsi que les domaines KH1 et 2 d'une part, et KH3 et 4 d'autre part, ont été clonés dans le plasmide pET28 afin de produire les protéines KH1-4, KH1-2 et KH3-4 respectivement.

La production de la protéine recombinante KH1-4 a été réalisée dans la souche d'*E. coli* BL21 codon+, par induction à l'IPTG. Comme le montre la **Figure 72**, un haut degré de pureté a pu être obtenu après purification sur résine Ni-NTA. L'absence d'acides nucléiques a été vérifiée par mesure du rapport d'absorbance  $A_{260}/A_{280}$  et par dichroïsme circulaire qui a permis également de vérifier la bonne structuration des protéines produites. Le spectre exposé dans la Figure 72.C présente un pic négatif entre 210 et 230 nm. Ce pic correspond aux structures secondaires en hélices  $\alpha$  et en feuillets  $\beta$ . La présence de ce pic est donc une preuve de la structuration de la protéine.

La protéine KH3-4 n'a pas pu être produite chez *E.coli*, tandis que, bien que nous ayons réussi à produire des quantités importantes de KH1-2, nous n'avons pas été en mesure de la purifier avec un degré de pureté suffisant pour nos expériences.

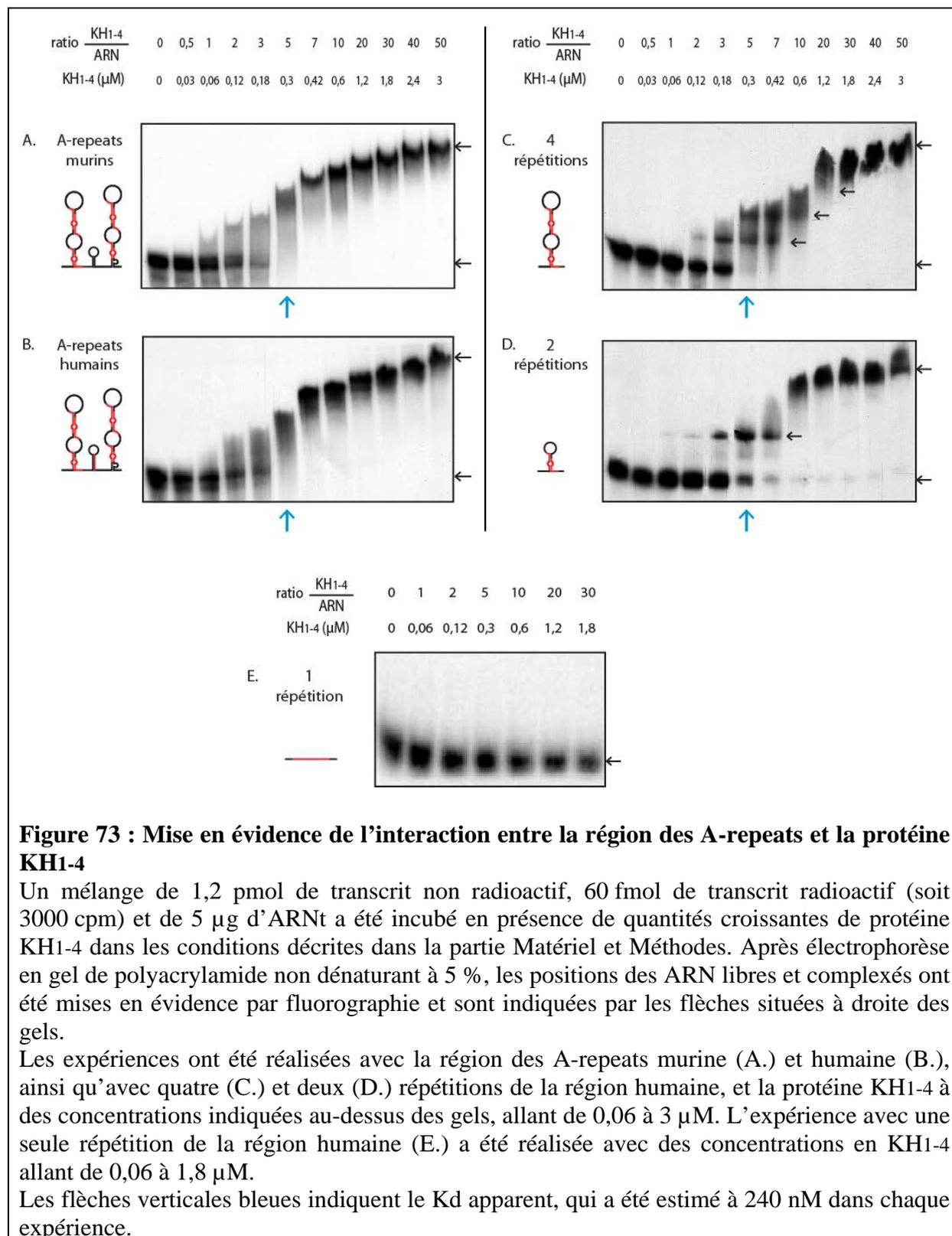


### III.2.2 Expériences de retard sur gel

Afin de vérifier le caractère direct de l'interaction entre la région des A-repeats et la protéine KH1-4, nous avons ensuite réalisé des expériences de retard sur gel. Pour cela, nous avons utilisé les ARN de la région des A-repeats entière murine et humaine, ainsi que les ARN correspondant aux répétitions 1 à 4 et aux répétitions 3 et 4 humaines. Nous avons incubé les différents ARN avec des concentrations croissantes en protéine pour obtenir des rapports molaires [protéines] / [ARN] allant de 0 à 50. La **Figure 73** présente les résultats obtenus.

Dans les panneaux A et B correspondants aux régions entières, on peut observer un retard dès l'ajout de 0,06  $\mu\text{M}$  de KH1-4, soit un ratio [protéine:ARN] de 1:1. L'ARN libre disparaît totalement à une concentration de 0,42  $\mu\text{M}$  (ratio de 7). Nous pouvons donc affirmer que la protéine KH1-4 interagit directement avec la région des A-repeats. Pour les concentrations en protéine supérieures, le complexe formé est plus fortement retardé, ce qui indique qu'il existe plusieurs sites de fixation sur l'ARN. La gamme de concentrations en protéine étant suffisamment large et les paliers [ARN libre] et [ARN complexé] étant bien visibles, nous avons pu apprécier la constante de dissociation apparente de l'interaction. Elle correspond à la concentration en protéine pour laquelle 50 % de l'ARN est sous forme complexée. Dans le cas des régions entières murine et humaine, nous l'avons estimée à 240 nM. Pour comparaison, le  $K_d$  observé avec la protéine PTB est de 120 nM (expériences réalisées par S. Maenner).

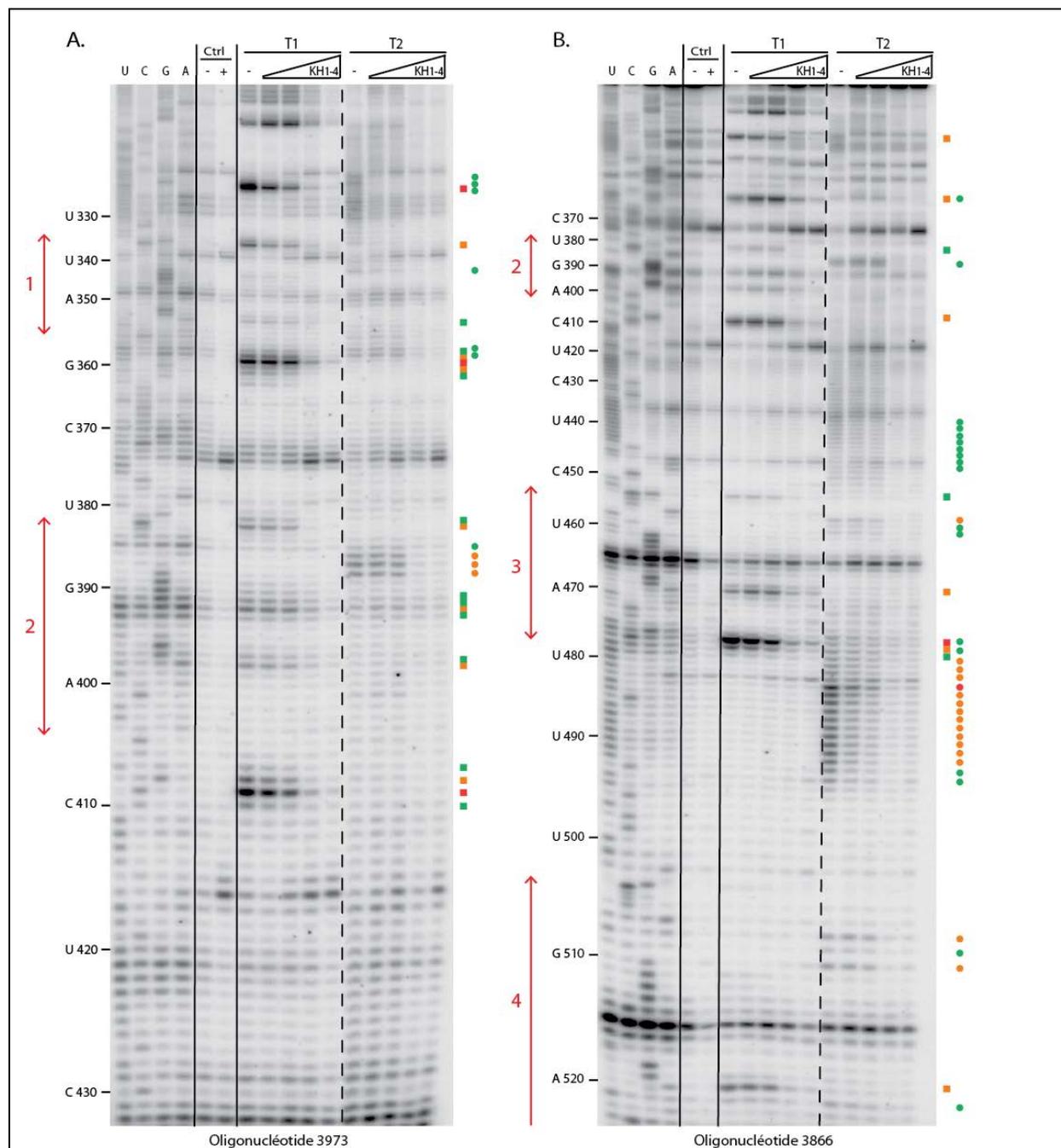
Les panneaux C et D représentent les expériences avec les régions tronquées à 4 ou 2 répétitions. Le retard de migration, même s'il ne débute qu'à partir de l'ajout de 0,12  $\mu\text{M}$  de protéine, indique une interaction directe entre ces régions et la protéine KH1-4. En revanche, le profil général de migration n'est pas progressif comme avec les régions entières, mais sous forme de paliers. On peut observer la présence de 4 paliers au total dans l'expérience avec 4 répétitions (**Figure 73.C**), et de 2 paliers avec 2 répétitions (**Figure 73.D**). Ces résultats semblent indiquer que chaque élément répété contient un site de fixation pour la protéine KSRP. Cependant, l'expérience réalisée avec une seule répétition ne montre aucune formation de complexe (**Figure 73.E**). La présence d'au moins 2 répétitions est donc nécessaire à la fixation de la protéine. Nous pouvons mettre cette information en relation avec la structure de l'ARN. En effet, dès 2 répétitions, l'ARN peut adopter une structure en tige-boucle. Cet élément structural serait ainsi essentiel à la fixation de la protéine KSRP sur la région des A-repeats.



### III.2.3 Etude par empreinte aux RNases T1 et T2 des sites de fixation de la protéine KH1-4 sur la région des A-repeats

J'ai ensuite réalisé des expériences d'empreinte en utilisant les RNases T1 et T2, afin de définir les sites de fixation de KH1-4 sur la région des A-repeats entière, murine et humaine. En fonction des expériences de retard sur gel, les hydrolyses ont été réalisées sur des complexes formés à quatre ratios différents [protéine:ARN] : 5, 10, 30 et 50. Pour cela, 1,2 pmoles d'ARN ont été incubés en présence de 0, 6, 12, 36 et 60 pmol de protéine, puis les complexes formés ont été soumis à des hydrolyses ménagées dans les conditions décrites dans la partie Matériel et Méthodes. Les positions de clivage ont été identifiées par extension d'amorce. Le fractionnement des produits d'extension d'amorce en gel de polyacrylamide dénaturant obtenu avec les oligonucléotides amorce 3973 et 3866 qui permettent d'analyser les 4 premières répétitions de la région murine est illustré dans la **Figure 74**. Les protections détectées à partir de ces extensions d'amorce sont représentées sur la structure représentant la première tige-boucle dans la **Figure 76**. Comme nous l'attendions, une grande partie des régions riches en résidus U et A sont protégées par la protéine KH1-4.

Les expériences sur la région humaine ont révélé globalement les mêmes sites de fixation même si les résultats obtenus sont plus partiels (**Figure 75** et **Figure 76**). Ils sont localisés au niveau des boucles, internes et terminales, des structures tige-boucle ainsi qu'au niveau des hélices qui les jouxtent. Il se pourrait que la fixation de KSRP sur ces régions permette de stabiliser la naissance des boucles et donc la structure entière de la région des A-repeats. Cependant, d'autres rôles de cette protéine sont envisageables, notamment des rôles plus directement fonctionnels dans le mécanisme de l'inactivation.

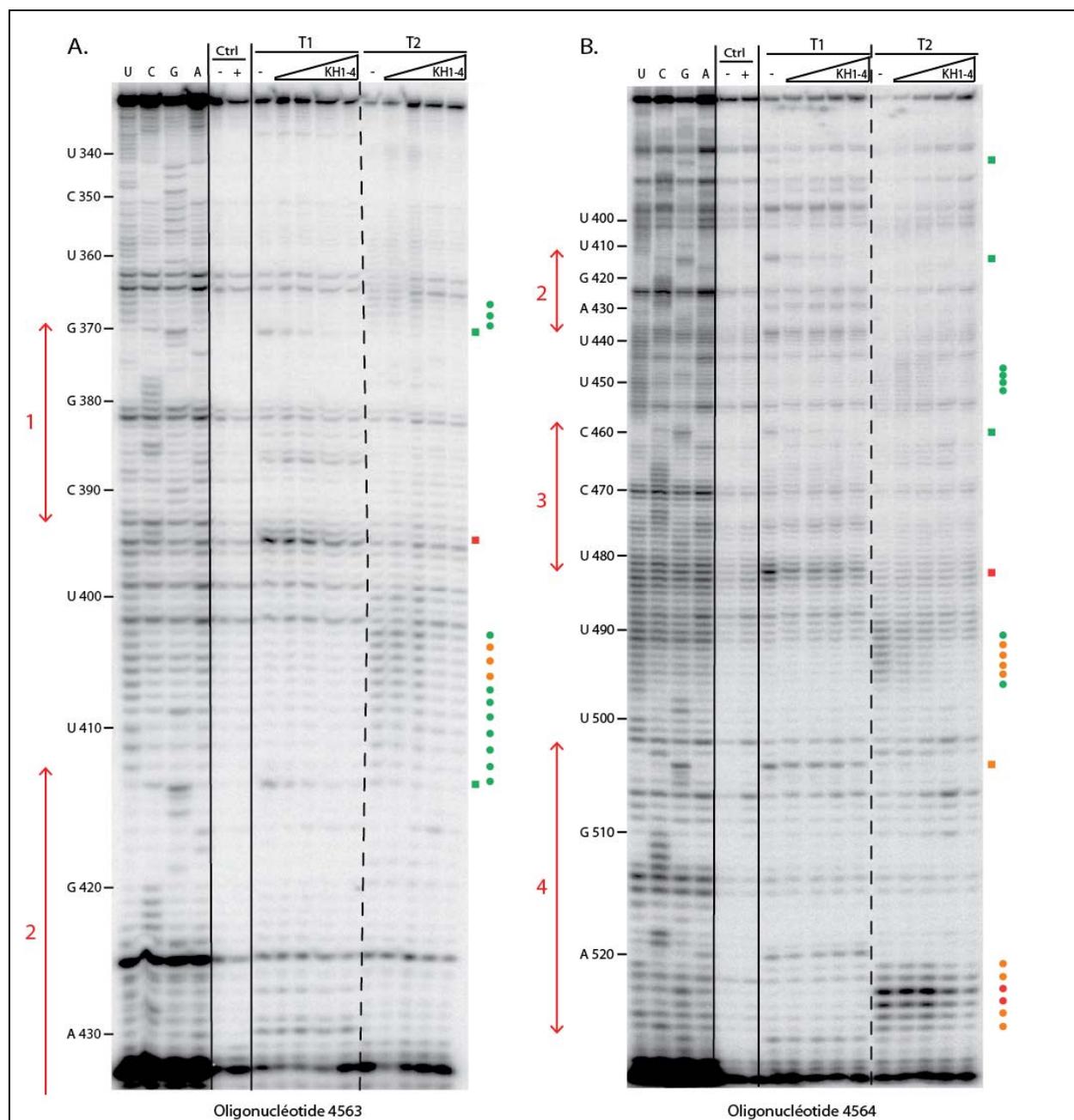


### Figure 74 : Identification des séquences de la région des A-repeats murine fixant la protéine KH1-4

L'ARN de la région des A-repeats murine a été soumis à des coupures par les RNases T1 et T2 en conditions ménagées, comme décrit dans la partie Matériel et Méthodes. Différentes concentrations de protéine KH1-4 ont été testées pour chaque sonde enzymatique et chaque oligodésoxyribonucléotide : 0 ; 0,3 ; 0,6 ; 1,8 et 3  $\mu$ M. Les pistes contrôle (Ctrl) correspondent à des réactions sans enzyme, sans (-) ou avec 3  $\mu$ M (+) de protéine KH1-4. Les positions des coupures ont été identifiées par extension des oligodésoxyribonucléotides 3973 (correspondant aux répétitions 1 et 2, Panneau A.) ou 3866 (correspondant aux répétitions 2 à 4, Panneau B.). Le séquençage du transcrit a été réalisé avec les mêmes amorces (pistes U, G, C, A). Les ADNc ont été fractionnés sur gel de polyacrylamide dénaturant à 7 %.

Les positions des nucléotides sont indiquées à gauche des autoradiographies. Les répétitions présentes dans la région des A-repeats sont indiquées en rouge par leur numéro, à droite des

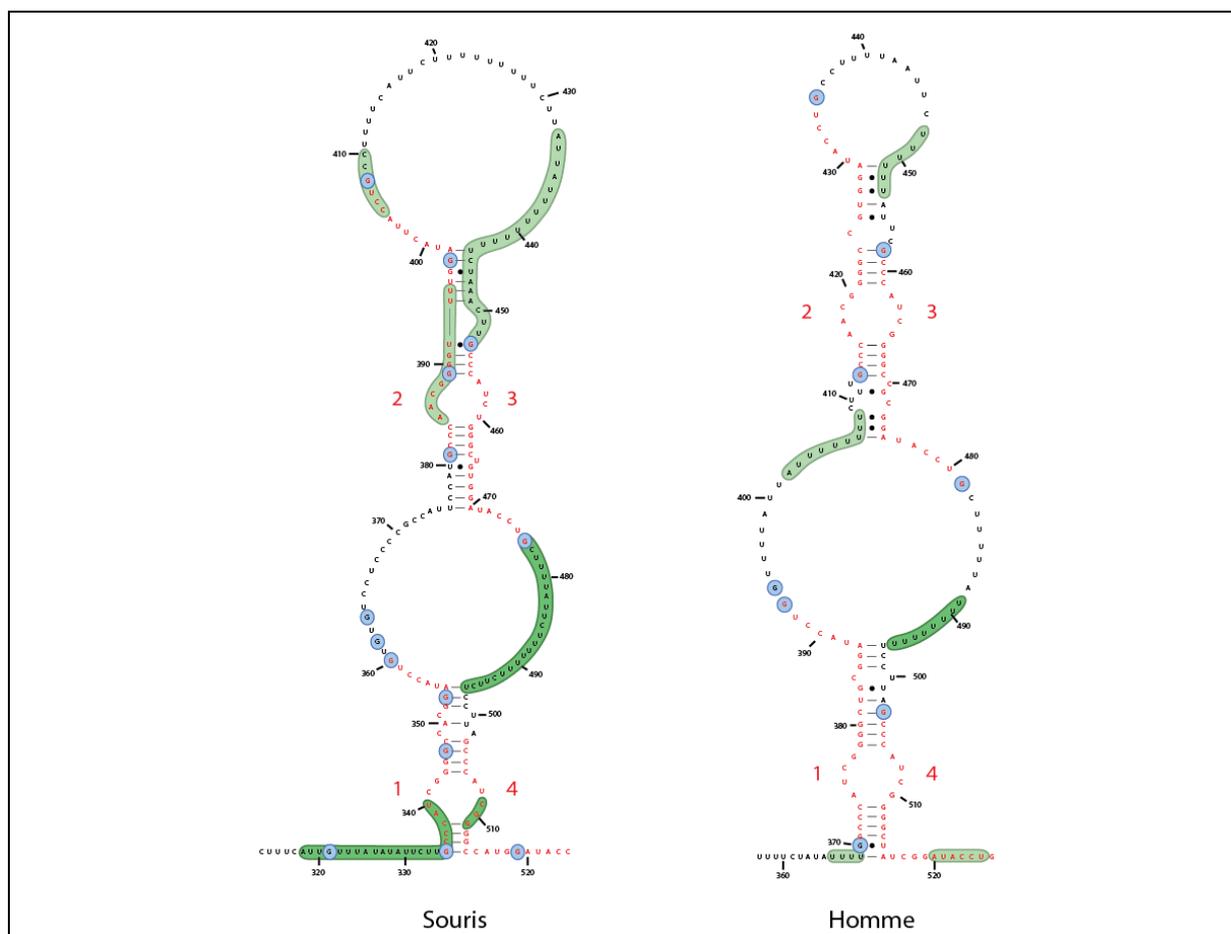
autoradiographies. Les clivages par les RNases T1 et T2 sont représentés par des carrés et des cercles respectivement. L'intensité du clivage est indiquée par la couleur (vert pour faible, orange pour moyenne et rouge pour forte).



**Figure 75 : Identification des séquences de la région des A-repeats humaine fixant la protéine KH1-4**

L'ARN de la région des A-repeats humaine a été soumis à des coupures par les RNases T1 et T2 en conditions ménagées, comme décrit dans la partie Matériel et Méthodes. Différentes concentrations de protéine KH1-4 ont été testées pour chaque sonde enzymatique et chaque oligodésoxyribonucléotide : 0 ; 0,3 ; 0,6 ; 1,8 et 3  $\mu$ M. Les pistes contrôle (Ctrl) correspondent à des réactions sans enzyme, sans (-) ou avec 3  $\mu$ M (+) de protéine KH1-4. Les positions des coupures ont été identifiées par extension des oligodésoxyribonucléotides 4563 (correspondant aux répétitions 1 et 2, Panneau A.) ou 4564 (correspondant aux répétitions 2 à 4, Panneau B.).

Le séquençage du transcrit a été réalisé avec les mêmes amorces (pistes U, G, C, A). Les ADNc ont été fractionnés sur gel de polyacrylamide dénaturant à 7 %. Les positions des nucléotides sont indiquées à gauche des autoradiographies. Les répétitions présentes dans la région des A-repeats sont indiquées en rouge par leur numéro, à droite des autoradiographies. Les clivages par les RNases T1 et T2 sont représentés par des carrés et des cercles respectivement. L'intensité du clivage est indiquée par la couleur (vert pour faible, orange pour moyenne et rouge pour forte).



**Figure 76 : Localisation des zones de la région des A-repeats murine et humaine protégées par la protéine KH1-4**

Les zones protégées ont été déterminées à partir des résultats d'empreintes, sur la région des A-repeats murine et humaine correspondant aux 4 premières répétitions. Les cercles bleus correspondent aux signaux observés avec la RNase T1 et celles en vert aux signaux observés avec la RNase T2. Plus la couleur est intense, plus le site de fixation est fort.

S. Maenner avait également déterminé les sites de fixation de la protéine PTB sur la région des A-repeats. Ils se situent sur l'ensemble de la longueur des boucles internes et terminales des structures tige-boucle. Les sites de fixation de KSRP les chevauchent sur une petite zone, correspondant à la base des boucles. Même si les sites ne se recouvrent pas en totalité, la fixation de l'une des deux protéines peut gêner la fixation de la seconde. Une autre

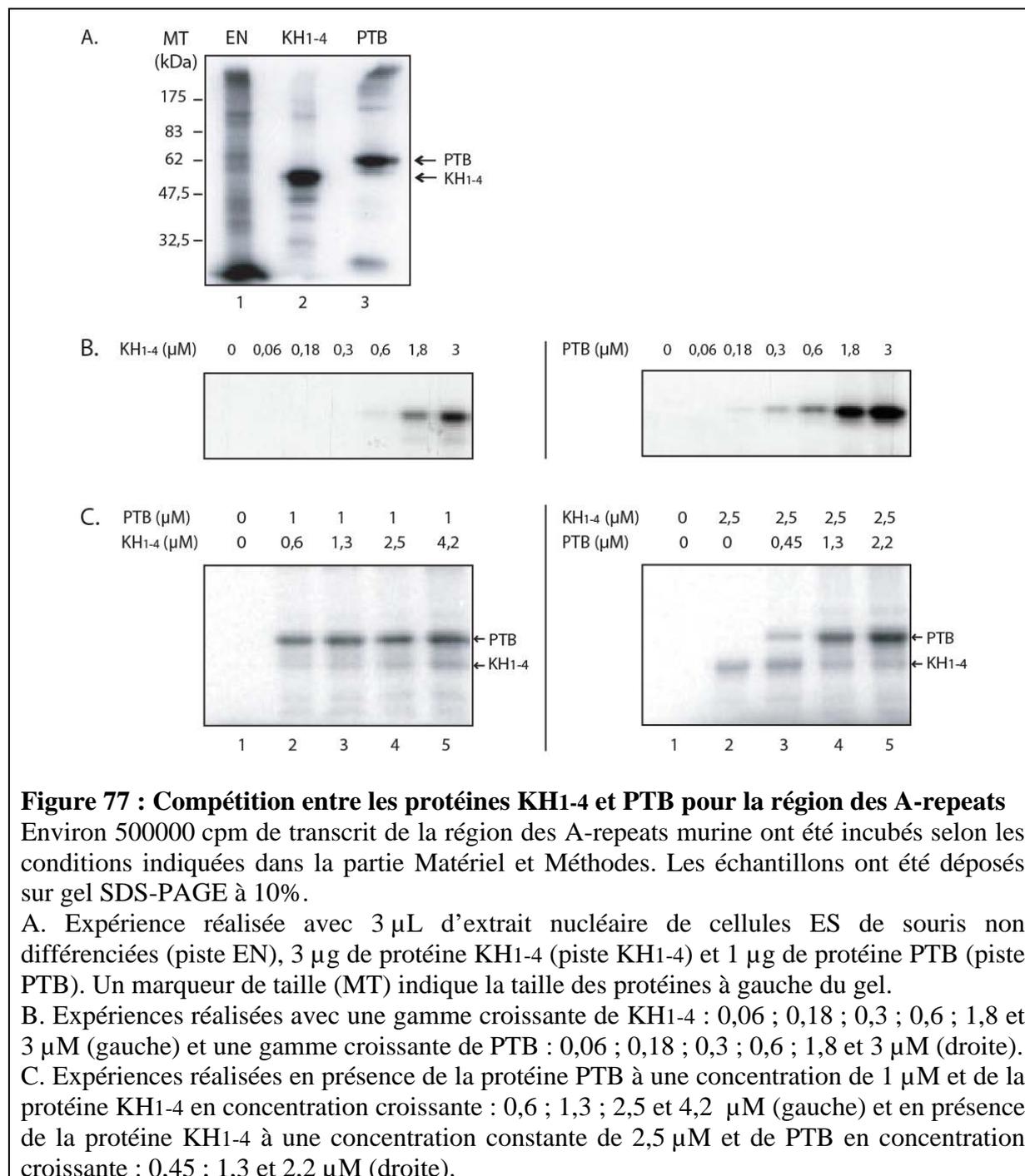
possibilité serait que la fixation de l'une facilite celle de la seconde. C'est pour cette raison que nous avons ensuite étudié la compétition entre ces 2 protéines pour leur fixation sur la région des A-repeats.

#### III.2.4 Etude de la compétition KSRP/PTB par pontage aux UV

J'ai d'abord vérifié que dans nos conditions expérimentales, les deux protéines se fixent sur l'ARN Xist (**Figure 77.A**). La présence de bandes intenses aux tailles attendues pour les deux protéines indique que celles-ci se fixe sur la région des A-repeats en interagissant directement avec l'ARN. J'ai ensuite effectué les mêmes expériences mais en utilisant une gamme de concentration en protéines afin de déterminer la concentration à laquelle le signal obtenu est suffisamment fort mais non saturé. Ainsi les diminutions ou les augmentations du signal dans les expériences de compétition seront facilement observables. Nous avons opté pour une concentration de KH1-4 à 2,5  $\mu$ M (**Figure 77.B**) et une concentration de PTB à 1  $\mu$ M (**Figure 77.C**) dans nos conditions.

Dans l'expérience suivante, la concentration de PTB a été maintenue à 1  $\mu$ M alors que la concentration de KH1-4 varie de 0 à 4,2  $\mu$ M. Comme nous pouvons le constater sur la **Figure 77.D**, l'intensité du signal de la protéine PTB ne varie pas alors que celui de KH1-4 augmente en fonction de la concentration. Dans ces conditions, la protéine KH1-4 n'entre donc pas en compétition avec la protéine PTB mais se fixe très probablement sur des sites éventuellement non occupés par PTB, qui rappelons le, n'est pas utilisée à des concentrations saturantes.

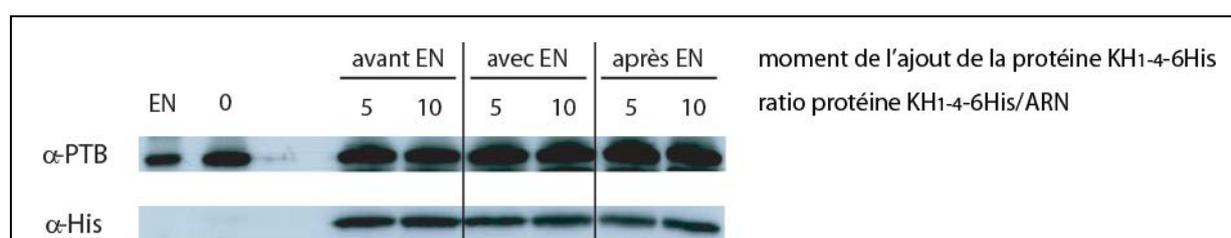
En revanche, lorsque la concentration de la protéine KH1-4 est maintenue stable (2,5  $\mu$ M) et la concentration de PTB augmente, le signal correspondant à cette dernière s'intensifie alors que celui de la KH1-4 diminue (**Figure 77.E**). Ces résultats indiquent que la protéine PTB déplacerait KH1-4 pour s'y fixer.



J'ai ensuite réalisé des expériences de chromatographie d'affinité avec l'ARN de la région des A-repeats fusionné à l'ARN MS2, incubé avec un extrait nucléaire de cellules HEK dans lequel j'ai ajouté la protéine KH1-4-6His recombinante. Le but de ces expériences était de tester si l'incubation avec KH1-4-6His avant, pendant ou après l'incubation avec l'extrait nucléaire influençait la fixation de la protéine recombinante KH1-4-6His et de la protéine PTB, provenant elle de l'extrait nucléaire, sur la région des A-repeats. Les résultats obtenus sont présentés dans la **Figure 78**. La détection de la protéine KH1-4-6His a été

réalisée grâce à un anticorps dirigé contre son étiquette 6xHis. Deux quantités de protéine KH1-4-6His ont été testées (50 et 100 pmol, soit 5 ou 10 fois plus que d'ARN), afin de visualiser une éventuelle différence de fixation selon la concentration. Nous avons observé (**Figure 78**), on peut observer que l'intensité de la bande correspondant à PTB est légèrement plus faible dans le cas où la protéine KH1-4-6His est incubée avec l'ARN avant l'extrait nucléaire. Il est donc possible que la fixation de KH1-4-6His sur la région des A-repeats préalablement à PTB empêche la fixation de cette dernière.

Notons cependant que dans cette expérience, la fixation de PTB n'empêche pas celle de KH1-4-6His, même quand cette dernière est ajoutée après l'incubation avec l'extrait. Nous pouvons mettre cette observation en relation avec le fait que dans l'extrait nucléaire, la protéine PTB est moins concentrée que dans les expériences avec la protéine recombinante, et qu'elle est déjà utilisée à d'autres processus nucléaires, donc moins disponible pour la région des A-repeats. Il se pourrait que la protéine KH1-4-6His ait ainsi plus de sites libres, ou moins encombrés, pour se fixer. Il ne faut pas oublier non plus que la protéine PTB est cette fois issue d'un extrait nucléaire et a donc pu subir les modifications post-traductionnelles qui lui sont normalement apposées. Ces modifications, telles que la phosphorylation, peuvent influencer sur l'activité de la protéine mais aussi sur ses propriétés de fixation à l'ARN. Nous pouvons supposer que la compétition entre ces 2 protéines pour la région des A-repeats dépend de la concentration cellulaire de chacune d'entre elles ainsi que de leurs modifications post-traductionnelles respectives, et que la variation de ces paramètres dans la fenêtre de temps où se met en place l'inactivation du chromosome X peut jouer un rôle dans le processus.



**Figure 78 : Expériences de chromatographie d'affinité MBP/MS2 réalisées avec la région des A-repeats et un extrait nucléaire complétement en protéine KH1-4-6His**

Une quantité de 10 pmol d'ARN correspondant à la région des A-repeats murine fusionnée à l'ARN MS2 est incubée en présence de 100  $\mu$ g d'extrait nucléaire de cellules HEK. L'ARN est également incubé avec 50 ou 100 pmol de protéine KH1-4 (soit un ratio molaire [protéine] / [ARN] de 5 ou 10) avant l'incubation avec l'extrait nucléaire, en même temps, ou après. Après élution par du maltose, les protéines sont fractionnées par électrophorèse en gel SDS-PAGE et révélées par Western-blot.

Quant au rôle précis de ces protéines, plusieurs hypothèses sont envisageables. En effet, leurs sites de fixation, notamment ceux de PTB, laissent supposer une fonction dans la stabilisation de la structure de la région des A-repeats. Correctement structurée, elle pourrait alors recruter des facteurs ayant une action plus directe dans le processus d'inactivation. Les protéines KSRP et PTB peuvent jouer également un rôle de protéines plateforme pour le recrutement d'autres protéines partenaires. De plus, au vu des fonctions de ces protéines dans la cellule, nous ne pouvons pas exclure un rôle dans la régulation de l'épissage de l'ARN Xist (PTB) ou dans celle de sa dégradation (KSRP). Enfin, nous pouvons envisager que la plus forte présence de l'une de ces protéines par rapport à l'autre engagerait l'ARN dans des voies différentes.

### III.3 Autres protéines partenaires de la région des A-repeats

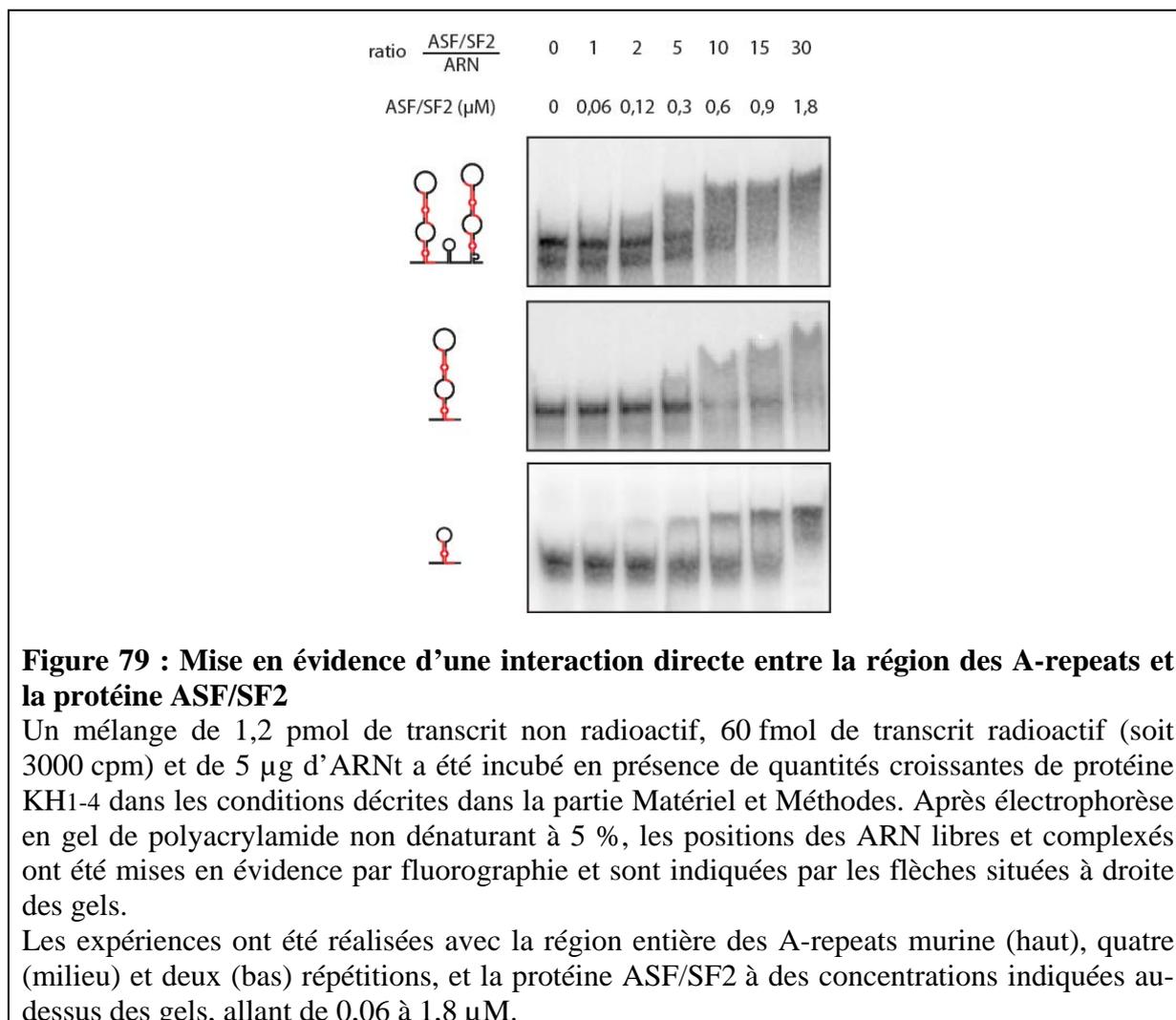
#### III.3.1 ASF/SF2

Récemment, il a été mis en évidence une interaction entre la protéine ASF/SF2 et la région des A-repeats (Royce-Tolland *et al.*, 2010). De plus, les auteurs ont montré qu'elle joue un rôle dans l'épissage de Xist *in vivo*. Les résultats publiés nous ont incités à caractériser son interaction avec la région des A-repeats et son rôle dans l'épissage de l'ARN Xist.

##### a. Etude de l'interaction par expérience de retard sur gel

De la même manière qu'avec la protéine KH1-4, j'ai réalisé des expériences de retard sur gel avec des concentrations croissantes de protéine ASF/SF2 en présence de la région des A-repeats. La protéine recombinante nous a été fournie par J. Stevenin (IGBMC, Strasbourg). Dans le but d'observer si elle se fixe également à une région plus courte des A-repeats, l'expérience a également été effectuée avec un ARN comprenant 4 et 2 répétitions (**Figure 79**). Le retard de migration prouve que l'interaction entre ASF/SF2 et la région des A-repeats est directe. Cependant, dans le cas de la région entière et de 4 répétitions, le palier maximal de complexes formés n'est pas tout à fait atteint, même avec un excès molaire de protéine de 30 fois par rapport à l'ARN. En raison d'une quantité insuffisante de protéine recombinante à notre disposition, l'expérience n'a pas pu être reproduite avec des concentrations supérieures. Dans le cas d'une incubation avec 2 répétitions, la totalité de l'ARN est sous forme de complexes avec la protéine lorsque celle-ci est en excès molaire de 30 fois, donc le nombre de molécules fixées dépend du nombre de répétitions de la région.

Au vu des profils obtenus, nous avons estimé le  $K_d$  de cette interaction à 700 nM. Par comparaison avec celui des protéines KH1-4 et PTB (240 et 120 nM respectivement), nous pouvons en déduire que l'interaction entre ASF/SF2 et la région des A-repeats est relativement faible, ce qui expliquerait pourquoi nous n'avons pas pu l'identifier dans les complexes analysés par spectrométrie de masse.

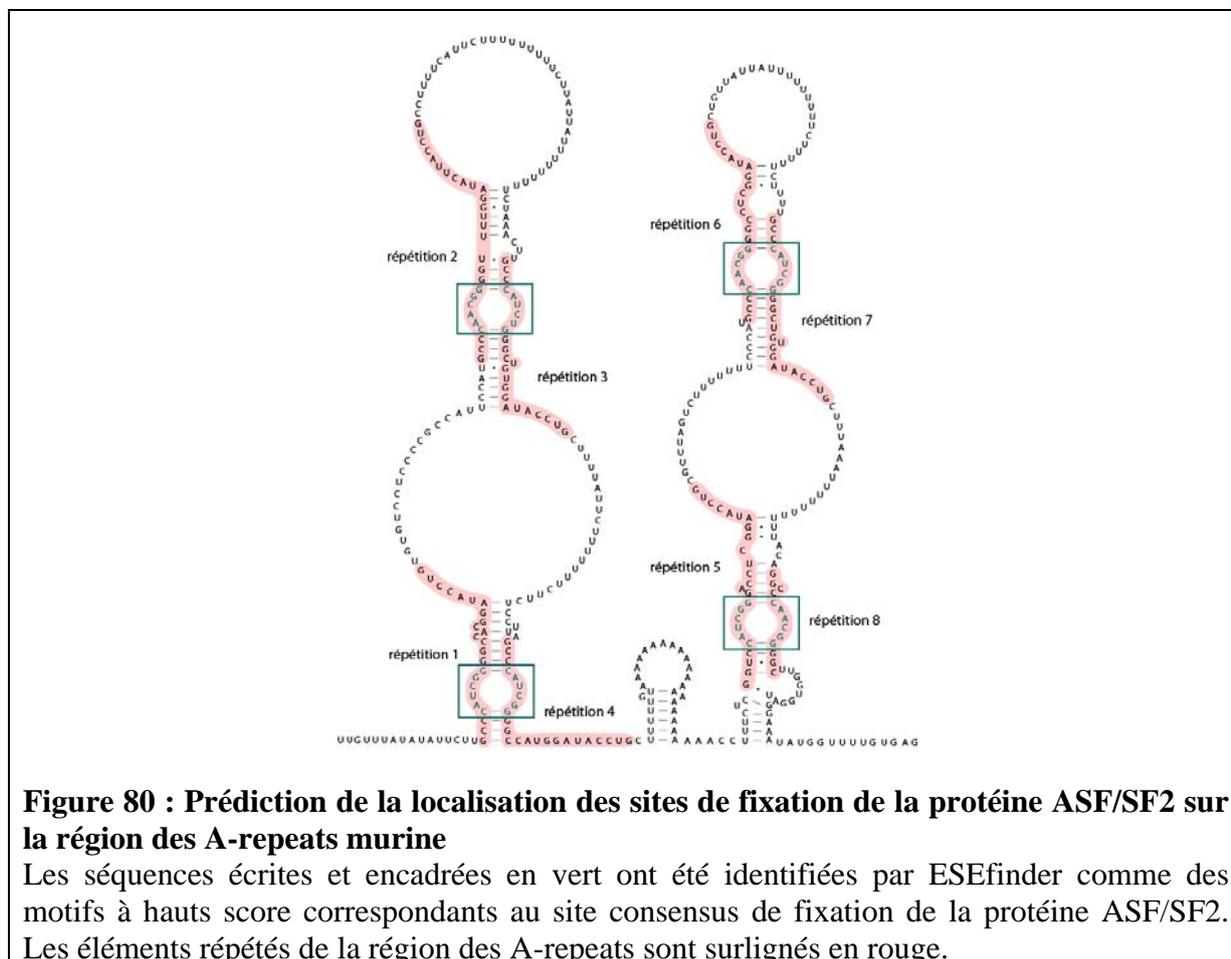


L'interaction entre ASF/SF2 et la région des A-repeats étant directe, l'étape suivante fût de chercher les sites de fixation de la protéine sur cette région.

#### b. Prédiction informatique des sites de fixation d'ASF/SF2

Le site de fixation consensus de la protéine ASF/SF2 selon ESEfinder (déterminé par SELEX fonctionnel) est : C(A/G)(C/G)(A/C)(C/G)G(A/U). En cherchant un tel site dans la région des A-repeats, nous trouvons la séquence CAUCGGG, située au début de chaque répétition. Par rapport à la structure 2D de cette région, nous pouvons observer que ces

séquences forment de petites boucles internes au sein des grandes structures tige-boucle (**Figure 80**). Ces structurations particulières pourraient ainsi constituer des sites favorables à la fixation de la protéine ASF/SF2.

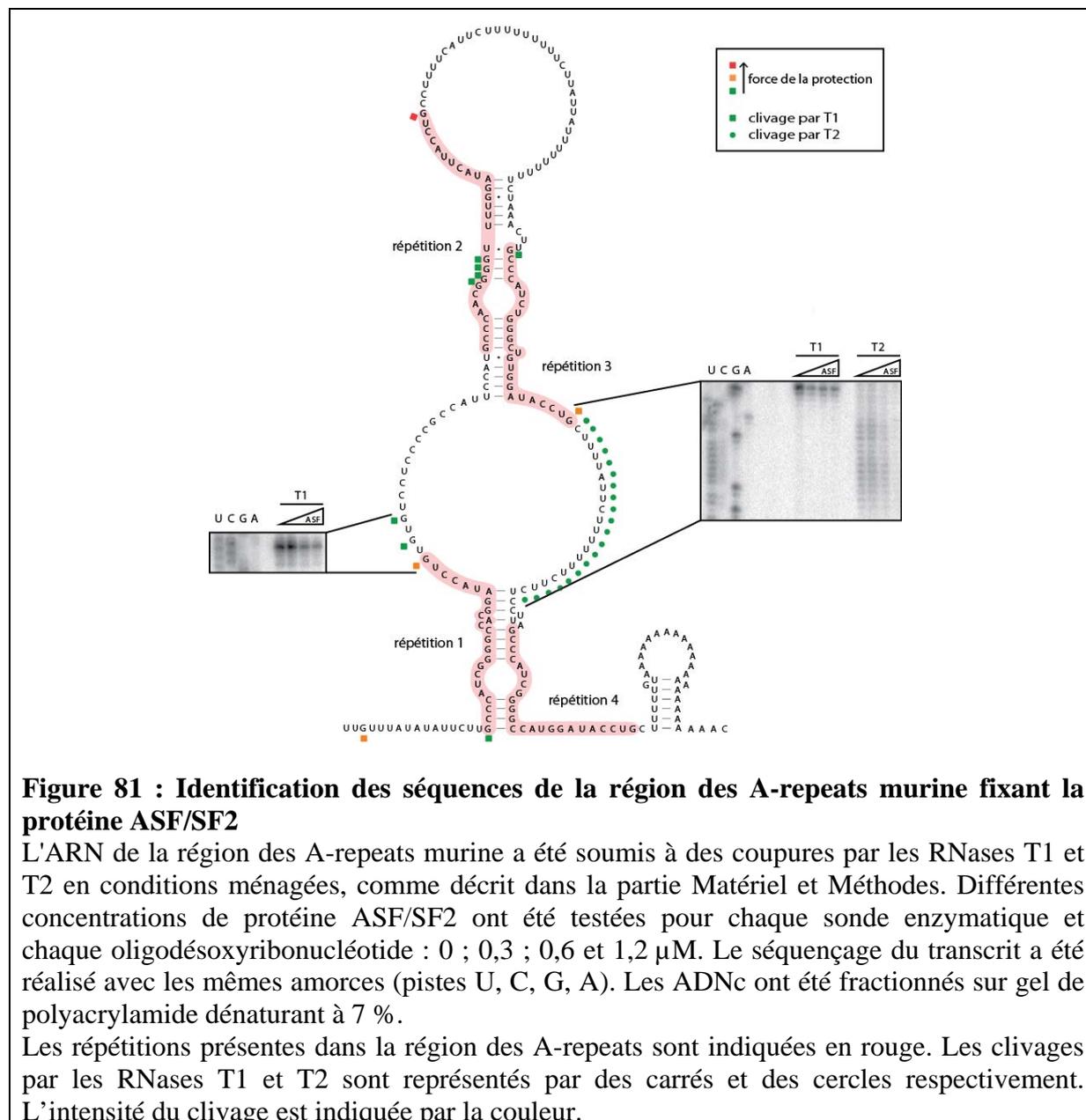


Il était donc cohérent de chercher ensuite les sites de fixation exacts. C'est pourquoi nous avons réalisé des expériences d'empreintes en présence de la protéine ASF/SF2 recombinante.

c. Analyse des sites de fixation d'ASF/SF2 sur la région des A-repeats par expériences d'empreintes

Des coupures par les RNases T1 et T2 ont été réalisées sur les 4 premières répétitions de la région des A-repeats murine, dans les mêmes conditions que pour la protéine KH1-4. De ce que nous pouvons observer dans la **Figure 81**, les résidus protégés par la présence de la protéine se trouvent essentiellement au niveau de la boucle interne de la structure tige-boucle formée par l'ARN. Or les séquences identifiées ne ressemblent pas à la séquence consensus des sites de fixation d'ASF/SF2, ni avec la séquence prédite dans le paragraphe précédent.

Cette expérience n'a été réalisée qu'une seule fois, avec une concentration maximale en protéine ASF/SF2 de 1,2  $\mu$ M. Elle mériterait d'être reproduite avec des concentrations supérieures afin de confirmer les résultats obtenus ici.



Nous avons donc montré que la protéine ASF/SF2 interagit de manière directe avec la région des A-repeats, dès la présence de 2 répétitions. De plus, sa présence protège fortement les résidus G se trouvant aux extrémités des répétitions mais aussi au sein de la répétition 2. Ces résultats expliquent ceux de l'équipe de B. Panning qui a montré par expérience de pull-down que la protéine ASF/SF2 pouvait se fixer sur un seul élément répété (Royce-Tolland *et*

*al.*, 2010). L'analyse de son rôle dans la régulation de l'épissage sera décrite dans le paragraphe IV.4.

Comme nous l'avons mentionné auparavant, S. Maenner, par des expériences de chromatographie d'affinité suivies d'analyses par spectrométrie de masse, avait identifié une centaine de protéines partenaires potentiels de la région des A-repeats. Certaines d'entre elles, de par leur rôle décrit dans la littérature, nous ont semblé intéressantes. Nous avons donc tout d'abord voulu confirmer les résultats par des expériences de chromatographie d'affinité suivies de Western-blot.

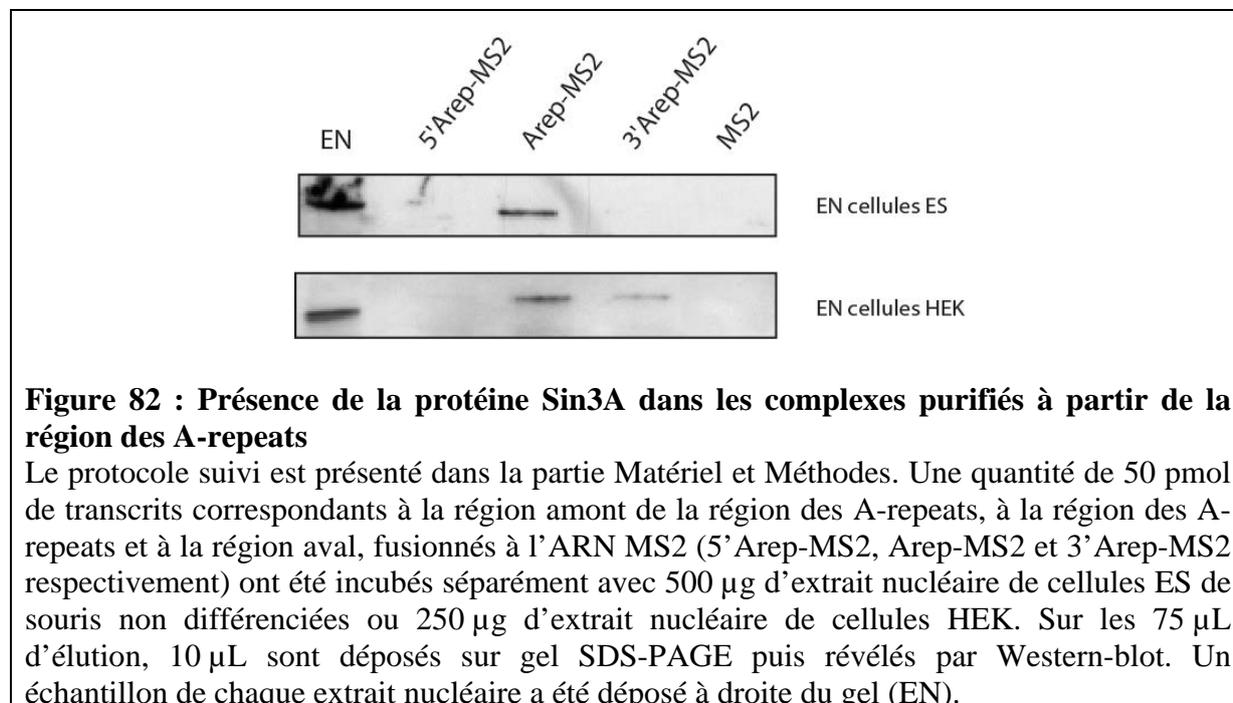
### III.3.2 Sin3A

Sin3A (SWI-independent) a d'abord été identifiée génétiquement comme un régulateur global de la transcription. Cette grosse protéine est composée essentiellement de domaines d'interactions protéine-protéine. Sa fonction est de fournir un support structural pour un complexe hétérogène Sin3A/HDAC (Hassig *et al.*, 1997 ; Laherty *et al.*, 1997 ; Zhang *et al.*, 1997). En plus des HDAC, Sin3A peut séquestrer d'autres fonctions enzymatiques, dont le remodelage des nucléosomes, la méthylation de l'ADN ou encore la méthylation des histones (Sif *et al.*, 2001 ; Yang *et al.*, 2003). Bien que la plupart des études se concentrent sur son rôle de co-répresseur, il a été montré que cette protéine peut également réguler la transcription de manière positive. Sin3A est donc une protéine plateforme permettant la modification de la chromatine par de multiples combinaisons, utilisant des mécanismes séquence-spécifiques mais aussi des mécanismes indépendants de la séquence.

Etant donné les modifications épigénétiques ayant lieu au cours de l'inactivation, il était naturel de se demander si Sin3A avait un rôle dans leur mise en place, et si la région des A-repeats pouvait servir à la recruter au début de la différenciation.

La **Figure 82** présente les résultats obtenus après purification de complexes RNP par la technique de chromatographie d'affinité MBP/MS2 en présence d'un extrait nucléaire de cellules ES de souris femelles non différenciées et d'un extrait nucléaire de cellules HEK. Pour cette expérience, en plus de la région des A-repeats, j'ai également utilisé les régions en amont et en aval, afin de vérifier la spécificité d'interaction de la protéine. Comme le montre la figure, la protéine Sin3A fait bien partie des protéines en complexe avec la région des A-repeats. Notons toutefois qu'une bande apparaît dans le cas d'une incubation de la région aval des A-repeats avec un extrait nucléaire de cellules HEK, mais pas avec l'extrait de cellules

ES. Cependant, aucune bande n'est présente lors d'une incubation avec la région amont ou avec l'ARN MS2 seul. Nous en avons conclu que Sin3A est en interaction, directe ou indirecte, spécifiquement avec la région des A-repeats de l'ARN Xist.



**Figure 82 : Présence de la protéine Sin3A dans les complexes purifiés à partir de la région des A-repeats**

Le protocole suivi est présenté dans la partie Matériel et Méthodes. Une quantité de 50 pmol de transcrits correspondants à la région amont de la région des A-repeats, à la région des A-repeats et à la région aval, fusionnés à l'ARN MS2 (5'Arep-MS2, Arep-MS2 et 3'Arep-MS2 respectivement) ont été incubés séparément avec 500 µg d'extrait nucléaire de cellules ES de souris non différenciées ou 250 µg d'extrait nucléaire de cellules HEK. Sur les 75 µL d'élution, 10 µL sont déposés sur gel SDS-PAGE puis révélés par Western-blot. Un échantillon de chaque extrait nucléaire a été déposé à droite du gel (EN).

Sin3A ne possédant pas de domaine de fixation à l'ARN ni à l'ADN, il serait intéressant de connaître par l'intermédiaire de quel autre facteur elle est en complexe avec cette région. Il faudrait ensuite chercher à comprendre son rôle dans l'inactivation, notamment son implication dans la mise en place des modifications épigénétiques liés à l'inactivation. De plus, il a été montré que le complexe Sin3A/HDAC interagit avec Sox2 et est impliqué dans la régulation de l'expression de Nanog, 2 facteurs que nous avons présentés parmi les régulateurs de l'expression *Xist* (Baltus *et al.*, 2009).

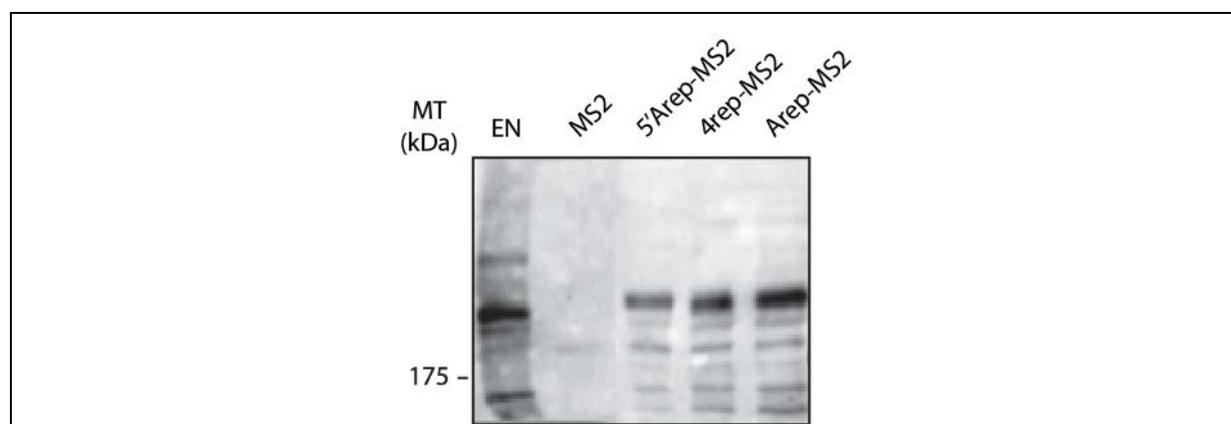
### III.3.3 Rif1

Rif1 (Rap1 interacting factor 1) a initialement été identifiée chez la levure comme une protéine de fixation aux télomères. Il a récemment été montré qu'elle affecte l'ordre temporel dans lequel les segments chromosomiques sont répliqués lors de la fission des levures. Un défaut en cette protéine avance la réplication de certains gènes normalement répliqués tardivement (Hayano *et al.*, 2012). Dans la continuité de cette étude, il a été montré que la protéine humaine affecte également les profils spatiaux-temporels de la réplication (Yamazaki *et al.*, 2012). La protéine Rif1 s'associe avec des structures d'architecture nucléaires durant

l'interphase et se lie fortement à la chromatine après la mitose, de manière concomitante avec le point de décision d'entrée en phase de réplication. Dans des cellules déplétées de Rif1, tous les profils de réplication nucléaires sont réarrangés, certains domaines présentant du retard de réplication et d'autres de l'avance. Des observations similaires ont été faites chez la souris (Cornacchia *et al.*, 2012). Il semble que Rif1 établit le programme de réplication en retardant la réplication de groupes spécifiques de régions chromatiniennes. Le rôle de Rif1 se ferait en lien avec une action sur la structure de la chromatine d'ordre supérieur, notamment par la formation de boucles (pour revue, Aladjem, 2012).

Rif1 a donc été identifiée parmi les protéines partenaires de la région des A-repeats. Son interaction avec cette région a été confirmée par Western-blot (Figure 83). Les résultats obtenus montrent également une présence dans les complexes formés avec la région amont des A-repeats, Rif1 n'est donc pas spécifique de la région des A-repeats. Elle serait néanmoins spécifique de l'ARN Xist car aucune bande n'est détectée lors des expériences menées avec l'ARN MS2 seul.

Un rôle potentiel de cette protéine dans le processus d'inactivation du chromosome X se situerait au niveau de la réplication tardive observée pour le Xi (cf. paragraphe III.1.2 de l'introduction).

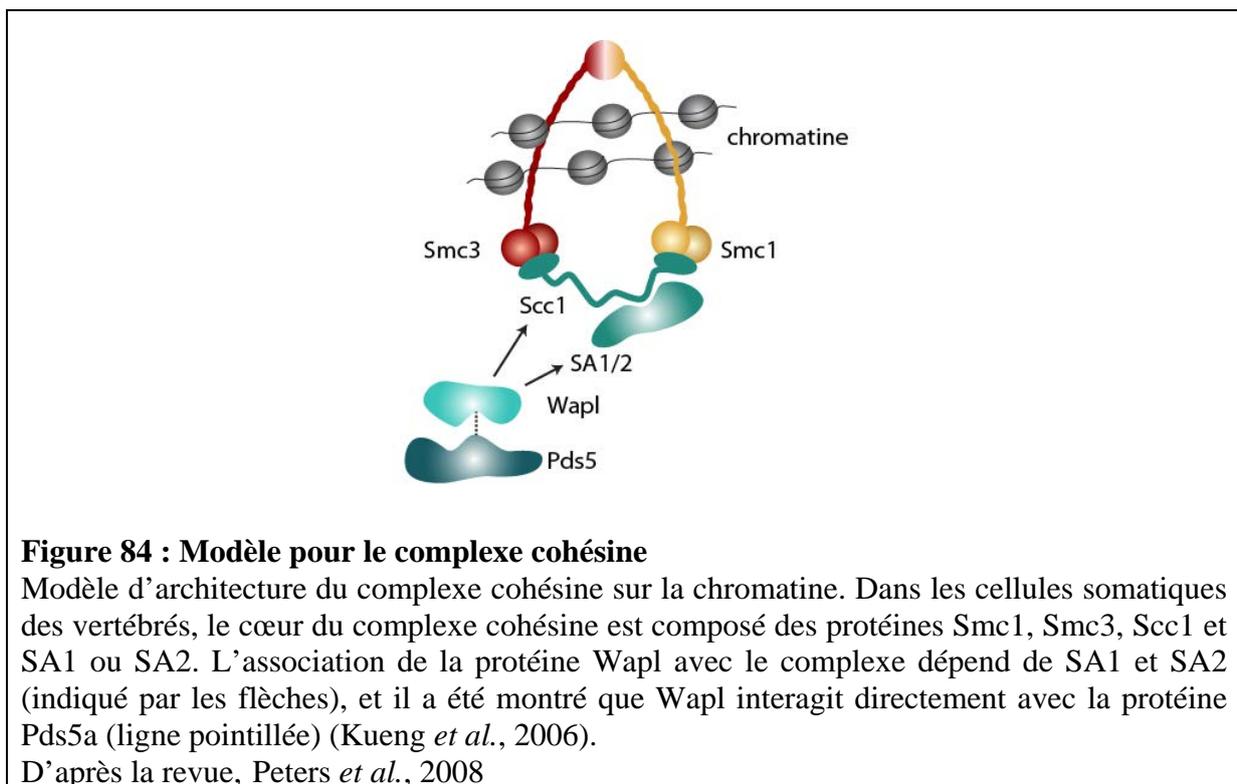


**Figure 83 : Mise en évidence de la protéine Rif1 parmi les protéines en complexe avec la région des A-repeats**

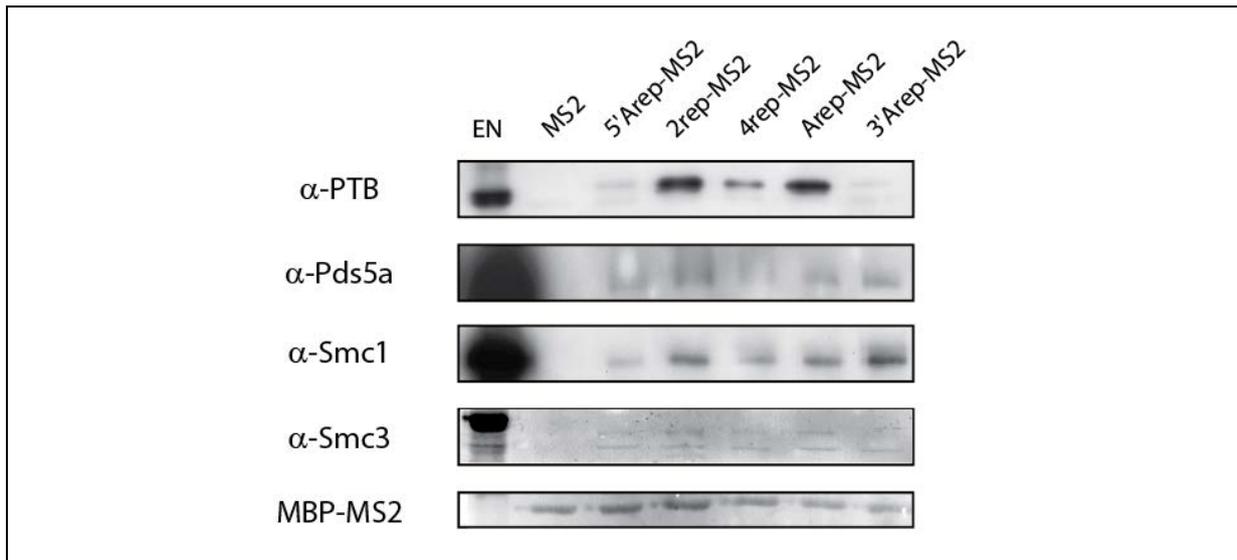
Le protocole suivi est présenté dans la partie Matériel et Méthodes. Une quantité de 100 pmol de transcrits correspondants à la région amont des A-repeats, à 4 répétitions et à la région entière des A-repeats, fusionnés à l'ARN MS2 (5'Arep-MS2, 4rep-MS2 et Arep-MS2 respectivement) ont été incubés séparément avec 500 µg d'extrait nucléaire de cellules ES de souris non différenciées. Sur les 75 µL d'élution, 10 µL sont déposés sur gel SDS-PAGE 6 % puis les bandes sont révélées par Western-blot. Un échantillon d'extrait nucléaire a été déposé à droite du gel (EN).

### III.3.4 Pds5a

Egalement connue sous le nom de SCC-12 (sister chromatid cohesion), cette protéine joue un rôle dans l'établissement, la maintenance et la dissolution de la cohésion des chromatides sœurs (Hartman *et al.*, 2000 ; Panizza *et al.*, 2000 ; Stead *et al.*, 2003). Pour cela, elle est en interaction avec le complexe cohésine à l'origine du rapprochement des chromatides sœurs. Ce complexe est notamment composé des protéines Smc1 et 3 qui constituent les « bras » entourant la chromatine (**Figure 84**). Ces 2 protéines font partie des partenaires de la région des A-repeats identifiées par spectrométrie de masse.



J'ai vérifié la présence de ces protéines sur la région des A-repeats, ainsi que sur les régions amont et aval par la méthode de chromatographie d'affinité MBP/MS2 suivie de Western-blot. Les résultats sont présentés dans la **Figure 85**. Ces protéines sont bien présentes parmi les protéines en complexe avec la région des A-repeats. Cependant, elles ne sont pas spécifiques de la stricte région puisqu'elles sont également présentes dans les complexes formés avec les régions amont et aval. Les anticorps commerciaux utilisés contre ses protéines ne sont pas assez spécifiques et les tests d'immunoprécipitation réalisés par la suite n'ont rien donné. Des expériences supplémentaires ou avec d'autres anticorps seraient nécessaires afin de caractériser cette association.



**Figure 85 : Présence de protéines du complexe cohésine parmi les protéines purifiées avec la région des A-repeats**

Le protocole suivi est présenté dans la partie Matériel et Méthodes. Une quantité de 100 pmol de transcrits de la région murine fusionnés à l'ARN MS2, correspondant à la région amont de la région des A-repeats (5'Arep-MS2), à deux répétitions (2rep-MS2), à 4 répétitions (4rep-MS2), à la région des A-repeats entière (Arep-MS2) et à la région aval (3'Arep-MS2) ont été incubés séparément avec 200 µg d'extrait nucléaire de cellules HEK. Sur les 75 µL d'élution, 10 µL sont fractionnés par électrophorèse en gel SDS-PAGE puis révélés par Western-blot avec des anticorps contre les protéines indiquées à gauche. Un échantillon de chaque extrait nucléaire a été déposé à droite du gel (EN).

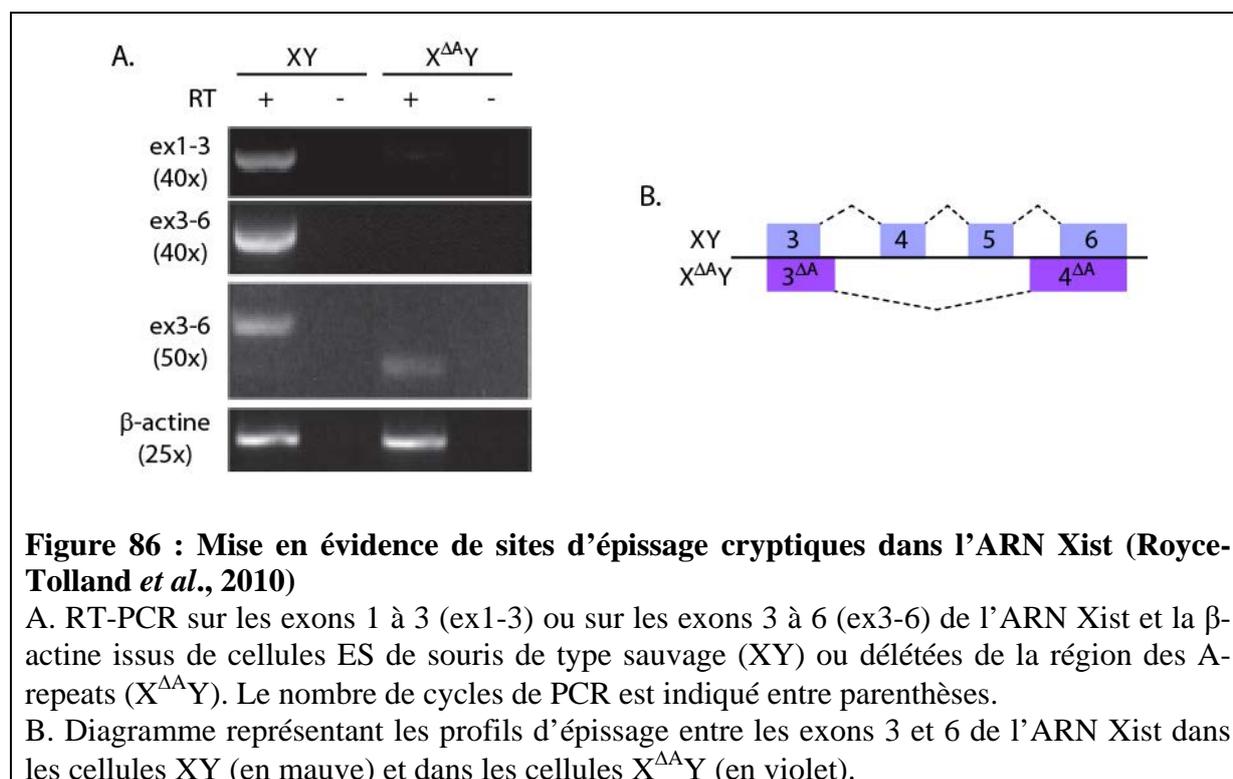
La région des A-repeats est associée à de nombreuses protéines, dont les fonctions sont variées mais dont le rôle dans l'inactivation du chromosome X reste pour la plupart à être déterminé. Si cette région permet le recrutement de facteurs protéiques directement impliqués dans les mécanismes mis en jeu dans l'inactivation, elle intervient également dans des processus moins directs, comme l'épissage de son propre ARN.

#### IV. Analyse du rôle de la région des A-repeats dans les processus de transcription et maturation de l'ARN Xist

##### IV.1 Etude de l'épissage de l'ARN Xist *in cellulo*

###### IV.1.1 Clonage des exons 2 à 6 du gène *Xist* dans le plasmide pSpliceExpress

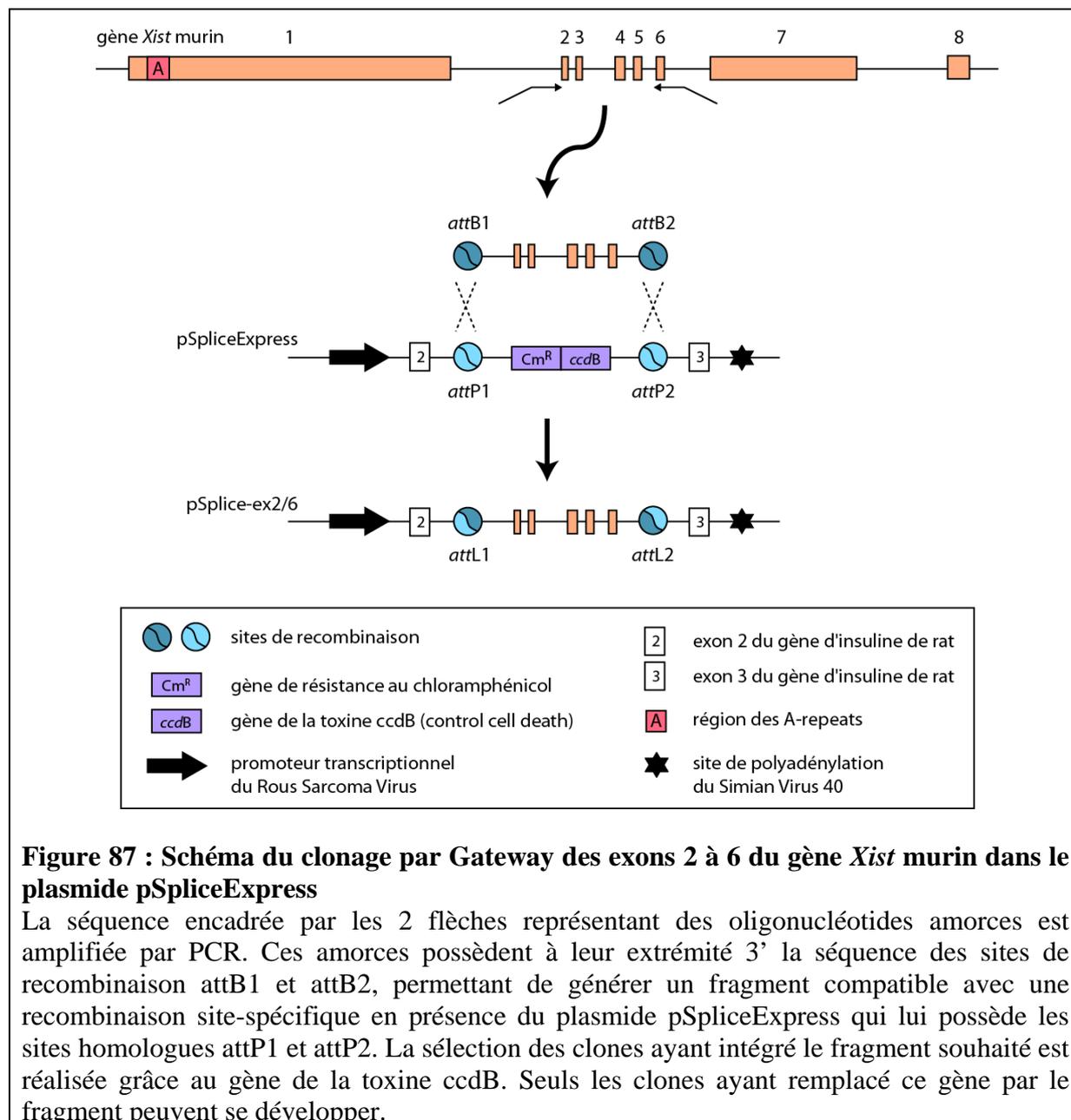
La région des A-repeats du gène *Xist* et de son ARN semble avoir plusieurs fonctions. Il a été postulé que cette région jouerait le rôle d'un « enhancer » de transcription. De plus, elle recrute des protéines du complexe PRC2 pour la mise en place de marques épigénétiques spécifiques du chromosome X inactif (Maenner *et al.*, 2010). Elle est également responsable du recrutement des gènes dans le territoire de Xist (Chaumeil *et al.*, 2006). Très récemment, un nouveau rôle a été mis en évidence : ASF/SF2 s'associe à cette région et l'absence ou la présence des A-repeats modifie l'épissage de l'ARN (Royce-Tolland *et al.*, 2010) (**Figure 86**).



Nous avons donc développé un modèle afin d'étudier le rôle de la région des A-repeats, d'ASF/SF2 ou d'autres protéines impliquées dans le « processing » de l'ARN (maturation, dégradation) des ARNm, dans la mise en place de l'inactivation. Le premier objectif fut de créer un minigène rapporteur et d'identifier des sites d'épissage alternatifs dans les séquences exoniques et introniques 2 à 6. Cette région inclut celle analysée par l'équipe de B. Panning qui a mis en évidence la présence d'un site 5' cryptique dans l'intron 3 et un site

3' cryptique dans l'intron 5, donnant lieu à un transcrit plus court dans lequel les exons 4 et 5 sont manquants (**Figure 86**).

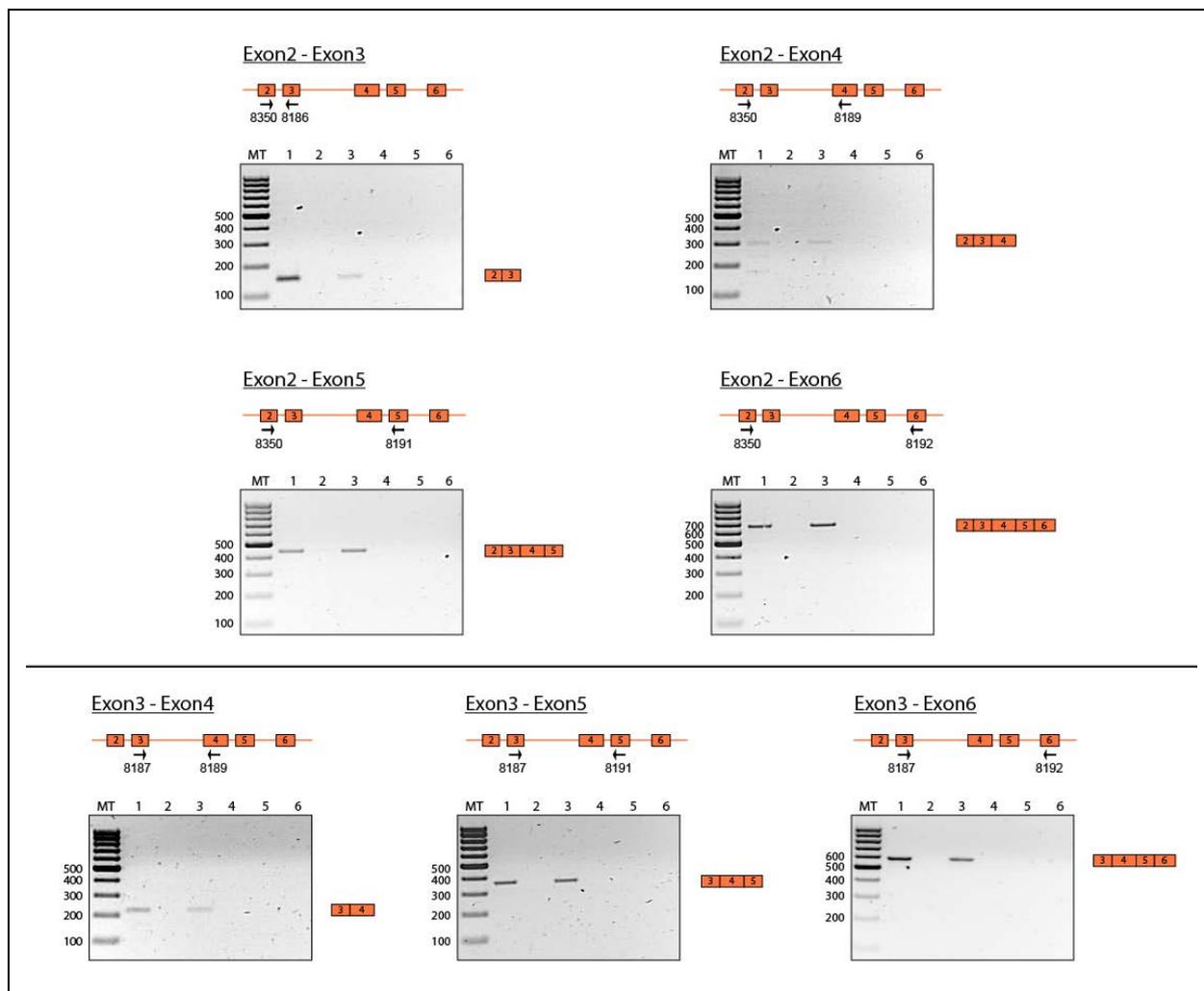
Pour cela, nous avons choisi de cloner la séquence correspondant aux exons 2 à 6 du gène *Xist* murin dans le plasmide pSpliceExpress. Ce plasmide nous a été fourni par S. Stamm qui l'avait déjà utilisé dans le but de créer un minigène rapporteur pour une analyse d'épissage (Kishore *et al.*, 2008). Le clonage dans ce plasmide fait appel à la technique Gateway, utilisant la recombinaison entre sites homologues spécifiques. Outre la rapidité de génération du minigène, le clonage par Gateway permet l'intégration de longs fragments avec une meilleure efficacité que le clonage classique. De plus, elle permet de s'affranchir de la sélection de sites de restriction, qui n'est pas une étape aisée avec des séquences longues de plusieurs kilobases. D'un point de vue fonctionnel, les résultats de profils d'épissage obtenus à partir de ce plasmide sont tout à fait similaires à ceux obtenus par un plasmide à clonage classique. Enfin, il a déjà été utilisé à plusieurs reprises pour des analyses de régulations de l'épissage alternatif et il a donné de très bons résultats (Kishore *et al.*, 2010 ; Sumanasekera *et al.*, 2012). Ce plasmide était donc parfaitement adapté à l'analyse de l'épissage entre les exons 2 à 6 de *Xist*. Le protocole suivi est présenté dans la partie Matériel et Méthodes et illustré dans la **Figure 87**.

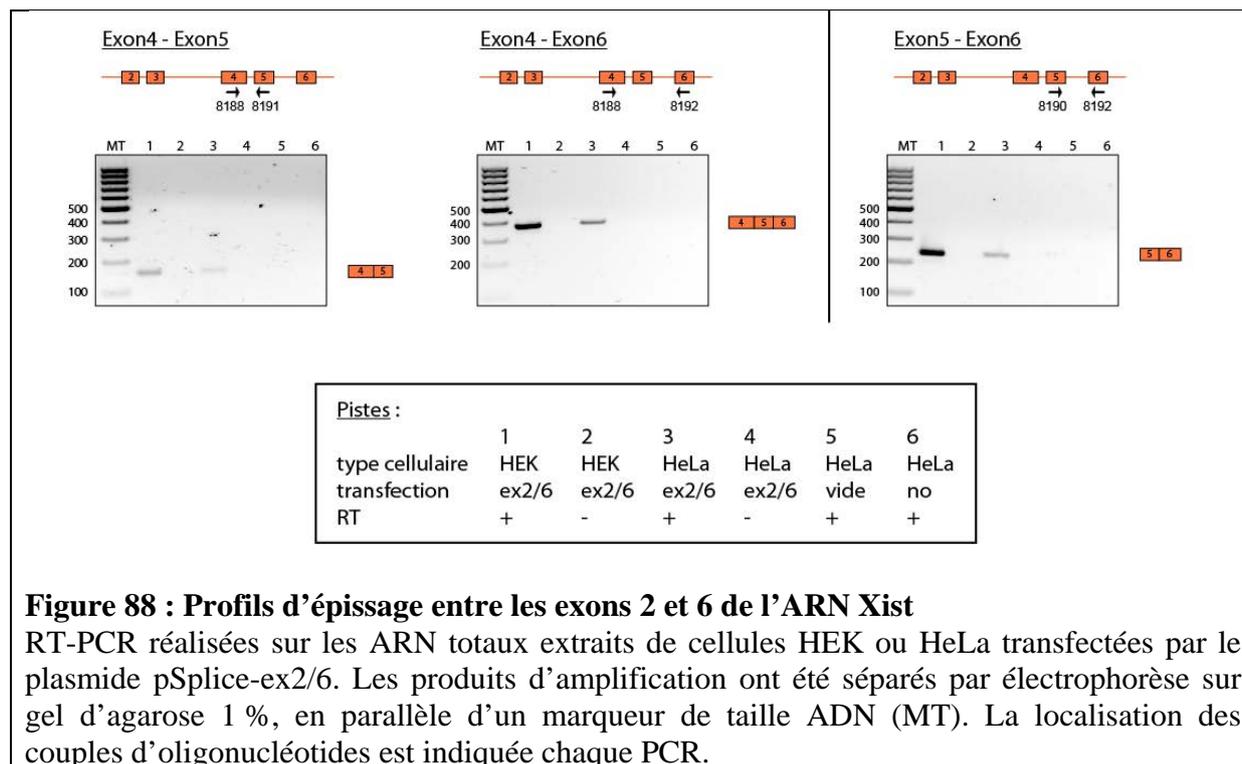


#### IV.1.2 Analyse des produits d'épissage entre les exons 2 et 6 en absence de la région des A-repeats

Le plasmide a ainsi créé été transfecté dans des cellules HEK ou HeLa afin d'exprimer l'ARN correspondant à l'exon 2 du gène de l'insuline, aux exons 2 à 6 de *Xist* et à l'exon 3 du gène de l'insuline. Après extraction des ARN totaux 48 heures après transfection, les ARN « *Xist* » sont analysés par RT-PCR. Les protocoles de transfection, extraction d'ARN et de RT-PCR sont décrits dans la partie Matériel et Méthodes (paragraphe II.5.4). Dans le but d'analyser le profil d'épissage entre tous les exons de *Xist*, j'ai réalisé plusieurs PCR utilisant des amorces qui couvrent la totalité des possibilités de combinaisons entre les différents exons

(Figure 88). Comme nous pouvons le constater, un seul produit d'épissage est amplifié dans chaque cas et qu'il correspond à l'épissage constitutif. Aucun intron n'est inclus et aucun site d'épissage alternatif n'est utilisé, quel que soit le type cellulaire utilisé. Nous avons réalisé l'expérience dans deux types cellulaires différents car ils ne possèdent pas le même caryotype. Les cellules HEK sont connues pour leur efficacité de transfection très élevée, nous avons donc débuté nos expériences avec ce type cellulaire. Cependant, ces cellules possèdent un caryotype complexe, contenant notamment trois copies du chromosome X. Afin de s'assurer que cette caractéristique n'influe pas sur nos résultats, nous avons reproduit l'expérience avec des cellules HeLa qui ne présentent pas cette particularité du nombre de chromosome X.





**Figure 88 : Profils d'épissage entre les exons 2 et 6 de l'ARN Xist**

RT-PCR réalisées sur les ARN totaux extraits de cellules HEK ou HeLa transfectées par le plasmide pSplice-ex2/6. Les produits d'amplification ont été séparés par électrophorèse sur gel d'agarose 1 %, en parallèle d'un marqueur de taille ADN (MT). La localisation des couples d'oligonucléotides est indiquée chaque PCR.

Par ailleurs, ces résultats vont à l'encontre de ceux obtenus par l'équipe de B. Panning. Dans son étude, cette équipe montre qu'en absence de la région des A-repeats, il se produit un épissage non constitutif entre les exons 3 et 6 de l'ARN Xist (Royce-Tolland *et al.*, 2010). Dans ce cas, l'épissage s'effectuerait entre un site 5' d'épissage en aval du site 5' constitutif de l'intron 3 et un site 3' d'épissage en amont du site constitutif de l'exon 5, supprimant les exons 4 et 5 du transcrit mûri (cf. **Figure 86**). Afin d'observer ce produit alternatif, la PCR avait été réalisée sur 50 cycles et aucun produit ne correspondant à un épissage constitutif n'était visible. Dans notre expérience, le produit de l'épissage constitutif apparaît dès 35 cycles. Plusieurs explications peuvent être proposées. D'une part, la différence est peut-être due au contexte physiologique d'une délétion de la région des A-repeats sur un chromosome X dans des cellules ES, comparé à la surexpression d'un fragment du gène *Xist* ne contenant pas cette région dans des cellules HEK. D'autre part, il manque dans notre modèle l'exon 1 et l'intron 1, qui peuvent eux aussi jouer un rôle dans la régulation de l'épissage. Nous pouvons donc conclure qu'en absence de l'exon 1 et de l'intron 1, l'épissage entre les exons 2 et 6 de Xist reste constitutif.

Puisque nous n'avons pas observé d'effet de l'absence de la région des A-repeats sur l'épissage des exons 2 à 6 de l'ARN Xist, nous avons ensuite voulu déterminer le rôle potentiel de cette région dans un environnement indépendant de Xist. Si cette région est en

effet un régulateur de l'épissage, elle doit également agir dans un contexte hétérologue, c'est-à-dire sur l'épissage d'exons d'un autre ARN. Pour cela, nous avons utilisé une construction hétérologue avec la séquence C2 de HIV-1.

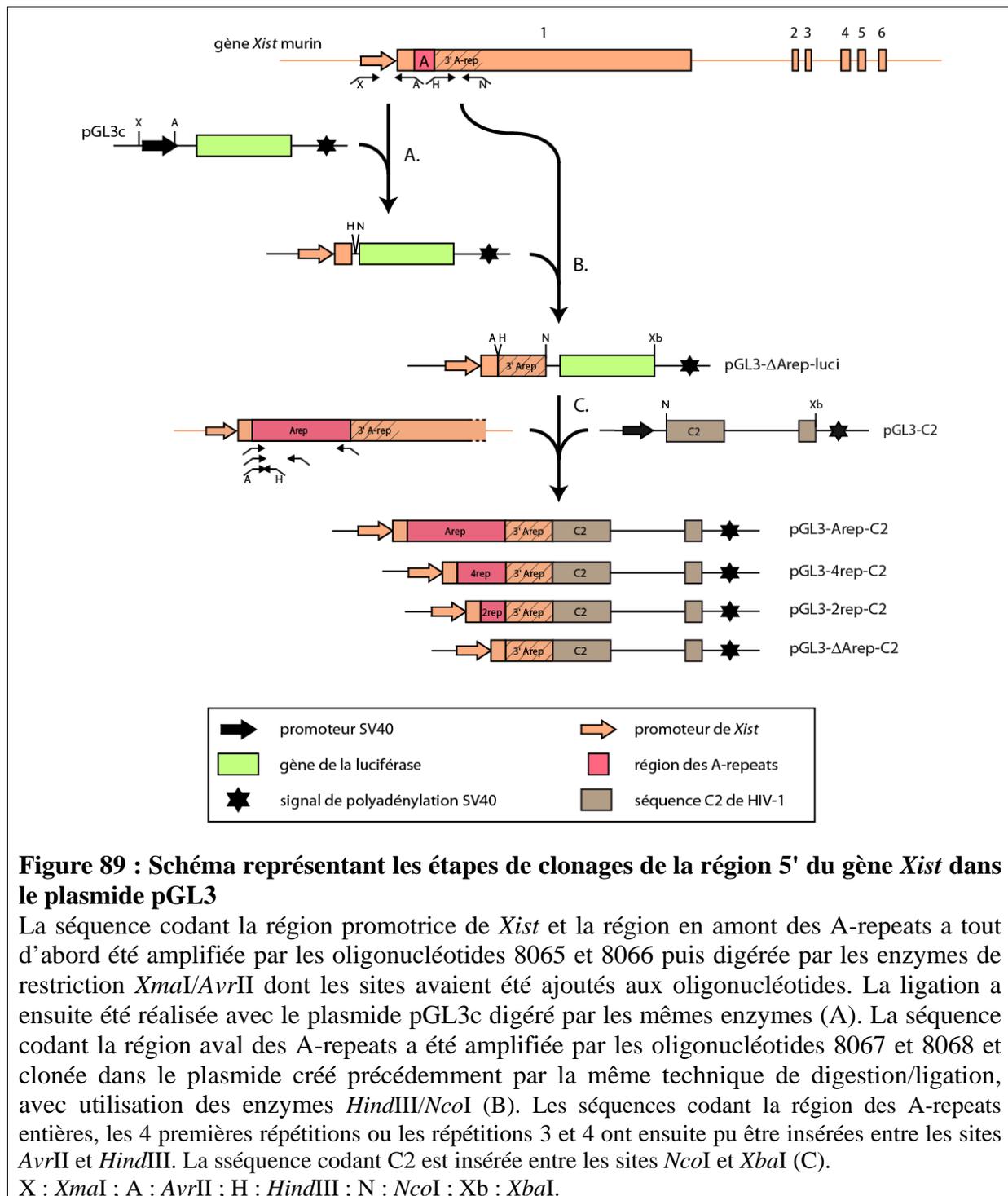
## IV.2 Etude de l'épissage de l'ARN C2 sous le contrôle de la région des A-repeats

La séquence C2 produit un pré-ARNm contenant un petit intron bordé des sites forts d'épissage 5' donneur D1 et 3' accepteur A2 (Damier *et al.*, 1997). L'épissage entre ces sites est connu pour être régulé par la fixation de protéines au niveau du site A2 (Bilodeau *et al.*, 2001 ; Ropers *et al.*, 2004). Cette construction a déjà été utilisée dans notre laboratoire dans le but d'étudier le régulateur négatif de l'épissage (NRS, Negative Regulator of Splicing) du virus de Sarcome de Rous, et il s'est révélé parfaitement adapté à ce genre d'étude (Bar *et al.*, 2011).

### IV.2.1 Construction du plasmide pGL3-Arep-C2

La région promotrice du gène *Xist* murin (du résidu -332 au résidu +1) ainsi que la partie 5' de l'exon 1 (du +1 au résidu +1080, incluant donc la région des A-repeats), ont été clonés dans le plasmide pGL3c en plusieurs étapes. Tout d'abord, la section correspondant à la région promotrice et à la partie en amont des A-repeats a été clonée après digestion par *XmaI/AvrII* (**Figure 89.A**). Ensuite, la séquence en aval de la région des A-repeats (357 paires de bases) a été clonée après digestion par *HindIII/NcoI* (**Figure 89.B**). Ce clonage en 2 parties nous a permis ensuite d'insérer des variants de la région des A-repeats entre les sites *AvrII* et *HindIII*. Nous y avons donc intégré la région entière (pGL3-Arep), 4 répétitions (pGL3-4rep) et 2 répétitions (pGL3-2rep). Nous avons également utilisé le plasmide sans répétition mais en incluant les séquences en amont et en aval des A-repeats (pGL3- $\Delta$ Arep).

A partir de ces 4 plasmides, j'ai ensuite inséré la séquence codant C2 entre les sites de restriction *XbaI* et *NcoI* se trouvant de part et d'autre du gène de la luciférase, créant ainsi les plasmides pGL3-Arep-C2 et ses variants de la région des A-repeats. Le plasmide pGL3-C2 a servi de plasmide contrôle lors des transfections. Tous ces plasmides sont présentés dans la **Figure 89.C**.

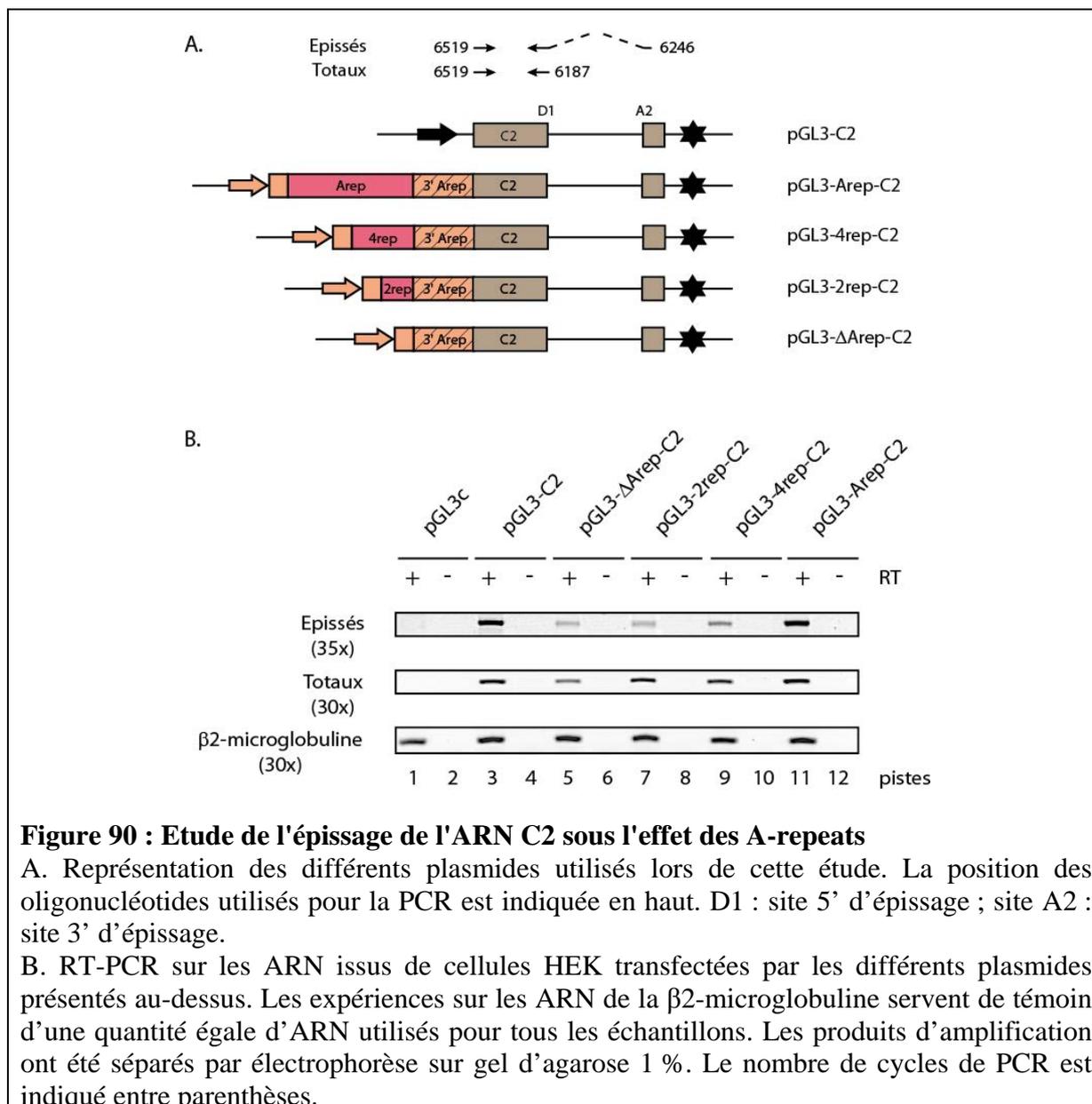


#### IV.2.2 Etude de l'épissage de C2 par RT-PCR

Comme indiqué dans la partie Matériel et Méthodes, des cellules HEK ont été transfectées par les différents plasmides présentés dans le panneau A de la **Figure 90**. Après 48 heures de culture, les ARN totaux ont été extraits et analysés par RT-PCR. L'étape de transcription inverse a été effectuée avec l'amorce 'dT' (15 résidus T). Notre plasmide contenant un signal de polyadénylation à l'extrémité 3' de la séquence clonée, les ARN issus

de ces constructions devraient être également inversement transcrits en ADNc. Pour la PCR, deux couples d'oligonucléotides amorces ont été utilisés afin de différencier les ARN épissés de ceux non épissés. Le couple 6519/6187 permet d'amplifier les ADNc des ARN épissés et non épissés (nommés « Totaux ») car les amorces s'hybrident dans l'exon 1 de la construction C2. Le couple 6519/6246 ne permet d'amplifier que les ADNc des ARN épissés car l'amorce 6246 s'hybride uniquement si les exons 1 et 2 sont contigus.

Les résultats obtenus sont présentés dans le panneau B de la **Figure 90**. Nous pouvons observer que la quantité d'ARN épissés mais aussi totaux varie selon le nombre de répétitions dans la région des A-repeats. Cette quantité diminue fortement lorsque C2 est sous le contrôle de la région promotrice de *Xist* en absence des A-repeats (piste 5), par rapport à celle obtenue avec C2 sous le contrôle du promoteur SV40 (piste 3). Elle augmente avec le nombre de répétitions (pistes 7, 9 et 11) jusqu'à atteindre celle obtenue avec la séquence C2 seule (piste 3). Comme ces résultats sont obtenus sur les ARN totaux (épissés et non épissés), nous pouvons en conclure que la région des A-repeats a un effet sur le taux de transcrits et que le nombre de répétitions est important. Elle peut agir sur la transcription en elle-même mais aussi sur la vitesse de dégradation des ARN. Nous approfondirons ce point dans un paragraphe ultérieur.

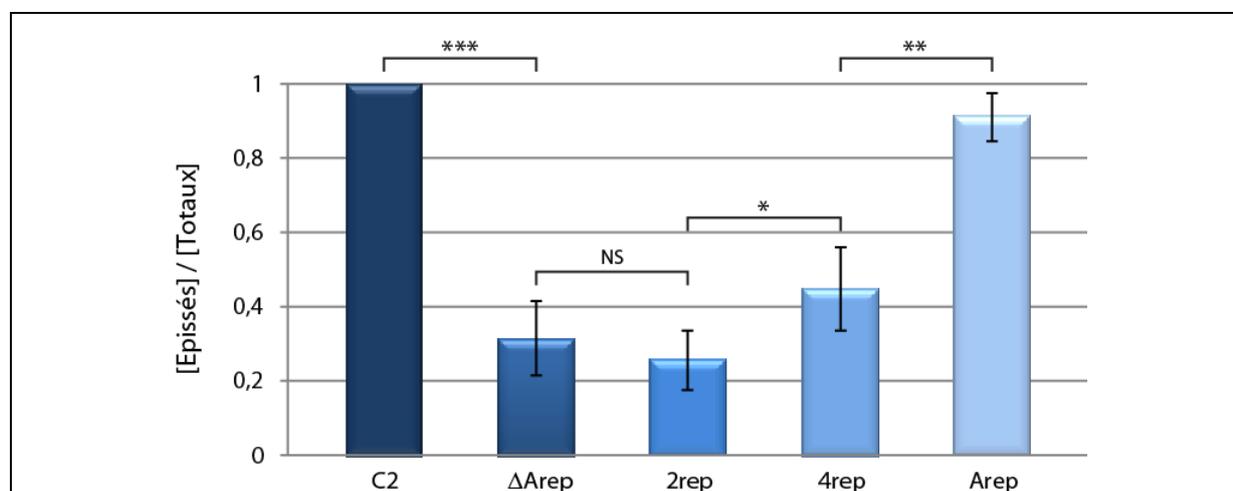


L'effet de la région des A-repeats sur l'épissage est plus difficile à examiner puisque les taux des ARN épissés et totaux semblent varier dans le même sens. Cependant, ces expériences ont été réalisées 3 fois et dans chaque cas, les bandes des produits PCR ont été quantifiées à l'aide du logiciel SynGene GeneTools. J'ai alors calculé le ratio [Epissés] / [Totaux] correspondant à chaque plasmide. Les valeurs obtenues sont présentées sous forme d'histogrammes dans la **Figure 91**.

Ce qui ressort de ces expériences est, d'une part, la forte diminution de l'épissage de C2 lorsque cette séquence est sous le contrôle de la région promotrice de *Xist* sans la région des A-repeats. D'autre part, la présence des 8 répétitions de la région des A-repeats rétablit le taux d'épissage atteint sous le contrôle d'un promoteur fort. Nous pouvons en conclure que la

région des A-repeats joue un rôle de régulateur de l'épissage dans un contexte hétérologue, bien qu'elle soit localisée bien en amont du site 5' donneur. Cela constitue une régulation non décrite à ce jour dans la littérature.

Par ailleurs, le nombre de répétitions semble également important. La présence de 4 répétitions permet d'atteindre 50 % du taux maximal d'épissage, tandis que l'épissage en présence de 2 répétitions est du même ordre qu'en absence de répétitions (**Figure 91**, 4rep et 2rep respectivement), bien que la significativité des résultats ne soit pas élevée. Nous pourrions mettre ces observations en relation avec la structure de la région. Comme nous l'avons présenté dans la première partie des Résultats, la région des A-repeats adopte une structure formée de 2 grandes tiges-boucles comprenant chacune 4 répétitions. La présence des 2 tiges-boucles serait la structure idéale pour la fonction de cette région dans régulation de l'épissage, par exemple par recrutement de différents facteurs d'épissage. Si une seule structure tige-boucle est formée, la région peut tout de même exercer sa fonction mais celle-ci sera plus faible car moins de facteurs pourront être recrutés. Enfin, la présence de 2 répétitions n'est certainement pas suffisante pour former une structure permettant le recrutement de ces facteurs.



**Figure 91 : Variation du taux des transcrits épissés par rapport à celui des transcrits totaux (épissés et non épissés) selon le nombre de répétitions de la région des A-repeats**

A partir des électrophorèses en gel d'agarose des produits PCR obtenus avec les couples 6519/6246 et 6519/6187, les bandes ont été quantifiées à l'aide du logiciel SynGene GeneTools. Le ratio [Epissés] / [Totaux] a ensuite été calculé pour chaque échantillon puis normalisé de manière à ce que le ratio de C2 soit égal à 1. Chaque barre de l'histogramme représente la moyenne de 3 expériences. Les écarts-types sont représentés par les barres verticales et les p-values sont représentées par les étoiles :  $p < 0,01$  \*\*\* ;  $0,01 < p < 0,05$  \*\* ;  $0,05 < p < 0,1$  \* ;  $p > 0,1$  NS (non significatif).

### IV.3 Etude du rôle de la région des A-repeats dans l'épissage de l'ARN Xist

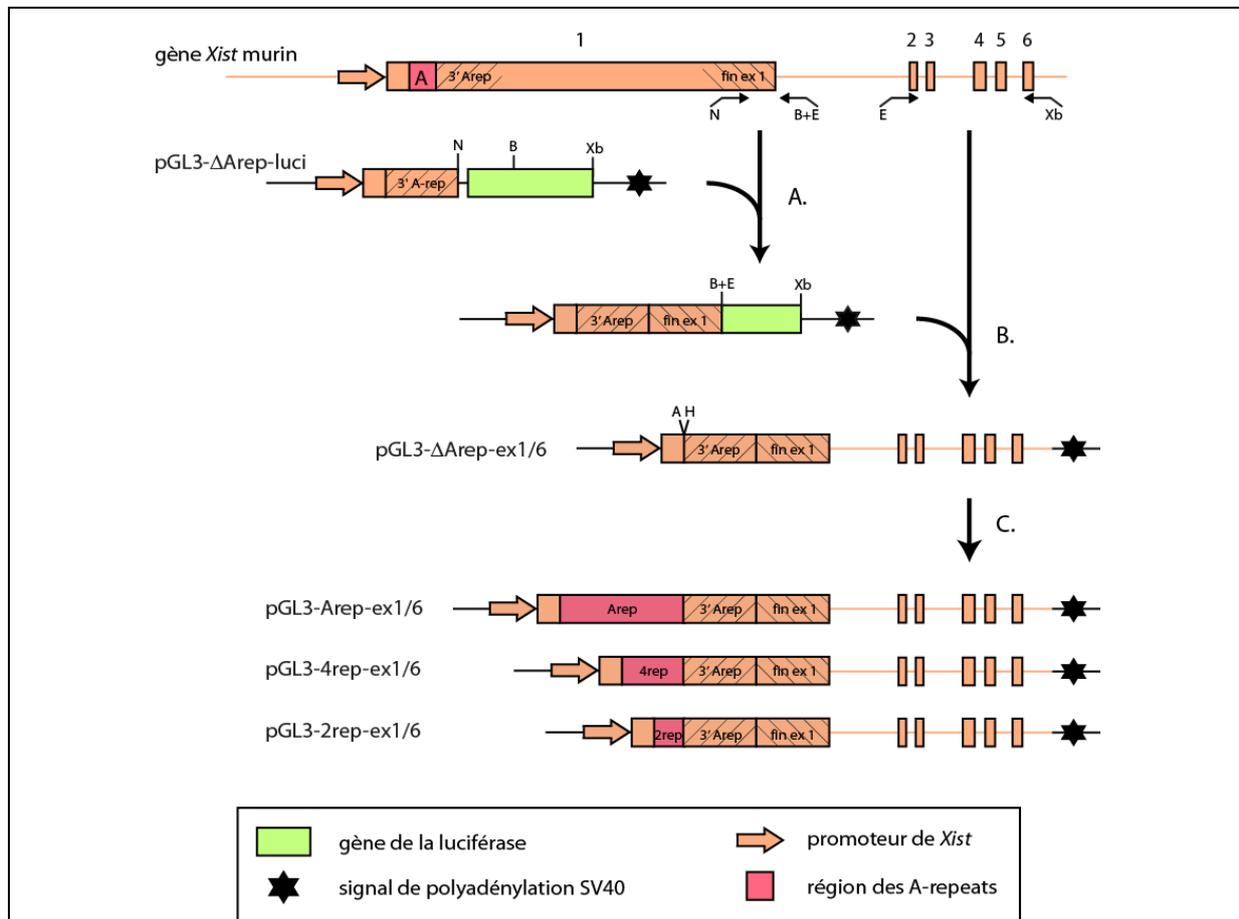
#### IV.3.1 Clonage du gène *Xist* murin dans le plasmide pGL3

La région des A-repeats pouvant réguler l'épissage dans un système hétérologue, nous avons ensuite cherché à étudier son rôle dans l'épissage des exons de l'ARN Xist. Il a donc fallu créer un plasmide contenant tous ces éléments. La longueur du gène *Xist* a nécessité de le tronquer de plusieurs régions, nous n'avons donc pas utilisé la technique de Gateway, mais la technique de clonage classique. Nous avons déjà la région promotrice (-332 à +1) ainsi que la partie 5' de l'exon 1 (+1 à +1080) dans le plasmide pGL3- $\Delta$ Arep-luci (cf. **Figure 89.B**), nous avons donc utilisé ce plasmide pour créer notre minigène.

- L'exon 1 étant long d'environ 10 kb, nous avons décidé d'en supprimer la partie centrale et de ne garder que les 318 derniers résidus, c'est-à-dire une région assez grande pour contenir les différents éléments essentiels à l'épissage.
- De même pour l'intron 1 de 2800 nucléotides, nous nous sommes limités à ses extrémités 5' et 3' (200 nucléotides pour chaque).
- En revanche, les exons 2 à 6 étant courts et rapprochés, nous avons pu les cloner ensemble. Cependant son clonage a nécessité de nombreuses tentatives car cette région est longue de 2600 nucléotides.

Le choix des sites de restriction n'a pas été une étape simple. En effet, les séquences clonées étant très longues, plusieurs sites de restriction naturels se trouvaient déjà présents et n'étaient donc pas utilisables. De plus, dans une première stratégie, nous avons pensé cloner les séquences correspondant à la fin de l'exon1/début de l'intron 1 et les exons 2 à 6 lors d'une même étape de ligation à 3 partenaires avec le plasmide. Cela nous permettait d'utiliser le site *EcoRI*, un des rares sites non présents dans les séquences à cloner mais absent dans le plasmide pGL3. Toutefois cette stratégie s'est révélée inefficace et nous avons dû en changer pour un clonage classique [1 séquence à cloner] pour [1 plasmide]. La séquence des exons 2 à 6 ayant déjà été difficilement sous-clonée dans le plasmide pCR2.1<sup>®</sup> avec la séquence du site *EcoRI* à son extrémité 5', nous avons rajouté un site *BclII* en aval du site *EcoRI* dans l'oligonucléotide antisens pour l'amplification de la séquence « fin exon 1-début intron 1 ». Ainsi cette séquence a finalement pu être clonée dans le plasmide pGL3- $\Delta$ Arep-luci entre les sites *NcoI* et *BclII*, apportant par la même un site *EcoRI*, puis la séquence des exons 2 à 6 a été clonée entre les sites *EcoRI* et *XbaI*. Les étapes du clonage sont détaillées dans la **Figure 92**.

Les plasmides ainsi créés contiennent le promoteur du gène *Xist*, la région en aval des A-repeats, la région en aval, l'extrémité 3' de l'exon 1, l'intron 1 tronqué de sa partie centrale et les exons 2 à 6. Le plasmide pGL3- $\Delta$ Arep-ex2/6 ne contient rien de la région des A-repeats, le pGL3-Arep-ex2/6 contient la région entière, le pGL3-4rep-ex2/6 les 4 premières répétitions et le pGL3-2rep-ex2/6 les deux premières.



**Figure 92 : Schéma des différentes étapes de clonage du gène *Xist* murin tronqué dans le plasmide pGL3**

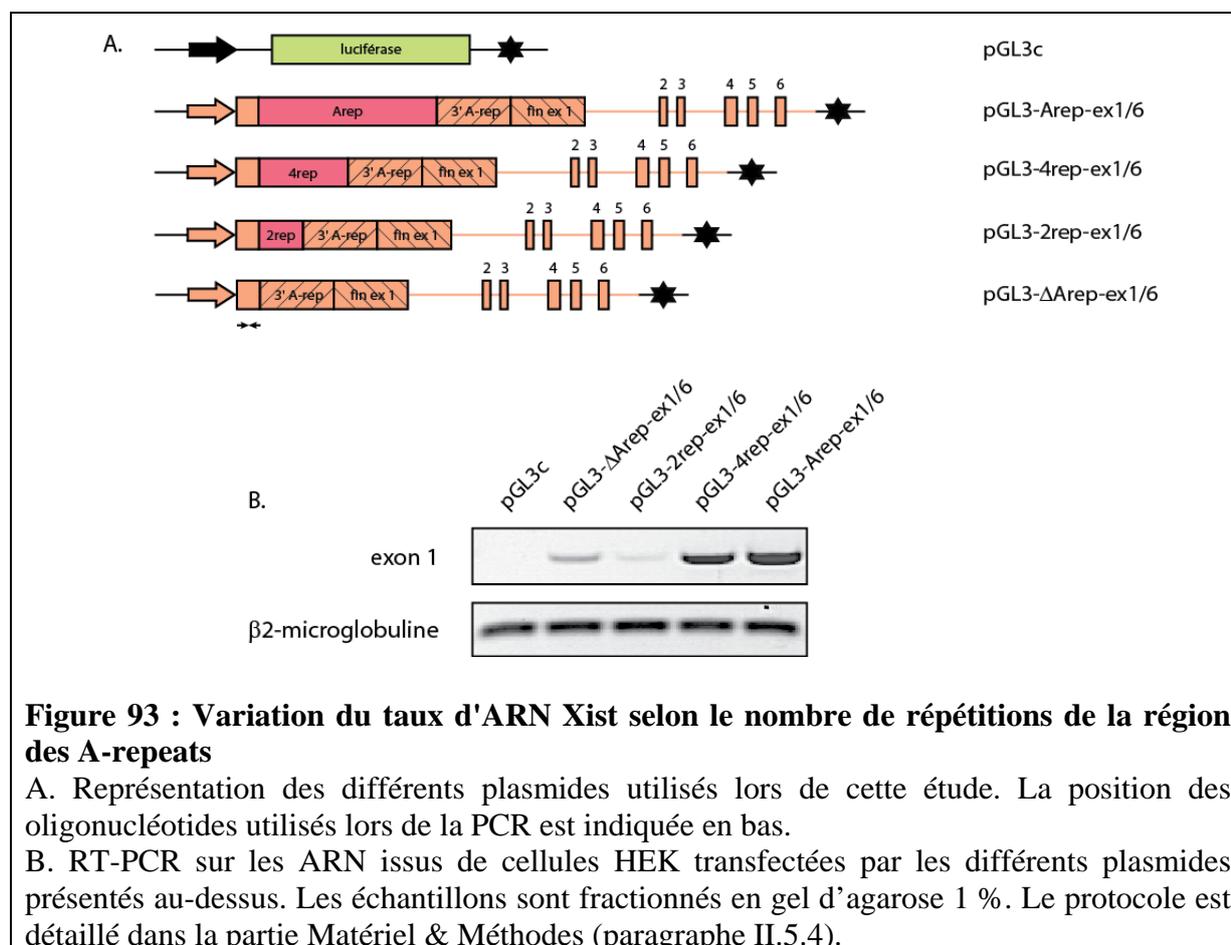
La séquence correspondant aux 318 dernières paires de bases de l'exon 1 de *Xist* et les 202 premières paires de bases de l'intron 1 a été amplifiée par les oligonucléotides 8069 et 8219 auxquels ont été ajoutés les sites de restriction de *NcoI* (8069) et *BclI* et *EcoRI* (8219). Le produit ainsi que le plasmide pGL3- $\Delta$ Arep-luci ont été digérés par les enzymes de restriction *NcoI* et *BclI* puis ligués entre eux (A). La séquence codant mes exons 2 à 6 ont ensuite été amplifiés par PCR avec les oligonucléotides 8071 (rallongé du site *EcoRI*) et 8072 (rallongé du site *XbaI*) pour être cloné dans le plasmide créé précédemment (B). Enfin, les séquences codant la région des A-repeats entières, les 4 premières répétitions ou les répétitions 3 et 4 ont été insérées entre les sites *AvrII* et *HindIII* (C).

N : *NcoI* ; B : *BclI* ; E : *EcoRI* ; A : *AvrII* ; H : *HindIII* ; Xb : *XbaI*.

## IV.3.2 Analyse de l'épissage entre les exons 1 et 6 de l'ARN Xist

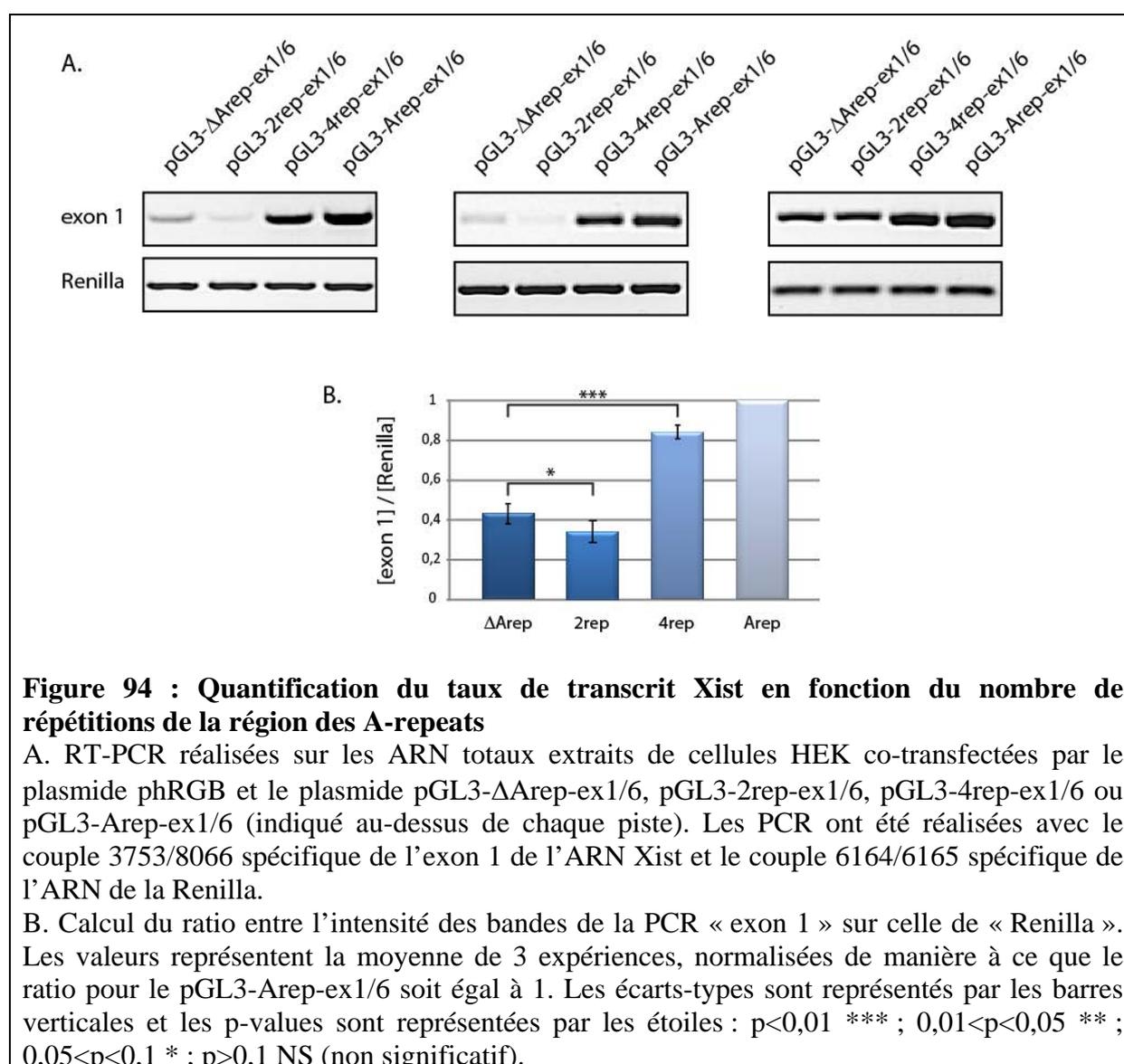
## a. Analyse du taux de transcrits Xist

Les différentes constructions ont été transfectées dans des cellules HEK, afin d'en extraire les ARN 48 heures après qui ont été analysés par RT-PCR avec oligonucléotides spécifiques. Dans un premier temps, j'ai réalisé une PCR de contrôle amplifiant une région commune à tous les ARN issus d'un plasmide contenant la séquence *Xist*. Les oligonucléotides s'hybridant au niveau de l'exon 1, en amont de la région des A-repeats, la PCR donne une information sur le taux d'ARN Xist incluant toutes les formes possibles. Les résultats obtenus, présentés dans la **Figure 93**, nous ont surpris car nous nous attendions à une intensité de bande constante. La variation observée entre les différentes pistes ne peut pas être due à une variation de la quantité d'ARN utilisée pour la RT puisque les PCR réalisées sur l'ARN de la  $\beta$ 2-microglobuline montrent un signal constant.



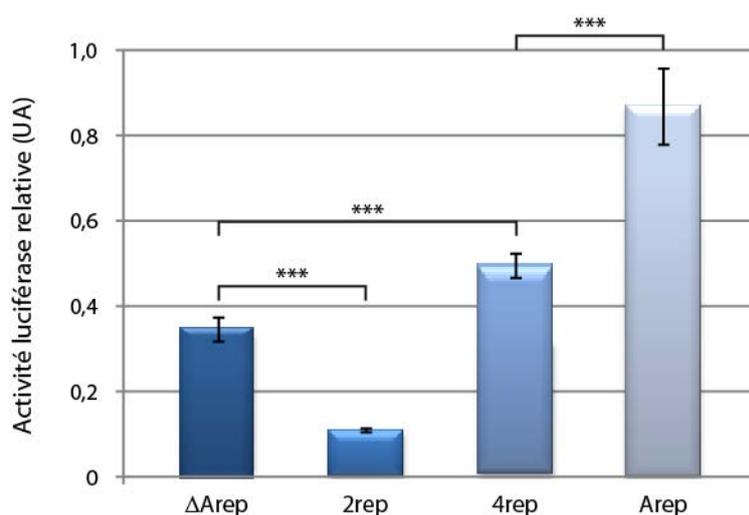
J'ai donc reproduit l'expérience avec de nouvelles dilutions de plasmides et en co-transfectant avec un plasmide contrôle permettant de vérifier l'efficacité de la transfection. J'ai opté pour le plasmide phRGB portant la séquence codant la luciférase Renilla. Comme le

montre la **Figure 94.A**, le profil de RT-PCR réalisé sur l'ARN de la Renilla ne présente aucune variation selon le plasmide pGL3 co-transfecté, donc l'efficacité de transfection des différents échantillons est très proche. En revanche, on observe que le profil de l'exon 1 est le même que celui obtenu précédemment. L'expérience a été réalisée 3 fois et j'ai quantifié l'intensité des bandes (**Figure 94.B**). Nous pouvons nettement observer que le nombre de répétitions influe fortement sur le taux de transcrit. La présence d'aucune ou de 2 répétitions ne permet même pas d'atteindre 50 % du taux de transcrit obtenu avec la région entière. La présence de 4 répétitions donne un résultat proche de celui avec la région entière, ce qui suggère que les éléments requis pour un taux élevé sont présents.



Afin de confirmer ces résultats, je me suis appuyée sur les constructions pGL3-Arep-lucif (ΔArep, 2rep, 4rep et Arep), produites lors de la création des plasmides pGL3-Arep-C2

(cf. **Figure 89.B**). Nous avons utilisé ces constructions en co-transfection avec le plasmide phRGB codant pour la luciférase Renilla. Grâce au système Dual-Luciferase<sup>®</sup> Reporter (Promega), les activités de la luciférase Firefly (celle du pGL3) et de la luciférase Renilla (celle du phRGB) peuvent être mesurées séquentiellement à partir d'un même échantillon. Le protocole qui a été suivi est présenté dans le Matériel & Méthodes. J'ai ensuite calculé le ratio [Firefly] / [Renilla] des valeurs d'activité. Les résultats sont présentés sous forme de graphique dans la **Figure 95**. Le profil obtenu est similaire à celui observé par quantification des bandes de RT-PCR et confirme donc les résultats précédents. En effet, la présence de 2 répétitions fait baisser l'activité luciférase, et donc la quantité de transcrits, par rapport à une absence totale de répétitions. De même, la présence de 4 répétitions a bien un effet positif, mais ne permet pas d'atteindre une valeur proche de celle obtenue avec la région entière.



**Figure 95 : Variation l'activité de la luciférase Firefly selon le nombre de répétitions de la région des A-repeats**

Des cellules HEK ont été co-transfectées par le plasmide phRGB et l'un des plasmides pGL3-ΔArep-luci, pGL3-2rep-luci, pGL3-4rep-luci ou pGL3-Arep-luci, indiqué sous chaque barre d'histogramme. Les activités luciférase de la Firefly et de la Renilla ont été mesurées selon le protocole décrit dans la partie Matériel et Méthodes. L'expérience a été réalisée 3 fois ; les valeurs représentent la moyenne du ratio des activités [Firefly] / [Renilla]. Les écarts-types sont représentés par les barres verticales et les p-values sont représentées par les étoiles.

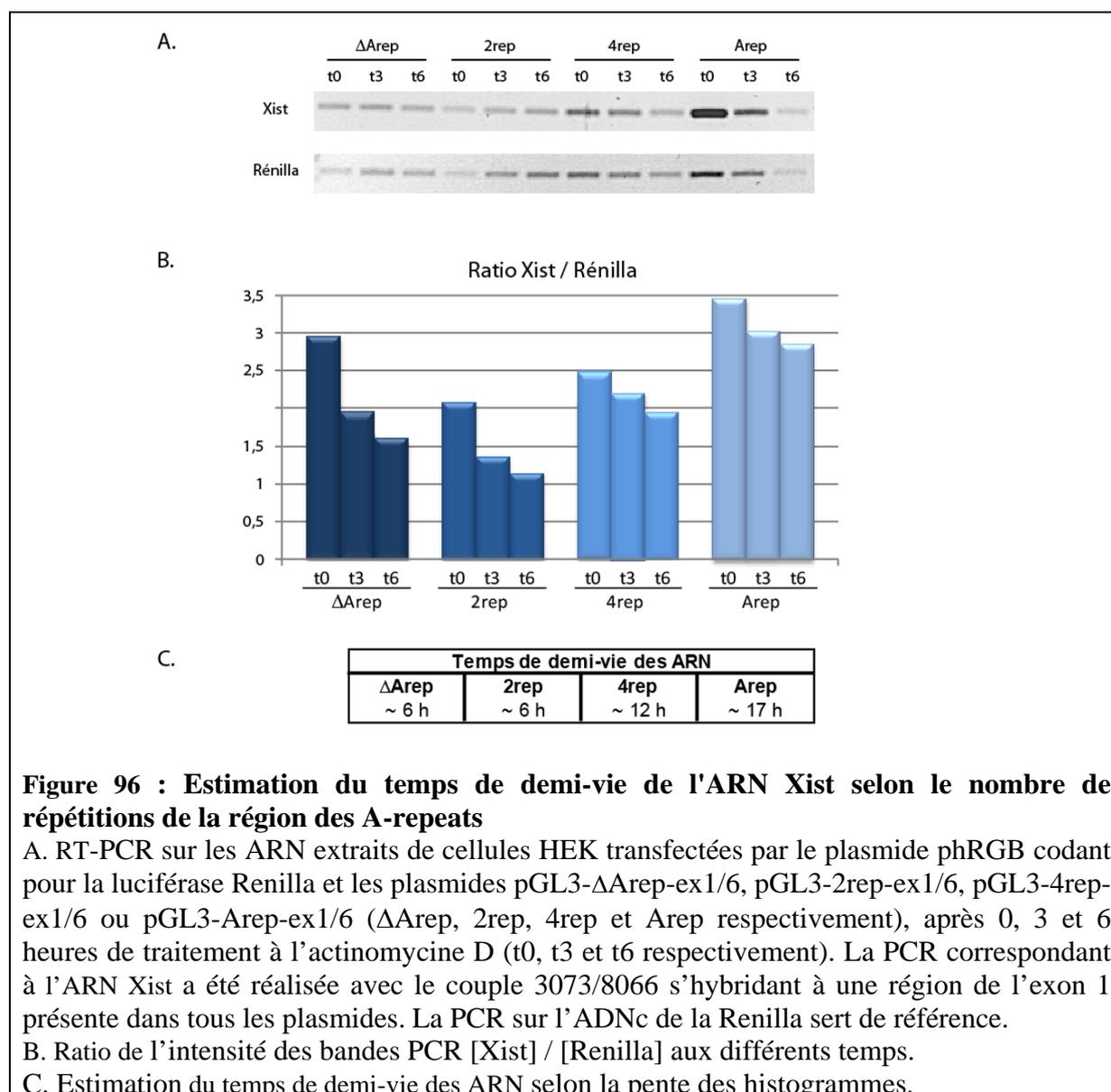
Ce résultat est cohérent avec la régulation de l'expression de *Xist* par la région des A-repeats que nous avons mentionnée dans le paragraphe IV.3.2.c de l'Introduction. En effet, une délétion de cette région sur le chromosome paternel chez la souris conduisait une diminution de la quantité de l'ARN *Xist* (Hoki *et al.*, 2009). Mis à part l'effet des A-repeats en lui-même, cette observation confirme que notre système d'étude convient et qu'il reflète au

moins en partie les situations et régulations naturelles. La régulation peut s'effectuer au niveau transcriptionnel mais il est également possible que ce soit la stabilité de l'ARN qui soit régulée. Bien qu'il ait été montré que l'ARN Tsix ne déstabilise pas l'ARN Xist (Sun *et al.*, 2006), il n'est pas exclu que la région des A-repeats, elle, augmente sa stabilité.

Afin de vérifier si l'action des A-repeats s'effectue au niveau de la stabilité de l'ARN Xist, nous avons cherché si le temps de demi-vie de ce dernier variait en absence de la région et selon le nombre de répétitions.

#### b. Analyse du temps de demi-vie de l'ARN Xist

Afin de déterminer si la vitesse de dégradation d'un ARN varie selon les conditions, il est possible d'arrêter la transcription par ajout d'actinomycine dans le milieu de culture et de regarder le taux de transcrits à différents temps. J'ai utilisé cette technique avec des cellules co-transfectées par le plasmide phRGB et le plasmide pGL3 contenant les variants du nombre de répétitions de la région des A-repeats. Les ARN ont été extraits avant traitement à l'actinomycine, puis après 3 et 6 heures de traitement. L'analyse du temps de demi-vie a été réalisée par quantification des bandes obtenues par RT-PCR (**Figure 96.A**) et calcul du ratio [Xist] / [Renilla] pour chaque échantillon (**Figure 96.B**). Nous retrouvons bien le profil de taux de transcrits obtenu précédemment (à savoir une baisse du taux avec 2 répétitions par rapport à une absence de répétition puis une augmentation avec 4 puis 8 répétitions). Ce que nous observons par ailleurs, c'est que le taux de transcrit diminue plus fortement entre 0 et 6 heures de traitement en présence de 0 ou 2 répétitions qu'en présence de 4 ou 8 répétitions (soit la région entière). La **Figure 96.C** présente les temps de demi-vie approximatifs, estimés par calcul de la vitesse de diminution de la quantité de transcrits Xist. Il semblerait donc que le temps de demi-vie soit fortement diminué si la région des A-repeats ne comporte pas au minimum 4 répétitions. Ces résultats expliquent ceux obtenus par RT-PCR sur l'exon 1 et par activité luciférase.



Cette expérience n'a été réalisée qu'une fois, et il serait intéressant de la reproduire avec une durée de traitement plus longue, afin d'estimer le temps de demi-vie plus avec une meilleure précision. De même, une analyse par RT-qPCR donnerait des résultats plus précis. Cependant, nous avons montré que la région des A-repeats est importante dans la régulation du taux de l'ARN Xist en agissant sur sa stabilité, bien qu'une régulation au niveau transcriptionnel supplémentaire ne peut pas être exclue. Si la régulation se joue sur le promoteur du gène *Xist*, il serait intéressant de reproduire les mêmes expériences mais avec un plasmide ne contenant pas le promoteur *Xist* (remplacé par le promoteur SV40 par exemple, promoteur initial du plasmide pGL3). De même, si la méthylation du promoteur est un élément essentiel de la régulation comme l'équipe de Sado a pu le montrer (Hoki *et al.*, 2009), nous pourrions étudier les effets d'une méthylation *in vitro* du plasmide.

D'une manière générale, l'ensemble de ces résultats concordent une fois de plus avec la structure adoptée par la région des A-repeats en 2 grandes tiges-boucles formées de 4 répétitions successives appariées 2 à 2. La présence des 2 tiges-boucles est la structure optimale pour les différents rôles de la région, une seule tige-boucle permet d'assurer tout de même ces fonctions mais en-dessous de 4 répétitions, la structure formée n'est pas en adéquation avec une activité correcte de la région.

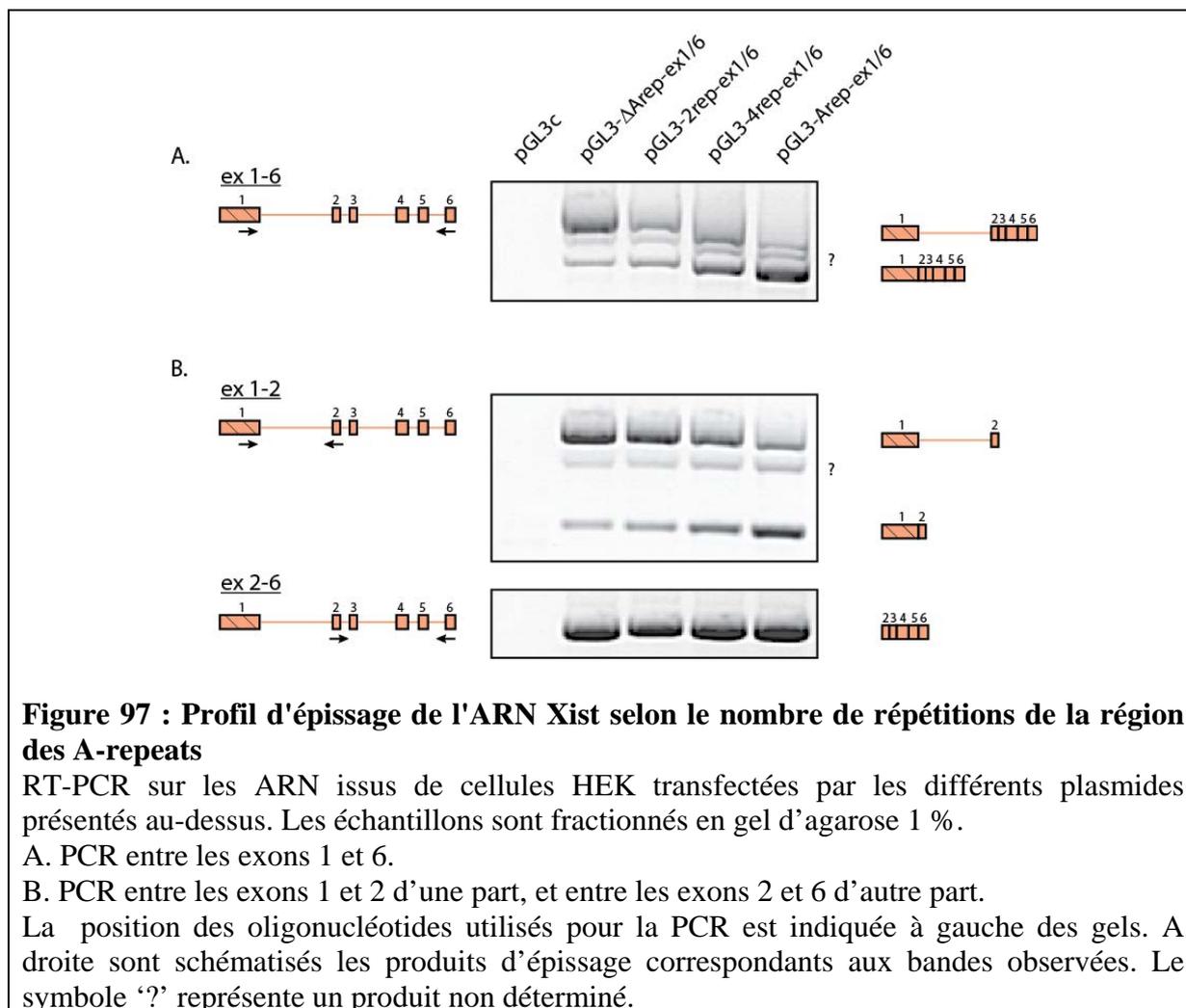
Nous avons donc montré un rôle de la région des A-repeats dans la régulation du taux de transcrits Xist, nous avons ensuite étudié son rôle dans celle de l'épissage.

### c. Analyse de l'épissage de l'ARN Xist

Afin de vérifier l'influence de la région des A-repeats et de ses mutants sur l'épissage de l'ARN Xist, des cellules HEK ont été transfectées par les différentes constructions, les ARN totaux ont été extraits et dans un premier temps, j'ai réalisé des RT-PCR afin de détecter les ARN Xist englobant les 6 exons (**Figure 97.A**). Nous pouvons observer la présence de 3 bandes pour tous les échantillons, excepté pour le témoin négatif. Pour identifier les différents produits, l'ADN de chaque bande a été extrait, cloné dans le plasmide pCR2.1<sup>®</sup> et séquencé. La bande inférieure correspond aux exons 1 à 6 constitutivement épissés (comme le laissait supposer la taille du fragment). La bande supérieure correspond aux exons 1 à 6 avec inclusion de l'intron 1. Le séquençage de la bande intermédiaire, quant à elle, est similaire à celui de la bande inférieure. Les séquençages sont présentés en partie Annexe.

En absence des A-repeats, le produit comprenant l'intron 1 est majoritaire par rapport à l'ARN constitutivement épissé, puis le rapport s'inverse en fonction de l'augmentation du nombre de répétitions. La région des A-repeats jouerait donc un rôle dans l'épissage de l'ARN Xist, au moins sur l'inclusion ou pas de l'intron 1. Notons tout de même la persistance de la bande supérieure dans le cas où la région entière est présente.

Afin de vérifier que l'intron 1 est le seul à être inclus comme l'a démontré le séquençage, j'ai réalisé des PCR entre les exons 1 et 2 d'une part et entre les exons 2 et 6 d'autre part (**Figure 97.B**). Nous avons bien la confirmation de l'inclusion de l'intron 1 en absence de la région des A-repeats (ex 1-2). La PCR réalisée entre les exons 2 et 6 (ex 2-6) ne présente qu'une seule bande correspondant en taille au produit de l'épissage constitutif, ce qui a été confirmé par séquençage. Le rôle de la région des A-repeats semble donc se concentrer sur la régulation de l'épissage de l'intron 1.



Rappelons que les profils d'épissage obtenus après transfection avec le plasmide pSplice-ex2/6 ne présentaient aucune forme non épissée (cf. **Figure 88**). Le fait d'avoir transfecté les cellules avec un plasmide contenant également le promoteur *Xist* et une partie de l'exon 1 a provoqué l'apparition d'une forme non épissée, justement entre les exons 1 et 2, tandis que l'épissage entre les exons 2 et 6 est resté constitutif. Il se pourrait donc qu'il manque des séquences nécessaires à l'épissage correct entre les exons 1 et 2. Ces séquences se situeraient soit dans la partie tronquée de l'exon1, soit dans celle de l'intron 1.

Par ailleurs, nous observons à nouveau le profil du taux d'ARN Xist variant selon le nombre de répétitions, facilement visible lorsque la région entre les exons 2 et 6 a été amplifiée. En ce qui concerne les PCR entre les exons 1 et 6 et entre les exons 1 et 2, le profil est retrouvé lorsque nous additionnons l'intensité des différents produits. Cela confirme une fois de plus les résultats précédents.

La région des A-repeats serait donc impliquée dans la régulation de l'épissage. Cependant, l'épissage est un processus complexe qui fait intervenir de nombreuses protéines, auxquelles se rajoutent les facteurs régulateurs. Afin de rechercher quels partenaires protéiques seraient impliqués dans la régulation de l'épissage de l'ARN Xist, nous avons également analysé l'effet de plusieurs protéines sur l'épissage de cet ARN.

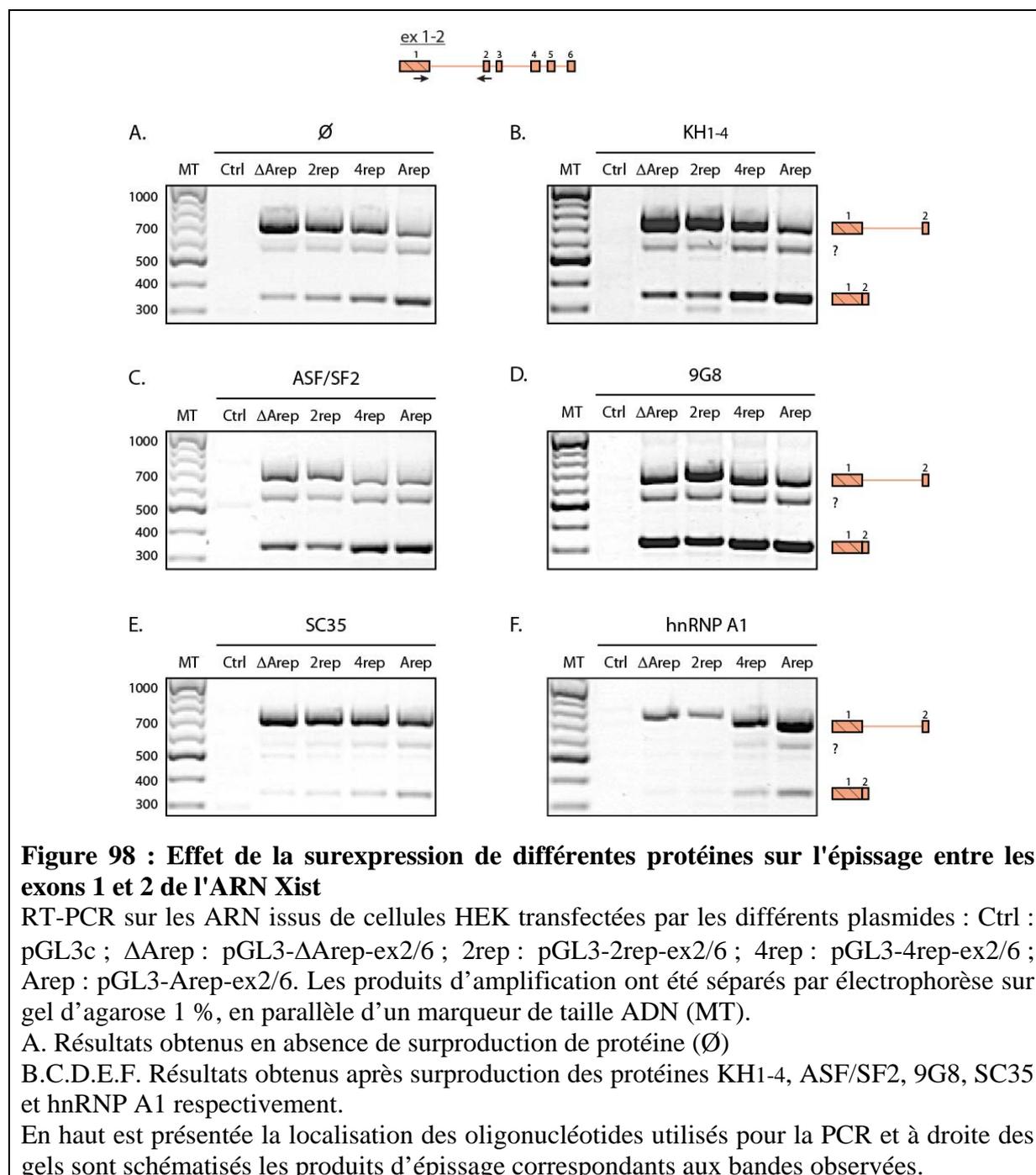
#### IV.4 Analyse du rôle de différentes protéines sur l'épissage de l'ARN Xist

Etant donné le nombre important de facteurs pouvant intervenir dans la régulation de l'épissage, nous nous sommes concentrés sur ceux qui nous paraissaient le plus susceptibles d'intervenir dans le cas de l'ARN Xist. Nous avons donc étudié l'effet de la protéine ASF/SF2 qui est un régulateur d'épissage très courant et dont nous avons mis en évidence son interaction avec la région des A-repeats. Nous avons également analysé le rôle de la protéine KH1-4, étudiée dans le paragraphe précédent. Enfin, puisque nous avons à notre disposition au laboratoire d'autres vecteurs codant pour des protéines régulatrices, nous avons étudié l'influence des protéines 9G8, SC35 et hnRNP A1, cette dernière étant souvent décrite pour avoir une action antagoniste à celle d'ASF/SF2.

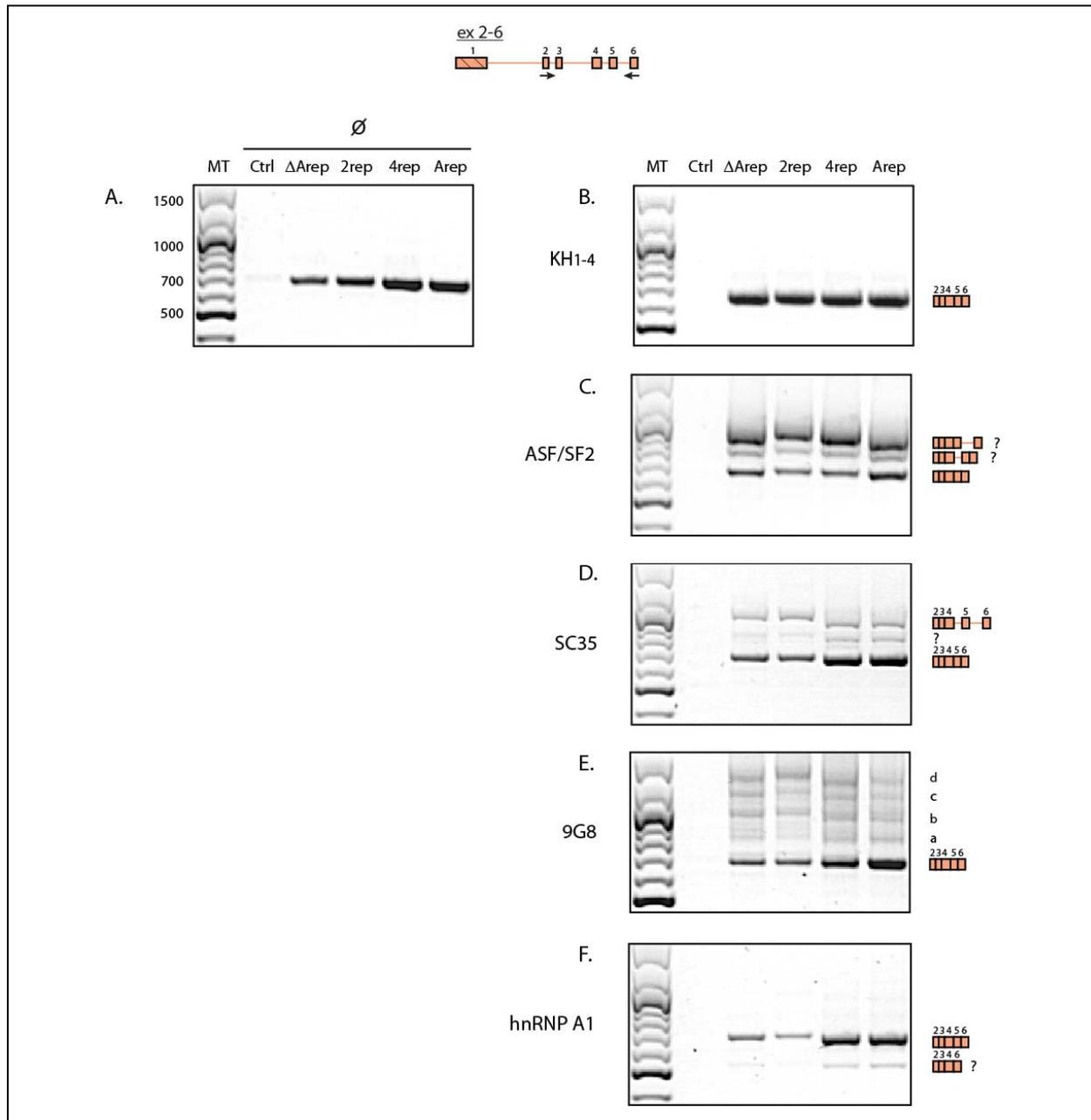
Nous avons donc réalisé les mêmes expériences que précédemment, en co-transfectant les cellules avec les différentes constructions et un plasmide codant la protéine régulatrice. Les différents plasmides d'expression des protéines sont spécifiés dans la partie Matériel et Méthodes (paragraphe I.1.3.c).

Nous avons tout d'abord réalisé une PCR pour amplifier la région incluant les exons 1 et 2 puisque les résultats obtenus sans surexpression de protéines présentaient déjà un profil particulier (**Figure 98.A**). Les résultats obtenus suite à la surproduction des différentes protéines ne diffèrent que très peu du profil initial. Notamment avec les protéines KH1-4, ASF/SF2 et 9G8, nous retrouvons les deux formes (celle correctement épissée et celle incluant l'intron 1), avec un ratio [épissé] / [non épissé] évoluant d'une faible proportion de la forme épissée vers une forte proportion en fonction du nombre de répétitions (**Figure 98.B, C et D** respectivement). Lorsque j'ai surproduit les protéines SC35 et hnRNP A1, le profil observé est légèrement différent dans le sens où, bien que le ratio [épissé] / [non épissé] évolue, il reste en faveur de la forme comprenant l'intron 1 (**Figure 98.E et F** respectivement). Nous pouvons ainsi supposer que l'effet de ces deux protéines est différent de celui des 3 précédentes. Puisque la forme correctement épissée reste en proportion plus

faible que la forme avec l'intron, nous proposons que les protéines SC35 et hnRNP A1 ont un effet négatif sur l'épissage de l'intron 1 de l'ARN Xist.



Dans un deuxième temps, nous avons poursuivi l'étude par l'analyse de l'épissage de la région entre les exons 2 et 6, qui, rappelons-le, n'était pas affecté par les différents mutants de la région des A-repeats (**Figure 99.A**).



**Figure 99 : Effet de la surproduction de différentes protéines sur l'épissage entre les exons 2 et 6 de l'ARN Xist**

RT-PCR sur les ARN issus de cellules HEK transfectées par les différents plasmides : Ctrl : pGL3c ; ΔArep : pGL3-ΔArep-ex2/6 ; 2rep : pGL3-2rep-ex2/6 ; 4rep : pGL3-4rep-ex2/6 ; Arep : pGL3-Arep-ex2/6.

A. Résultats obtenus en absence de surproduction de protéine (Ø)

B.C.D.E.F. Résultats obtenus après surproduction des protéines KH1-4, ASF/SF2, SC35, 9G8 et hnRNP A1 respectivement.

En haut est présentée la localisation des oligonucléotides utilisés pour la PCR et à droite des gels sont schématisés les produits d'épissage correspondants aux bandes observées. Les points d'interrogation indiquent que le produit d'épissage est présumé, par rapport à la taille de la bande.

Nous pouvons observer 3 types de comportement parmi les différentes protéines testées.

- Tout d'abord, la protéine KH1-4 ne semble pas avoir d'effet sur l'épissage entre les exons 2 et 6 puisqu'aucune autre bande que celle correspondant au produit correctement épissé n'est observée (**Figure 99.B**). Nous pouvons donc en conclure que dans notre modèle d'étude *in cellulo*, la protéine KH1-4 n'intervient pas dans le processus d'épissage de l'ARN Xist. Cependant, il faudrait poursuivre l'étude par des expériences utilisant des siARN, afin de vérifier si la diminution de cette protéine, quant à elle, entraîne un effet sur l'épissage.

- Ensuite, les protéines ASF/SF2, SC35 et 9G8 semblent favoriser l'inclusion de certains introns. En effet, nous pouvons observer l'apparition de bandes de taille plus élevée que celle du produit normalement attendu (**Figure 99.C, D et E**). Toutefois, nous remarquons plusieurs différences entre l'effet de ces protéines. Nous pouvons regrouper les protéines ASF/SF2 et SC35 d'une part et la protéine 9G8 d'autre part.

(i) La surproduction des protéines ASF/SF2 et SC35 (**Figure 99.B et C**), nous observons l'apparition de 2 bandes supplémentaires. Toutefois, les bandes obtenues par surproduction de la protéine ASF/SF2 migrent à des tailles différentes de celles obtenues par surproduction de SC35, donc les introns inclus ne sont pas les mêmes. Par rapport à la taille estimée des produits, il semblerait qu'ASF/SF2 favorise l'inclusion de l'intron 4 ou de l'intron 5, de manière mutuellement exclusive. En revanche, la bande supérieure observée avec SC35 pourrait correspondre à l'inclusion des 2 introns simultanément. De plus, l'effet de SC35 paraît moins prononcé que celui d'ASF/SF2 si nous comparons l'intensité des bandes. Les bandes correspondant aux produits alternatifs sont largement minoritaires dans le cas de SC35, tandis que qu'elles sont fortement représentées dans le cas d'ASF/SF2. Ainsi, en excès, ces deux protéines seraient inhibitrices de l'épissage, en favorisant l'inclusion d'introns. Cependant, même si elles agissent au niveau des mêmes introns, elles n'agissent pas exactement avec le même mécanisme.

Nous pouvons mettre ce résultat en relation avec le fait que, même si les protéines SR sont le plus souvent stimulatrices de l'épissage, elles peuvent également avoir un rôle répresseur. Rappelons ici que la protéine ASF/SF2 est aussi connue sous le nom de SRSF1 (Serine/arginine-Rich Splicing Factor 1) et la protéine SC35 est également appelée SRSF2. Il a été montré que la protéine ASF/SF2 inhibe l'épissage du pré-ARNm IIIa de l'adénovirus L1 en se fixant à l'élément répresseur appelé IIIa (3RE), localisé immédiatement en amont du point de branchement (Kanopka *et al.*, 1996). Son action s'effectue en empêchant le recrutement de la snRNP U2 sur le spliceosome. Cet effet est possible grâce à la haute affinité

d'ASF/SF2 pour l'élément 3RE. Dans notre cas, la protéine ASF/SF2 pourrait s'associer à des éléments similaires présents dans les introns de l'ARN Xist et inhiber l'épissage de ces derniers, conduisant aux profils observés.

(ii) Concernant la protéine 9G8, qui fait également partie des protéines SR (SRSF7), nous pouvons observer au moins 4 bandes supplémentaires (**Figure 99.E**). Selon leur taille de migration, il semblerait que la première bande supérieure (notée a) corresponde à l'inclusion de l'intron 2 ou de l'intron 4, les tailles de ces deux introns étant très proches. La bande encore au-dessus (notée b) correspondrait à l'inclusion de l'intron 5 et de l'un des deux introns 2 ou 4. La bande c correspond en taille à l'inclusion des trois introns 2, 4 et 5. Et enfin, la bande la plus haute (d) pourrait représenter l'inclusion de l'intron 3. Tous ces produits alternatifs d'épissage sont hypothétiques, et une analyse par séquençage ou par Southern blot serait nécessaire afin de confirmer ces suppositions. Néanmoins, le profil d'épissage est révélateur d'un effet plus général de la protéine 9G8 par rapport aux protéines ASF/SF2 et SC35. La protéine 9G8 agirait de manière plus globale sur le mécanisme d'épissage en tant que tel.

- Le troisième type d'effet est observé avec la protéine hnRNP A1. En effet, contrairement aux autres protéines, dans le cas de la surproduction de cette protéine, nous pouvons remarquer l'apparition d'une bande correspondant à un produit de plus petite taille que le produit constitutivement épissé (**Figure 99.F**). En estimant la taille de cette bande, nous avons émis l'hypothèse qu'il s'agissait de la séquence des exons 2 à 6, déléetée de l'exon 5. L'effet de cette protéine est relativement faible. En effet, nous ne l'observons qu'avec la région des A-repeats comportant 4 et 8 répétitions, et même dans ces deux cas, la bande correspondant au produit alternatif est largement minoritaire. En absence de la région ou avec 2 répétitions, le taux de transcrits est trop faible pour pouvoir visualiser le produit alternatif. La protéine hnRNP A1 semble donc avoir un effet négatif sur l'épissage, favorisant un épissage alternatif aboutissant à un transcrit final ne contenant pas l'exon 5.

L'épissage de l'ARN Xist pourrait donc être régulé par un ensemble de facteurs agissant de manière coopérative ou compétitive. La concentration de ces différentes protéines ainsi que leur localisation pourraient également influencer la régulation.

Ces expériences nous apportent également des indications quant au rôle de la région des A-repeats dans le mécanisme d'action des protéines. En effet, quelle que soit la protéine surproduite, son effet s'opère même en absence de cette région. La seule variabilité observée correspond à la quantité de produits, ce dont nous avons déjà discuté dans le paragraphe IV.3.2.a. Les protéines étudiées ont donc une influence sur l'épissage des exons 2 à 6 de l'ARN Xist, de manière non-dépendante de la région des A-repeats.

L'inactivation du chromosome X est un processus très complexe, faisant intervenir de nombreux mécanismes interdépendants. Ici nous avons pu montrer que même au niveau de la régulation de l'épissage de l'ARN Xist, qui est au cœur du processus, les régulations sont variées, même si le mécanisme global n'est pas encore totalement déchiffré.

---

# DISCUSSION & PERSPECTIVES

---



## Discussion et Perspectives

---

Ce travail a donc abouti, dans un premier temps, à la validation de la structure secondaire de la région des A-repeats de l'ARN Xist grâce à la technique de FRET appliqué à l'ARN. Peu de modèles de structure ont été proposés pour cette région. A. Wutz avait proposé une structure indépendante pour chaque répétition (Wutz *et al.*, 2002) puis des études par RMN ont montré que les répétitions peuvent se dimériser, pouvant ainsi créer une longue structure formée de l'hybridation des répétitions 2 à 2 (Duszczuk *et al.*, 2011). La méthode novatrice de FRET appliquée à l'ARN a fourni des informations de distances relatives qui ont permis de discriminer parmi plusieurs modèles de structure possibles, obtenus par S. Maenner au cours de ses expériences. Bien que ces études aient été réalisées *in vitro*, nous pouvons conclure que la région des A-repeats est formée de deux structures tige-boucle, constituée chacune de 4 éléments répétés s'hybridant 2 à 2. Il est d'ailleurs intéressant de noter que la structure proposée en 2011 à partir des données RMN ressemble à celle que nous avons mise en évidence. Cette structure particulière pourrait être à la base des différentes interactions de cette région avec ses partenaires protéiques. Une étude structurale *in cellulo* ou *in vivo* serait très difficile en raison du faible taux d'ARN Xist dans la cellule mais également du nombre important de protéines qui y sont certainement associées. La création, par l'équipe de C. Rougeulle, de lignées de cellules ES femelles possédant des mutants de la région des A-repeats est en cours. Ces lignées permettraient d'observer les effets des mutations sur la fonctionnalité de l'ARN Xist.

Dans un deuxième temps, ce travail a permis, à partir des partenaires protéiques potentiels de la région des A-repeats identifiés par chromatographie d'affinité, de mettre en évidence un certain nombre d'interactions. L'étude de ces protéines permettrait d'avoir une vue d'ensemble de tous les participants recrutés par cette région. De nombreuses protéines sont certainement impliquées dans des fonctions communes, et une analyse plus approfondie nous permettrait de parfaire la compréhension des mécanismes mis en jeu au cours de l'inactivation du chromosome X.

- C'est le cas par exemple de la protéine **Rif1** qui pourrait prendre part à l'événement de réplication tardive observée dès deux jours après le début de la différenciation. Cette réplication du Xi, qui a lieu plus tard que celle du Xa, correspond à la période d'induction de

l'expression de *Xist*, ainsi qu'au moment où les gènes liés au chromosome X deviennent transcriptionnellement silencieux. Bien que Rif 1 se fixe également sur les régions amont de la région des A-repeats, nous pourrions émettre l'hypothèse que cette protéine, en se fixant sur l'ARN *Xist*, dont la transcription est augmentée au début de l'initiation de l'inactivation et qui recouvre le futur Xi, permette de mettre en place le retard de réplication observé sur ce chromosome.

- Le facteur **TIF1 $\beta$**  a également été identifié parmi les partenaires potentiels. Il s'agit un co-répresseur pour les protéines à domaines KRAB (Krüppel-associated box) et contenant des motifs à doigts de zinc (KRAB-ZF proteins). Ces protéines se lient à des sites cibles de la chromatine de manière séquence-dépendante et interagissent ensuite avec le motif RING de TIF1 $\beta$  (Peng *et al.*, 2000). TIF1 $\beta$  induit alors la répression transcriptionnelle et la condensation de la chromatine par recrutement de HP1 (heterochromatin protein 1) (Lechner *et al.*, 2000), l'histone méthyltransférase SETDB1 qui méthyle les H3K9 (Schultz *et al.*, 2002), et le complexe de déacétylation des histones NuRD/Mi2 (Schultz *et al.*, 2001). HP1 est importante dans la formation de l'hétérochromatine. Elle peut se fixer aux H3K9me3, aux ADN méthyltransférases DNMT1 et DNMT3 impliquées dans la méthylation des îlots CpG, et à l'histone méthyltransférase Suv39H1 (Maison et Almouzni, 2004 ; Campos et Reinberg, 2009 ; Dinant et Luijsterburg, 2009). Etant donné que Suv39H1 peut triméthyle H3K9 et que HP1 peut interagir avec à la fois H3K9me3 et Suv39H1, il a été proposé que HP1 participe à l'expansion de l'hétérochromatine (Maison et Almouzni, 2004).

Bien qu'aucune donnée n'ait fait mention de Suv39H1 ou SETDB1 pour la méthylation des H3K9 des gènes liés au chromosome X, le fait de retrouver TIF1 $\beta$  parmi les protéines en complexe avec la région des A-repeats n'est sûrement pas anodin. Au laboratoire, S. Maenner a montré que cette région recrutait le complexe PRC2 responsable de la triméthylation des H3K27. Peut-être existe-t-il un mécanisme similaire par lequel la région des A-repeats recrute le(s) complexe(s) responsable(s) de la méthylation des H3K9.

- La spectrométrie de masse a également permis de mettre en évidence la présence de composants du complexe cohésine (**Smc1 et 3**) et de la protéine **Pds5a** parmi les partenaires potentiels de la région des A-repeats. Le complexe cohésine est impliqué dans le rapprochement des chromatides sœurs, comme nous l'avons déjà mentionné dans la partie Résultats (paragraphe III.3.4). De plus, une grande partie des sites de reconnaissance de la cohésine chevauche les sites de CTCF (Parelho *et al.*, 2008). Il a été montré que le complexe

cohésine agit en collaboration avec le facteur CTCF pour former des boucles intrachromosomiques entre des régions très éloignées (Hadjur *et al.*, 2009). Dans cette étude, il a été mis en évidence la formation d'une boucle entre des sites de fixation de CTCF et de la cohésine séparés par une dizaine de kb. De plus, cette interaction longue-distance est nécessaire pour la régulation de l'expression du gène cible, en l'occurrence le gène *IFNG*. Le taux de fixation de ces protéines est directement corrélé à l'induction de l'expression du gène et il peut être modulable selon l'environnement ou le stade de développement.

Le facteur CTCF est déjà impliqué dans la mise en place de l'inactivation, comme cela a été décrit dans l'Introduction (paragraphe I.2.1 et IV.3.2). Il serait très intéressant de chercher à savoir s'il est également engagé dans la formation de longues boucles intrachromosomiques, mettant notamment en jeu la région des A-repeats et le complexe cohésine. Une hypothèse quand à la fonction de ces boucles serait de relocaliser les gènes à inactiver dans le domaine Xist transcriptionnellement silencieux – rappelons que cette étape de relocalisation est dépendante de la région des A-repeats (Chaumeil *et al.*, 2006). Cette hypothèse est supportée par le mécanisme similaire qui aurait lieu chez *C. elegans*. En effet, comme nous l'avons vu dans le paragraphe II.2.2 de l'Introduction, le complexe condensine interagirait avec le complexe DCC pour la formation de boucles de fibres chromatinienne permettant le rapprochement des gènes régulés (Jans *et al.*, 2009)

Afin de vérifier la présence de telles boucles, la technique de capture de la conformation des chromosomes (3C) pourrait être utilisée. Elle permettrait non seulement de vérifier la présence d'une boucle, mais également d'identifier les sites impliqués dans sa formation. Nous pourrions alors déterminer quelles protéines y participent précisément.

- Parmi les protéines identifiées, nous nous sommes particulièrement intéressés aux protéines **PTB** et **KSRP**. L'étude de ces protéines a révélé une association directe avec cette région au niveau de sites de fixation communs. La différence d'affinité pour des sites semblables observée peut participer à la fonction de ces protéines. En effet, nous savons que la protéine KSRP est impliquée dans la dégradation des ARN (Gherzi *et al.*, 2004), et dorénavant qu'elle se fixe sur la région des A-repeats. Or il a été montré que la machinerie de dégradation des ARN joue un rôle dans la régulation du taux d'ARN Xist épissés et dans l'initiation de l'inactivation (Ciaudo *et al.*, 2006). La protéine KSRP serait-elle impliquée dans la régulation de la stabilité/dégradation de Xist ? Nous pourrions supposer que les protéines PTB et KSRP agissent de manière séquentielle. Avant l'initiation de l'inactivation, l'ARN Xist est produit en faible quantité à partir des 2 chromosomes X. La protéine KSRP

pourrait interagir durant cette période avec la région des A-repeats afin de contrôler le taux de transcrits. En recrutant l'exosome nucléaire, elle dirigerait l'ARN vers la dégradation et empêcherait ainsi son accumulation. Très rapidement après le début de la différenciation, l'ARN *Xist* s'accumule sur le futur Xi. Cette transition, comme nous l'avons vu dans l'Introduction, est liée à de nombreux facteurs de régulation qui agissent sur l'expression de *Xist* et de *Tsix* notamment. Dans le but de stabiliser l'ARN ou inhiber sa dégradation, la protéine PTB pourrait intervenir en interagissant avec la région des A-repeats en remplaçant KSRP au niveau des mêmes sites de fixation. D'ailleurs, La protéine PTB, en se fixant au niveau des boucles internes et terminales des structures tige-boucle de la région des A-repeats, pourrait stabiliser la région dans cette conformation. S. Maenner avait réalisé des tests de stabilité par dichroïsme circulaire, comparant les spectres de complexes ARN/protéine enregistrés à des températures croissantes. Il avait observé que la présence de PTB induisait un changement de conformation différent de celui obtenu avec l'ARN seul. Cette protéine pourrait assurer la bonne structuration de la région des A-repeats ainsi que sa stabilité, permettant à la région d'être fonctionnelle. Son affinité étant meilleure, la protéine KSRP serait écartée et l'ARN *Xist* pourrait s'accumuler sur le chromosome. Cette hypothèse pourrait être vérifiée par comparaison de la fixation de ces 2 protéines sur la région des A-repeats à différents temps de la différenciation, J-1, J0, J1 et J2 par exemple, l'accumulation de l'ARN *Xist* commençant à J1. Par ailleurs, PTB pourrait avoir un rôle de protéine plateforme pour le recrutement d'autres protéines.

Par ailleurs, d'autres protéines ont retenu notre attention bien qu'elles n'aient pas été identifiées par spectrométrie de masse. C'est le cas de la protéine **ASF/SF2**, qui a été identifiée par l'équipe de B. Panning comme un partenaire potentiel de la région des A-repeats (Royce-Tolland *et al.*, 2010). Nous avons confirmé que cette protéine est en interaction directe avec la région des A-repeats *in vitro*, même si une étude plus approfondie serait nécessaire pour déterminer avec précision ses sites de fixation. Cette protéine faisant partie de la famille des protéines SR impliquées dans l'épissage, nous nous sommes demandé quel était le rôle d'ASF/SF2 mais également de la région des A-repeats dans la régulation de l'épissage de *Xist*.

- Tout d'abord, nous avons mis en évidence un rôle de la région des A-repeats elle-même dans la régulation de l'épissage de *Xist*, et plus particulièrement sur l'exclusion de l'intron 1. En effet, nous avons observé que l'absence répétitions ou la présence de seulement 2 répétitions induisait une forte proportion d'ARN *Xist* contenant l'intron 1, contrairement à la présence de 4 répétitions ou la région entière pour lesquelles le rapport est inversé. De plus,

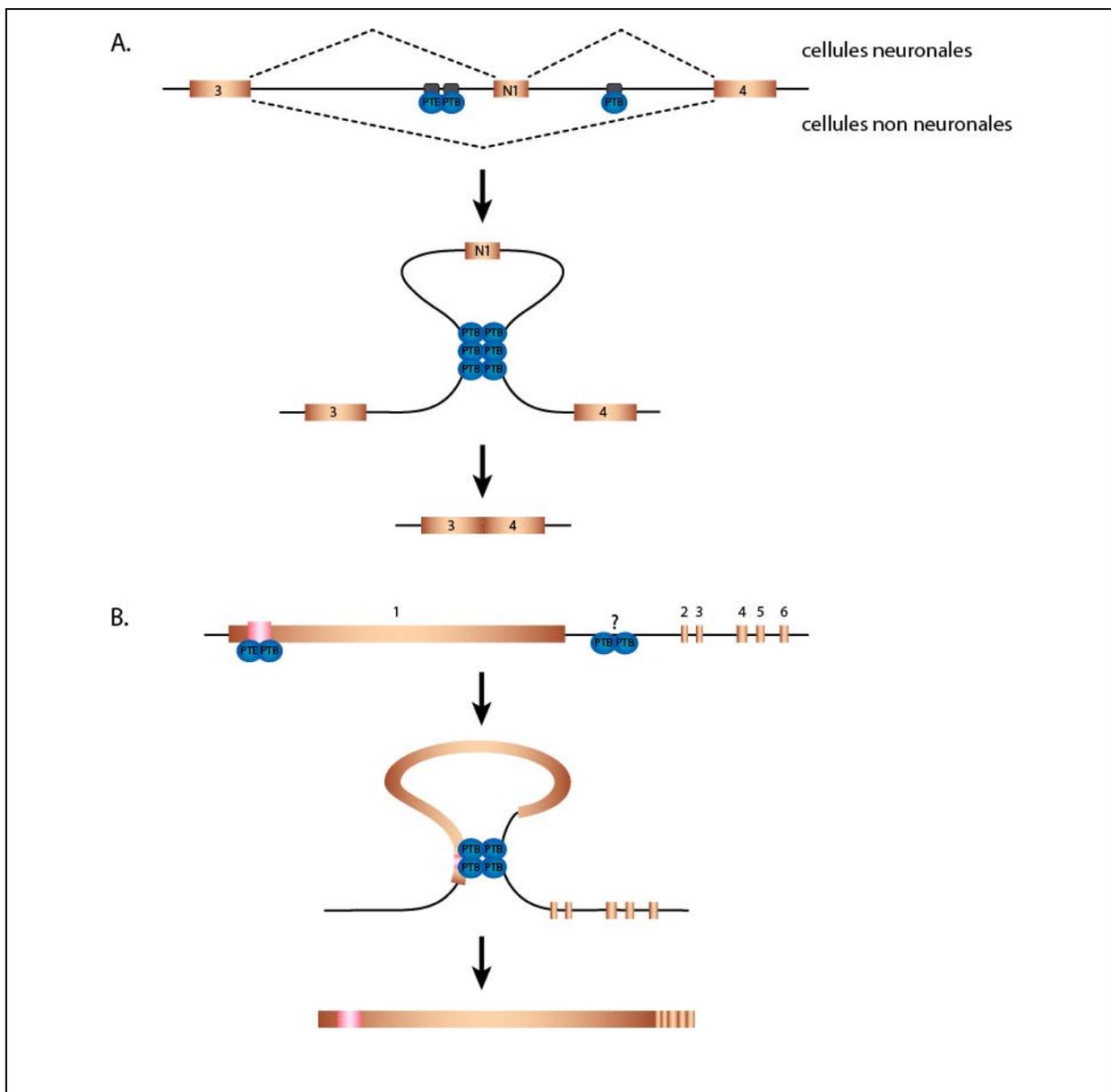
nous avons montré que dans un système hétérologue comportant la région des A-repeats largement en amont du site 5' donneur d'épissage de la région C2 de HIV-1, le nombre de répétitions incluses dans la construction influence également le ratio d'ARN [épissés] / ARN [non épissés]. Nous pouvons alors nous demander comment une région aussi éloignée du premier site 5' d'épissage (rappelons qu'elle est située à l'extrémité 5' de l'intron 1 long d'environ 10 kb) peut être impliquée dans ce processus. Il est maintenant admis que la régulation de l'expression des gènes peut s'effectuer par l'intermédiaire de régions situées à longue distance des gènes cibles. Cependant cette observation n'est pas courante dans les cas de régulation de l'épissage. Pour autant, il a été montré que les pré-ARNm forment des structures bi et tridimensionnelles qui constituent des mécanismes de régulation de nombreux processus des ARN (pour revue, (Buratti et Baralle, 2004).

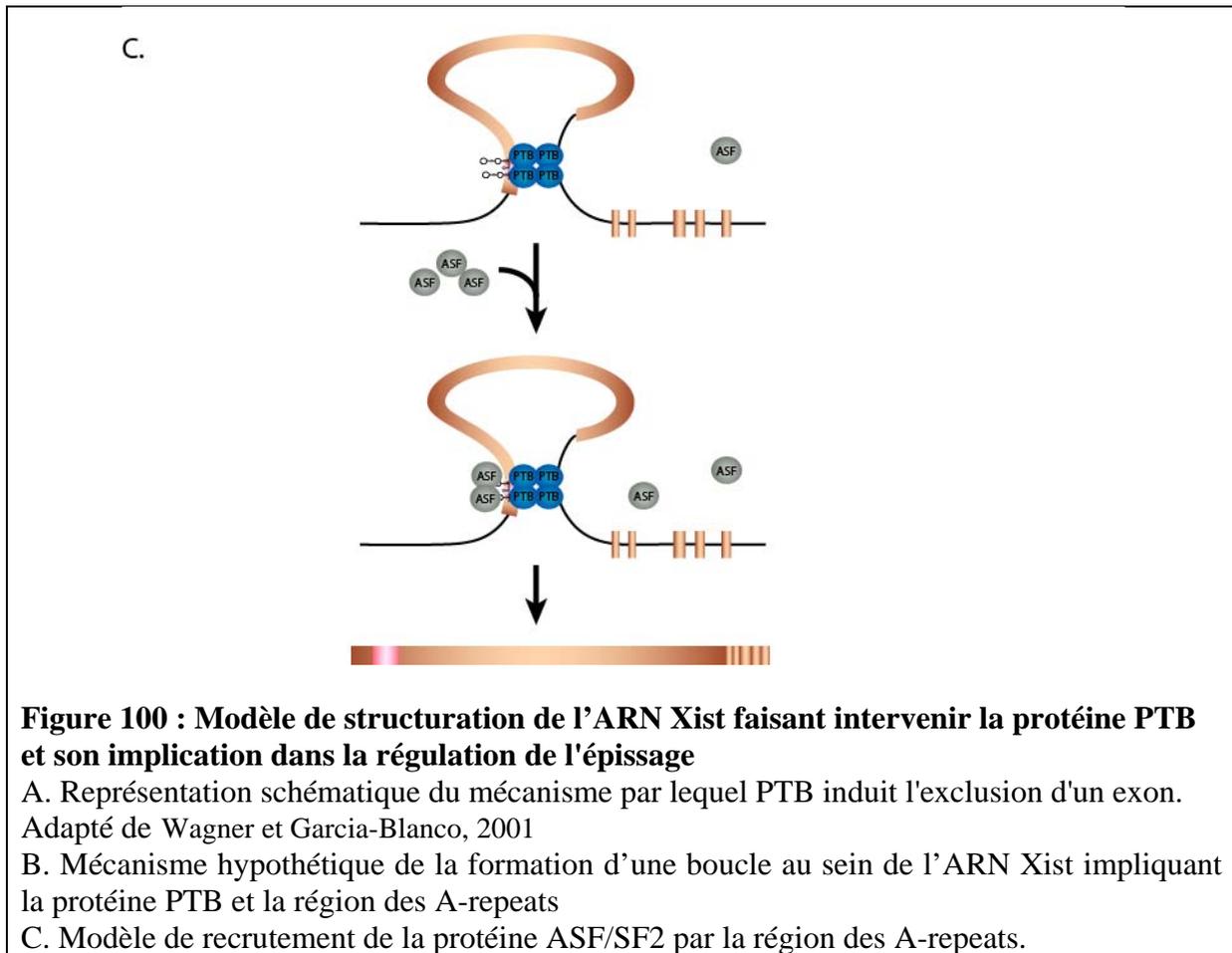
En général, ces structurations affectent l'épissage dans un environnement proche. Nous pouvons illustrer cet aspect par le cas du pré-ARNm rp51 de *S. cerevisiae*. La formation d'une hélice entre le site 5' donneur d'épissage et la boîte de branchement y facilite l'assemblage de la snRNP U1 (Charpentier et Rosbash, 1996). De même, la structure secondaire du pré-ARNm peut influencer sur la fixation de facteurs régulateurs, qu'ils soient activateurs ou répresseurs, la fixation dépendant ainsi de la séquence mais aussi de la structure (Shi *et al.*, 1997 ; Nagel *et al.*, 1998). La présence d'une structure 2D peut masquer certains sites de fixation, comme dans le cas où un site se retrouve en double brin par la formation d'une hélice. Au contraire, elle peut exposer ou rapprocher des sites, facilitant ainsi le recrutement de protéines régulatrices (Libri *et al.*, 1995). La plupart des membres majeurs de la famille des protéines SR est ainsi affectée par la conformation de l'ARN cible (Buratti *et al.*, 2004).

A une autre échelle, la structure secondaire d'une région d'un ARN peut affecter la structure d'ordre plus élevé de la molécule entière. La formation de grandes structures impliquant des introns et des exons conduit à une régulation de l'épissage par inclusion/suppression d'exons. Nous pouvons citer l'exemple du pré-ARNm de la  $\beta$ -tropomyosine dans lequel les exons 6A et 6B sont mutuellement exclusifs. Une structure engageant une partie des introns en amont et en aval de l'exon 6 ainsi que l'exon 6 lui-même provoque l'exclusion de cet exon de l'ARN mature dans les cellules non musculaires (Clouet d'Orval *et al.*, 1991 ; Libri *et al.*, 1991).

L'exclusion d'exon peut également être induite par des protéines régulatrices. C'est le cas de l'exon N1 de l'ARNm de c-src spécifiquement intégré dans les cellules neuronales. Dans les autres cellules, il est exclu par formation d'une boucle, formée par la fixation de la protéine PTB sur des sites situés de part et d'autre de l'exon (Wagner et Garcia-Blanco, 2001) (**Figure 100.A**).

Puisque la protéine PTB interagirait avec la région des A-repeats, et au vu de l'implication de cette région dans l'épissage de l'ARN *Xist*, nous pouvons supposer un mécanisme similaire. La boucle pourrait impliquer la région des A-repeats d'une part, et une région éloignée en aval d'autre part (**Figure 100.B**). Cette structure pourrait rapprocher le promoteur du gène *Xist* du premier site 5' donneur se trouvant 10 kb en aval. L'épissage étant un processus co-transcriptionnel, ce rapprochement spatial permettrait de faciliter la détermination du premier intron et donc de faciliter l'épissage de ce long ARN. Elle pourrait également permettre ou tout du moins faciliter le recrutement de protéines essentielles, comme par exemple la protéine ASF/SF2 (**Figure 100.C**).





- Ensuite, nous avons montré que la protéine ASF/SF2 est impliquée dans la régulation de l'épissage de l'ARN Xist. Une surproduction de cette protéine induit un défaut d'épissage de l'ARN Xist *in cellulo*, dans le modèle que nous avons élaboré et utilisé. En présence d'un excès de cette protéine, nous avons observé l'apparition de molécules d'ARN dont les introns n'ont pas été éliminés. ASF/SF2 est connue pour avoir un rôle activateur ou inhibiteur selon les ARN cibles. Dans notre cas, en forte concentration tout du moins, elle aurait donc un rôle négatif sur l'épissage en inhibant l'exclusion de certains introns. Plusieurs hypothèses peuvent être envisagées. Tout d'abord, elle pourrait masquer des sites essentiels pour la reconnaissance des introns et donc leur élimination. Dans une autre hypothèse, ASF/SF2 pourrait recruter d'autres protéines régulatrices, qui elles auraient une action directe sur l'épissage. Dans les 2 cas, l'effet serait lié à la quantité de protéine. La quantité présente dans les cellules de notre modèle ne serait pas suffisante pour recouvrir les sites ou recruter assez de protéines, l'effet serait ainsi visible à partir d'un certain seuil de concentration en ASF/SF2.

Afin d'approfondir le rôle d'ASF/SF2 dans la régulation de l'épissage de Xist, il serait intéressant de réaliser des expériences utilisant des siARN. Au lieu de surproduire cette protéine, nous diminuerions fortement sa concentration cellulaire et nous pourrions observer les effets sur l'épissage des exons de l'ARN Xist.

La régulation de l'épissage est en général assurée par un ensemble de protéines agissant en collaboration et/ou en compétition. C'est le cas par exemple dans la régulation du site A7 d'épissage de HIV-1 requis pour la production des ARNm *tat* et *rev* (Marchand *et al.*, 2002). La protéine hnRNP A1 s'oligomérisse au niveau de plusieurs sites dans la région du site A7 et inhibe l'utilisation de ce site. Cependant, les protéines antagonistes hnRNP A1 et ASF/SF2 partagent au moins un site de fixation. Lorsqu'ASF/SF2 entre en compétition avec hnRNP A1, cette dernière se dissocie du site de fixation pour laisser la place à ASF/SF2 qui peut alors recruter les composants du spliceosome. Ainsi, selon le ratio de concentration des 2 protéines dans la cellule, le site A7 est activé ou réprimé.

Il serait intéressant de chercher si ASF/SF2 agit de la même manière dans le cas de l'épissage de l'ARN Xist. L'analyse de protéines connues pour avoir un effet opposé à cette protéine, telles que les protéines hnRNP, devrait nous donner des informations à ce sujet. De même, la protéine Sam68 est connue pour son rôle dans la régulation du cycle cellulaire et de l'apoptose (Taylor *et al.*, 2004). Il a été montré que cette protéine est impliquée dans la régulation de l'épissage alternatif de l'ARN Bcl-x (Paronetto *et al.*, 2007). De plus, que la protéine ASF/SF2 contrecarre ses effets, tandis qu'elle interagit avec la protéine hnRNP A1 de manière coopérative. Or Sam68 fait partie des protéines identifiées par spectrométrie de masse comme partenaires potentiels de la région des A-repeats. Cette protéine serait donc un bon candidat pour une étude plus poussée de la régulation de l'épissage de l'ARN Xist.

Outre la protéine ASF/SF2, nous avons également montré l'influence des protéines **SC35**, **9G8** et **hnRNP A1** dans l'épissage de l'ARN Xist. Bien que leur influence semble indépendante de la région des A-repeats, nous avons observé que la protéine 9G8 pourrait avoir une action ciblée sur les exons 2 à 6, tandis que les protéines SC35 et hnRNP A1 agiraient aussi sur l'épissage de l'intron 1. Globalement, ces 3 protéines inhibent l'épissage correct de l'ARN Xist, mais indépendamment de la région des A-repeats. Il serait intéressant de réaliser aussi des expériences de knock-down afin de confirmer ces rôles et d'essayer d'en comprendre les mécanismes. Nous pouvons également nous demander si ces protéines agissent en synergie ou en compétition. Pour répondre à cette question, il faudrait tout d'abord

déterminer les sites de fixation de ces protéines au niveau des différents exons et introns, puis étudier l'effet de la variation de concentration de chaque protéine sur le profil d'épissage de Xist.

La région des A-repeats, qui avait été identifiée par A. Wutz comme essentielle à la mise en place de l'inactivation du chromosome X, paraît donc être impliquée dans ce processus complexe à plusieurs niveaux grâce au recrutement d'une large variété de protéines. Que ce soit en agissant sur la stabilité de l'ARN Xist, sur la régulation de son épissage, ou encore sur la relocalisation des gènes à inactiver dans le domaine Xist, cette région semble être au cœur de nombreux mécanismes cruciaux pour l'inactivation. Cependant, certains mécanismes ne sont qu'à l'état d'hypothèses et la poursuite des travaux sur cette région, et sur l'ARN Xist plus globalement, permettront d'élucider un peu plus le processus encore mystérieux de l'inactivation du chromosome X chez les mammifères femelles, apportant par là même de précieuses informations sur la régulation épigénétique impliquant de longs ARN non-codants.



---

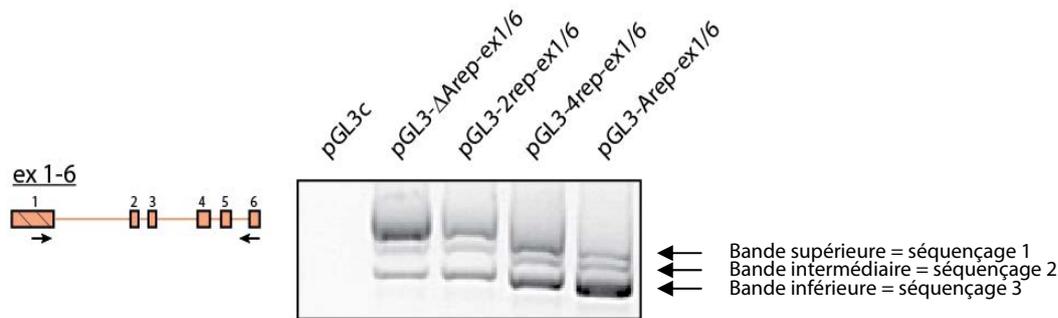
# ANNEXES

---



## Annexes

Séquençage entre les exons 1 et 3 des bandes issues de la PCR amplifiant la séquence entre les exons 1 et 6 de l'échantillon transfecté par le plasmide pGL3-Arep-ex1/6



### - Séquençage de la bande supérieure :

pGL3-Arep-ex1/6	GATGTTCAAGTAACTCATCCAGTGCAGGATCCTGCTTGAACTACTGCTCCTCCGTTACAT
Séquençage 1	
	GATGTTCAAGTAACTCATCCAGTGCAGGATCCTGCTTGAACTACTGCTCCTCCGTTACAT
pGL3-Arep-ex1/6	CAGACTCTGGCTGTTTGTAGACTACAGtaagtaactaccttgtagtactaagctatctttt
Séquençage 1	
	CAGACTCTGGCTGTTTGTAGACTACAGTAAGTACTACCTTGTTAGCTACTAAGCTATCTTTT
pGL3-Arep-ex1/6	tctcattcgagtttctttctgtgccaccactgaaaagtcaaagtgactaataagtc...
Séquençage 1	
	TCTCATTTCGAGTTTCTTTCTGTGCCACCCTGAAAAGTCAAAGTGACTAATAAGTC...
pGL3-Arep-ex1/6	...tgagaacatatcctcaccctgagttcttaccaccaattgaaaacgctgacatgtgca
Séquençage 1	
	...TGAGAACATATCCTCACCTGAGTTCTTACCACCAATTGAAAACGCTGACATGTGCA
pGL3-Arep-ex1/6	tatgaagattgtattgctagtaatacaattatcattttagagtttagtataattatcat
Séquençage 1	
	TATGAAGATTGTATTGCTAGTAATCAAATTATCATTGTAGAGTTAGTATAATTATCAT
pGL3-Arep-ex1/6	ttgtttccaggGATGAATTTGGAGTCTGTTTGTGCTCCTGCCTCAAGAAGAAGGA...
Séquençage 1	
	TTGTTTTCCAGGGATGAATTTGGAGTCTGTTTGTGCTCCTGCCTCAAGAAGAAGGA...
pGL3-Arep-ex1/6	...GAGTCTGTTTTGTGCTCCTGCCTCAAGAAGAAGGATTGCCTGGATTTAGAGGAGTGA
Séquençage 1	
	...GAGTCTGTTTTGTGCTCCTGCCTCAAGAAGAAGGATTGCCTGGATTTAGAGGAGTGA
pGL3-Arep-ex1/6	AGAGTGCTGGAGAGAGCCCAAAGGtatttccggttacttggttgactgagatagtatcttc
Séquençage 1	
	AGAGTGCTGGAGAGAGCCCAAAGG-----
pGL3-Arep-ex1/6	ccttaaggtctaattgtgtcaagttatacctcctaagtaaacctcattatcttcatgt
Séquençage 1	-----
pGL3-Arep-ex1/6	ggaatcttcaacaattttttagtagtctaattgatagccttgattttcttcacaggGACAAAC
Séquençage 1	-----
	GACAAAC

pGL3-Arep-ex1/6 AATCCCTATGTGAGACTCAAGGACTGCCAGCAGCCTATACAGCTACATTACATCTCAGCA  
 Séquençage 1 AATCCCTATGTGAGACTCAAGGACTGCCAGCAGCCTATACAGCTACATTACATCTCAGCA

→ L'intron 1 n'a pas été épissé.

- Séquençage de la bande intermédiaire :

pGL3-Arep-ex1/6 GATGTTCAAGTAACTCATCCAGTGCAGGATCCTGCTTGAAGTACTGCTCCTCCGTTACAT  
 Séquençage 2 GATGTTCAAGTAACTCATCCAGTGCAGGATCCTGCTTGAAGTACTGCTCCTCCGTTACAT

pGL3-Arep-ex1/6 CAGACTCTGGCTGTTTAGACTACAGtaagtactaccttgtagctactaagctatctttt  
 Séquençage 2 CAGACTCTGGCTGTTTAGACTACAG-----

pGL3-Arep-ex1/6 tctcattcgagtttcttcttctgtgccaccactgaaaagtcaaagtgactaataagtc...  
 Séquençage 2 -----... .

pGL3-Arep-ex1/6 ...tgagaacatatcctcaccctgagttcttaccaccaattgaaaacgctgacatgtgca  
 Séquençage 2 ...-----

pGL3-Arep-ex1/6 tatgaagattgtattgctagtaatacaattatcattttagagtttagtataattatcat  
 Séquençage 2 -----

pGL3-Arep-ex1/6 ttgtttccaggGATGAATTTGGAGTCTGTTTTGTGCTCCTGCCTCAAGAAGAAGGA...  
 Séquençage 2 -----GATGAATTTGGAGTCTGTTTTGTGCTCCTGCCTCAAGAGGAAGGA... .

pGL3-Arep-ex1/6 ...GAGTCTGTTTTGTGCTCCTGCCTCAAGAAGAAGGATTGCCTGGATTTAGAGGAGTGA  
 Séquençage 2 ...GAGTCTGTTTTGTGCTCCTGCCTCAAGAAGAAGGATTGCCTGGATTTAGAGGAGTGA

pGL3-Arep-ex1/6 AGAGTGCCTGGAGAGAGCCCAAAGGtatttccggttacttggttgactgagatagtatcttc  
 Séquençage 2 AGAGTGCCTGGAGAGAGCCCAAAGG-----

pGL3-Arep-ex1/6 ccttaaaggctaatgtgtcaagttatacctcctaagtaaatacctcattatcttcatgt  
 Séquençage 2 -----

pGL3-Arep-ex1/6 ggaatcttcaacaattttttagtctaatgatagccttgtatttcttcacaggGACAAAC  
 Séquençage 2 -----GACAAAC

pGL3-Arep-ex1/6 AATCCCTATGTGAGACTCAAGGACTGCCAGCAGCCTATACAGCTACATTACATCTCAGCA  
 Séquençage 2 AATCCCTATGTGAGACTCAAGGACTGCCAGCAGCCTATACAGCTACATTACATCTCAGCA

→ Les introns 1 et 2 ont été épissés.





---

# BIBLIOGRAPHIE

---



## Bibliographie

---

- Akusjarvi, G. and Stevenin, J.** (2003). "Remodelling of the host cell RNA splicing machinery during an adenovirus infection." *Curr Top Microbiol Immunol* **272**: 253-286.
- Aladjem, M. I.** (2012). "Rif1 choreographs DNA replication timing." *EMBO J.*
- Augui, S., Fillion, G. J., Huart, S., Nora, E., Guggiari, M., Maresca, M., . . . Heard, E.** (2007). "Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic." *Science (New York, N.Y.)* **318**(5856): 1632-1636.
- Bacher, C. P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., . . . Heard, E.** (2006). "Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation." *Nature cell biology* **8**(3): 293-299.
- Baetz, K. K., Krogan, N. J., Emili, A., Greenblatt, J. and Hieter, P.** (2004). "The ctf13-30/CTF13 genomic haploinsufficiency modifier screen identifies the yeast chromatin remodeling complex RSC, which is required for the establishment of sister chromatid cohesion." *Molecular and cellular biology* **24**(3): 1232-1232.
- 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 research* **26**(23): 5343-5350.
- Baltus, G. A., Kowalski, M. P., Tutter, A. V. and Kadam, S.** (2009). "A positive regulatory role for the mSin3A-HDAC complex in pluripotency through Nanog and Sox2." *J Biol Chem* **284**(11): 6998-7006.
- Bar, A., Marchand, V., Khoury, G., Dreumont, N., Mouglin, A., Robas, N., . . . Branlant, C.** (2011). "Structural and functional analysis of the Rous Sarcoma virus negative regulator of splicing and demonstration of its activation by the 9G8 SR protein." *Nucleic acids research* **39**(8): 3388-3403.
- Barakat, T. S., Gunhanlar, N., Pardo, C. G., Achame, E. M., Ghazvini, M., Boers, R., . . . Gribnau, J.** (2011). "RNF12 activates Xist and is essential for X chromosome inactivation." *PLoS genetics* **7**(1): e1002001-e1002001.
- Barr, M. L. and Bertram, E. G.** (1949). "A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis." *Nature* **163**(4148): 676-676.
- Berger, S. L.** (2002). "Histone modifications in transcriptional regulation." *Current opinion in genetics & development* **12**(2): 142-148.
- Berletch, J. B., Yang, F. and Disteche, C. M.** (2010). "Escape from X inactivation in mice and humans." *Genome biology* **11**(6): 213-213.
- Bestor, T., Laudano, A., Mattaliano, R. and Ingram, V.** (1988). "Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases." *J Mol Biol* **203**(4): 971-983.

- Bestor, T. H. and Ingram, V. M.** (1983). "Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA." *Proc Natl Acad Sci U S A* **80**(18): 5559-5563.
- Bilodeau, P. S., Domsic, J. K., Mayeda, A., Krainer, A. R. and Stoltzfus, C. M.** (2001). "RNA splicing at human immunodeficiency virus type 1 3' splice site A2 is regulated by binding of hnRNP A/B proteins to an exonic splicing silencer element." *J Virol* **75**(18): 8487-8497.
- Black, D. L.** (2003). "Mechanisms of alternative pre-messenger RNA splicing." *Annu Rev Biochem* **72**: 291-336.
- Boggs, B. A., Cheung, P., Heard, E., Spector, D. L., Chinault, A. C. and Allis, C. D.** (2002). "Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes." *Nature genetics* **30**(1): 73-76.
- Borsani, G., Tonlorenzi, R., Simmler, M. C., Dandolo, L., Arnaud, D., Capra, V., . . . Ballabio, A.** (1991). "Characterization of a murine gene expressed from the inactive X chromosome." *Nature* **351**(6324): 325-329.
- Boutz, P. L., Stoilov, P., Li, Q., Lin, C. H., Chawla, G., Ostrow, K., . . . Black, D. L.** (2007). "A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons." *Genes & Development* **21**(13): 1636-1652.
- Briggs, S. D., Xiao, T., Sun, Z.-W., Caldwell, J. a., Shabanowitz, J., Hunt, D. F., . . . Strahl, B. D.** (2002). "Gene silencing: trans-histone regulatory pathway in chromatin." *Nature* **418**(6897): 498-498.
- Brockdorff, N., Ashworth, A., Kay, G. F., Cooper, P., Smith, S., McCabe, V. M., . . . Rastan, S.** (1991). "Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome." *Nature* **351**(6324): 329-331.
- Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., . . . Rastan, S.** (1992). "The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus." *Cell* **71**(3): 515-526.
- Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R. and Willard, H. F.** (1991). "A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome." *Nature* **349**(6304): 38-44.
- Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafrenière, R. G., Xing, Y., Lawrence, J. and Willard, H. F.** (1992). "The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus." *Cell* **71**(3): 527-542.
- Brown, C. J. and Willard, H. F.** (1994). "The human X-inactivation centre is not required for maintenance of X-chromosome inactivation." *Nature* **368**(6467): 154-156.
- Buratti, E. and Baralle, F. E.** (2004). "Influence of RNA secondary structure on the pre-mRNA splicing process." *Mol Cell Biol* **24**(24): 10505-10514.
- Buratti, E., Dork, T., Zuccato, E., Pagani, F., Romano, M. and Baralle, F. E.** (2001). "Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping." *EMBO Journal* **20**(7): 1774-1784.

- Buratti, E., Muro, A. F., Giombi, M., Gherbassi, D., Iaconcig, A. and Baralle, F. E.** (2004). "RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon." *Mol Cell Biol* **24**(3): 1387-1400.
- Burdon, T., Smith, A. and Savatier, P.** (2002). "Signalling, cell cycle and pluripotency in embryonic stem cells." *Trends Cell Biol* **12**(9): 432-438.
- Caceres, J. F. and Krainer, A. R.** (1993). "Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains." *EMBO Journal* **12**(12): 4715-4726.
- Caceres, J. F., Stamm, S., Helfman, D. M. and Krainer, A. R.** (1994). "Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors." *Science* **265**(5179): 1706-1709.
- Campos, E. I. and Reinberg, D.** (2009). "Histones: annotating chromatin." *Annu Rev Genet* **43**: 559-599.
- Cao, R., Tsukada, Y. and Zhang, Y.** (2005). "Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing." *Mol Cell* **20**(6): 845-854.
- Caparros, M. L., Alexiou, M., Webster, Z. and Brockdorff, N.** (2002). "Functional analysis of the highly conserved exon IV of XIST RNA." *Cytogenetic and Genome Research* **99**(1-4): 99-105.
- Carrel, L. and Willard, H. F.** (2005). "X-inactivation profile reveals extensive variability in X-linked gene expression in females." *Nature* **434**(7031): 400-404.
- Castelo-Branco, P., Furger, A., Wollerton, M., Smith, C., Moreira, A. and Proudfoot, N.** (2004). "Polypyrimidine tract binding protein modulates efficiency of polyadenylation." *Molecular and Cellular Biology* **24**(10): 4174-4183.
- Chadwick, B. P. and Willard, H. F.** (2004). "Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome." *Proceedings of the National Academy of Sciences of the United States of America* **101**(50): 17450-17455.
- Chan, R. C. and Black, D. L.** (1997). "The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream." *Mol Cell Biol* **17**(8): 4667-4676.
- Chan, R. C., Severson, A. F. and Meyer, B. J.** (2004). "Condensin restructures chromosomes in preparation for meiotic divisions." *The Journal of cell biology* **167**(4): 613-625.
- Chandler, S. P., Guschin, D., Landsberger, N. and Wolffe, A. P.** (1999). "The methyl-CpG binding transcriptional repressor MeCP2 stably associates with nucleosomal DNA." *Biochemistry* **38**(22): 7008-7018.
- Chao, W., Huynh, K. D., Spencer, R. J., Davidow, L. S. and Lee, J. T.** (2002). "CTCF, a candidate trans-acting factor for X-inactivation choice." *Science (New York, N.Y.)* **295**(5553): 345-347.
- Charlesworth, D., Charlesworth, B. and Marais, G.** (2005). "Steps in the evolution of heteromorphic sex chromosomes." *Heredity* **95**(2): 118-128.
- Charpentier, B. and Rosbash, M.** (1996). "Intramolecular structure in yeast introns aids the early steps of in vitro spliceosome assembly." *Rna* **2**(6): 509-522.

- Chaumeil, J., Le Baccon, P., Wutz, A. and Heard, E.** (2006). "A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced." *Genes & development* **20**(16): 2223-2237.
- Chaumeil, J., Okamoto, I., Guggiari, M. and Heard, E.** (2002). "Integrated kinetics of X chromosome inactivation in differentiating embryonic stem cells." *Cytogenetic and Genome Research* **99**(1-4): 75-84.
- Chen, C. Y., Gherzi, R., Ong, S. E., Chan, E. L., Raijmakers, R., Pruijn, G. J., . . . Karin, M.** (2001). "AU binding proteins recruit the exosome to degrade ARE-containing mRNAs." *Cell* **107**(4): 451-464.
- Chen, H. Y., Sun, J. M., Zhang, Y., Davie, J. R. and Meistrich, M. L.** (1998). "Ubiquitination of histone H3 in elongating spermatids of rat testes." *J Biol Chem* **273**(21): 13165-13169.
- Chou, M. Y., Underwood, J. G., Nikolic, J., Luu, M. H. and Black, D. L.** (2000). "Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing." *Mol Cell* **5**(6): 949-957.
- Chow, J. and Heard, E.** (2009). "X inactivation and the complexities of silencing a sex chromosome." *Current opinion in cell biology* **21**(3): 359-366.
- Chow, J. C., Hall, L. L., Baldry, S. E., Thorogood, N. P., Lawrence, J. B. and Brown, C. J.** (2007). "Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible." *Proceedings of the National Academy of Sciences of the United States of America* **104**(24): 10104-10109.
- Chow, J. C., Hall, L. L., Clemson, C. M., Lawrence, J. B. and Brown, C. J.** (2003). "Characterization of expression at the human XIST locus in somatic, embryonal carcinoma, and transgenic cell lines." *Genomics* **82**(3): 309-322.
- Chuang, P. T., Lieb, J. D. and Meyer, B. J.** (1996). "Sex-specific assembly of a dosage compensation complex on the nematode X chromosome." *Science* **274**(5293): 1633-1634.
- Chureau, C., Chantalat, S., Romito, A., Galvani, A., Duret, L., Avner, P. and Rougeulle, C.** (2011). "Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region." *Human molecular genetics* **20**(4): 705-718.
- Chureau, C., Prissette, M., Bourdet, A., Barbe, V., Cattolico, L., Jones, L., . . . Duret, L.** (2002). "Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine." *Genome Research* **12**(6): 894-908.
- Ciaudo, C., Bourdet, A., Cohen-Tannoudji, M., Dietz, H. C., Rougeulle, C. and Avner, P.** (2006). "Nuclear mRNA degradation pathway(s) are implicated in Xist regulation and X chromosome inactivation." *PLoS genetics* **2**(6): e94-e94.
- Clemson, C. M., Hall, L. L., Byron, M., McNeil, J. and Lawrence, J. B.** (2006). "The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences." *Proceedings of the National Academy of Sciences of the United States of America* **103**(20): 7688-7693.
- Clemson, C. M., McNeil, J. A., Willard, H. F. and Lawrence, J. B.** (1996). "XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure." *The Journal of cell biology* **132**(3): 259-275.
- Clerc, P. and Avner, P.** (1998). "Role of the region 3' to Xist exon 6 in the counting process of X-chromosome inactivation." *Nature genetics* **19**(3): 249-253.

- Clerc, P. and Avner, P.** (1998). "Role of the region 3' to Xist exon 6 in the counting process of X-chromosome inactivation." *Nat Genet* **19**(3): 249-253.
- Clouet d'Orval, B., d'Aubenton Carafa, Y., Sirand-Pugnet, P., Gallego, M., Brody, E. and Marie, J.** (1991). "RNA secondary structure repression of a muscle-specific exon in HeLa cell nuclear extracts." *Science* **252**(5014): 1823-1828.
- Cohen, H. R. and Panning, B.** (2007). "XIST RNA exhibits nuclear retention and exhibits reduced association with the export factor TAP/NXF1." *Chromosoma* **116**(4): 373-383.
- Coles, L. S., Bartley, M. A., Bert, A., Hunter, J., Polyak, S., Diamond, P., . . . Goodall, G. J.** (2004). "A multi-protein complex containing cold shock domain (Y-box) and polypyrimidine tract binding proteins forms on the vascular endothelial growth factor mRNA. Potential role in mRNA stabilization." *Eur J Biochem* **271**(3): 648-660.
- Cooper, P., Keer, J. T., McCabe, V. M., Hamvas, R. M. J., Brown, S. D. M., Rastan, S. and Brockdorff, N.** (1993). "Physical Mapping of 2000 kb of the Mouse X Chromosome in the Vicinity of the Xist Locus." *Genomics* **15**(3): 570-575.
- Cornacchia, D., Dileep, V., Quivy, J. P., Foti, R., Tili, F., Santarella-Mellwig, R., . . . Buonomo, S. B.** (2012). "Mouse Rif1 is a key regulator of the replication-timing programme in mammalian cells." *EMBO J*.
- Crispino, J. D., Blencowe, B. J. and Sharp, P. A.** (1994). "Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP." *Science* **265**(5180): 1866-1869.
- Crosio, C., Cermakian, N., Allis, C. D. and Sassone-Corsi, P.** (2000). "Light induces chromatin modification in cells of the mammalian circadian clock." *Nat Neurosci* **3**(12): 1241-1247.
- Csankovszki, G., Colette, K., Spahl, K., Carey, J., Snyder, M., Petty, E., . . . Hagstrom, K.** (2009). "Three distinct condensin complexes control *C. elegans* chromosome dynamics." *Current biology : CB* **19**(1): 9-19.
- Csankovszki, G., Nagy, A. and Jaenisch, R.** (2001). "Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation." *Journal of cellular biochemistry* **153**(4): 773-784.
- Csankovszki, G., Panning, B., Bates, B., Pehrson, J. R. and Jaenisch, R.** (1999). "Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation." *Nature genetics* **22**(4): 323-324.
- Damier, L., Domenjoud, L. and Branlant, C.** (1997). "The D1-A2 and D2-A2 pairs of splice sites from human immunodeficiency virus type 1 are highly efficient in vitro, in spite of an unusual branch site." *Biochem Biophys Res Commun* **237**(1): 182-187.
- Davis-Smyth, T., Duncan, R. C., Zheng, T., Michelotti, G. and Levens, D.** (1996). "The far upstream element-binding proteins comprise an ancient family of single-strand DNA-binding transactivators." *J Biol Chem* **271**(49): 31679-31687.
- de Napoles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., . . . Brockdorff, N.** (2004). "Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation." *Developmental Cell* **7**(5): 663-676.
- Dinant, C. and Luijsterburg, M. S.** (2009). "The emerging role of HP1 in the DNA damage response." *Mol Cell Biol* **29**(24): 6335-6340.

- Donohoe, M. E., Silva, S. S., Pinter, S. F., Xu, N. and Lee, J. T.** (2009). "The pluripotency factor Oct4 interacts with Ctfc and also controls X-chromosome pairing and counting." *Nature* **460**(7251): 128-132.
- Donohoe, M. E., Zhang, L. F., Xu, N., Shi, Y. and Lee, J. T.** (2007). "Identification of a Ctfc cofactor, Yy1, for the X chromosome binary switch." *Molecular Cell* **25**(1): 43-56.
- Dover, J., Schneider, J., Tawiah-Boateng, M. A., Wood, A., Dean, K., Johnston, M. and Shilatifard, A.** (2002). "Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6." *J Biol Chem* **277**(32): 28368-28371.
- Dreyfuss, G.** (1986). "Structure and function of nuclear and cytoplasmic ribonucleoprotein particles." *Annu Rev Cell Biol* **2**: 459-498.
- Dreyfuss, G., Kim, V. N. and Kataoka, N.** (2002). "Messenger-RNA-binding proteins and the messages they carry." *Nature reviews. Molecular cell biology* **3**(3): 195-205.
- 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.
- Dreyfuss, G., Swanson, M. S. and Pinol-Roma, S.** (1988). "Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation." *Trends in biochemical sciences* **13**(3): 86-91.
- Du, C., McGuffin, M. E., Dauwalder, B., Rabinow, L. and Mattox, W.** (1998). "Protein phosphorylation plays an essential role in the regulation of alternative splicing and sex determination in *Drosophila*." *Molecular Cell* **2**(6): 741-750.
- Duszczuk, M. M., Wutz, a., Rybin, V. and Sattler, M.** (2011). "The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization." *RNA*.
- Duszczuk, M. M., Zanier, K. and Sattler, M.** (2008). "A NMR strategy to unambiguously distinguish nucleic acid hairpin and duplex conformations applied to a Xist RNA A-repeat." *Nucleic acids research* **36**(22): 7068-7077.
- Eperon, I. C., Ireland, D. C., Smith, R. A., Mayeda, A. and Krainer, A. R.** (1993). "Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF." *EMBO Journal* **12**(9): 3607-3617.
- Fackelmayer, F. O.** (2005). "A stable proteinaceous structure in the territory of inactive X chromosomes." *The Journal of biological chemistry* **280**(3): 1720-1723.
- Felsenfeld, G. and Groudine, M.** (2003). "Controlling the double helix." *Nature* **421**(6921): 448-453.
- Filippova, G. N., Cheng, M. K., Moore, J. M., Truong, J. P., Hu, Y. J., Nguyen, D. K., . . . Disteche, C. M.** (2005). "Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development." *Developmental cell* **8**(1): 31-42.
- Filippova, G. N., Fagerlie, S., Klenova, E. M., Myers, C., Dehner, Y., Goodwin, G., . . . Lobanenkov, V. V.** (1996). "An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes." *Mol Cell Biol* **16**(6): 2802-2813.
- Fitzpatrick, G. V., Pugacheva, E. M., Shin, J.-Y., Abdullaev, Z., Yang, Y., Khatod, K., . . . Higgins, M. J.** (2007). "Allele-specific binding of CTCF to the multipartite imprinting control region KvDMR1." *Molecular and cellular biology* **27**(7): 2636-2647.

- Fluckiger, A. C., Dehay, C. and Savatier, P.** (2003). "[Embryonic stem cells and cell replacement therapies in the nervous system]." *Med Sci (Paris)* **19**(6-7): 699-708.
- Franke, a. and Baker, B. S.** (2000). "Dosage compensation rox!" *Current opinion in cell biology* **12**(3): 351-354.
- Garcia-Blanco, M. A., Jamison, S. F. and Sharp, P. A.** (1989). "Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns." *Genes & Development* **3**(12A): 1874-1886.
- García-Mayoral, M. F., Hollingworth, D., Masino, L., Díaz-Moreno, I., Kelly, G., Gherzi, R., . . . Ramos, A.** (2007). "The structure of the C-terminal KH domains of KSRP reveals a noncanonical motif important for mRNA degradation." *Structure (London, England : 1993)* **15**(4): 485-498.
- Gartler, S. M. and Riggs, A. D.** (1983). "Mammalian x-chromosome inactivation." *Control.*
- Ge, H. and Manley, J. L.** (1990). "A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro." *Cell* **62**(1): 25-34.
- Gelbart, M. E. and Kuroda, M. I.** (2009). "Drosophila dosage compensation: a complex voyage to the X chromosome." *Development* **136**: 1399-1410.
- Gherzi, R., Lee, K.-Y., Briata, P., Wegmüller, D., Moroni, C., Karin, M. and Chen, C.-Y.** (2004). "A KH domain RNA binding protein, KSRP, promotes ARE-directed mRNA turnover by recruiting the degradation machinery." *Molecular cell* **14**(5): 571-583.
- Giannelli, F.** (1963). "The pattern of X-chromosome deoxy-ribonucleic acid synthesis in two women with abnormal sex-chromosome complements." *Lancet* **1**(7286): 863-865.
- Goldknopf, I. L., Taylor, C. W., Baum, R. M., Yeoman, L. C., Olson, M. O., Prestayko, A. W. and Busch, H.** (1975). "Isolation and characterization of protein A24, a "histone-like" non-histone chromosomal protein." *J Biol Chem* **250**(18): 7182-7187.
- Goll, M. G. and Bestor, T. H.** (2005). "Eukaryotic cytosine methyltransferases." *Annu Rev Biochem* **74**: 481-514.
- Gontan, C., Achame, E. M., Demmers, J., Barakat, T. S., Rentmeester, E., van Ijcken, W., . . . Gribnau, J.** (2012). "RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation." *Nature* **485**(7398): 386-390.
- Gonzalez, S., Pisano, D. G. and Serrano, M.** (2008). "Mechanistic principles of chromatin remodeling guided by siRNAs and miRNAs." *Cell Cycle* **7**(16): 2601-2608.
- Goto, H., Tomono, Y., Ajiro, K., Kosako, H., Fujita, M., Sakurai, M., . . . Inagaki, M.** (1999). "Identification of a novel phosphorylation site on histone H3 coupled with mitotic chromosome condensation." *J Biol Chem* **274**(36): 25543-25549.
- Graham, F. L., Smiley, J., Russell, W. C. and Nairn, R.** (1977). "Characteristics of a human cell line transformed by DNA from human adenovirus type 5." *J Gen Virol* **36**(1): 59-74.
- Grant, J., Mahadevaiah, S. K., Khil, P., Sangrithi, M. N., Royo, H., Duckworth, J., . . . Turner, J. M.** (2012). "Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation." *Nature*.
- Graveley, B. R.** (2000). "Sorting out the complexity of SR protein functions." *RNA* **6**(9): 1197-1211.

- Grewal, S. I. S. and Elgin, S. C. R.** (2002). "Heterochromatin: new possibilities for the inheritance of structure." *Current opinion in genetics & development* **12**(2): 178-187.
- Gribnau, J., Luikenhuis, S., Hochedlinger, K., Monkhorst, K. and Jaenisch, R.** (2005). "X chromosome choice occurs independently of asynchronous replication timing." *The Journal of cell biology* **168**(3): 365-373.
- Grimaud, C. and Becker, P. B.** (2009). "The dosage compensation complex shapes the conformation of the X chromosome in *Drosophila*." *Genes & development* **23**(21): 2490-2495.
- Grumbach, M. M., Morishima, A. and Taylor, J. H.** (1963). "Human sex chromosome abnormalities in relation to DNA replication and heterochromatinization." *Proceedings of the National Academy of Sciences of the United States of America* **49**(5): 581-589.
- Gurley, L. R., D'Anna, J. A., Barham, S. S., Deaven, L. L. and Tobey, R. A.** (1978). "Histone phosphorylation and chromatin structure during mitosis in Chinese hamster cells." *Eur J Biochem* **84**(1): 1-15.
- Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., . . . Lander, E. S.** (2009). "Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals." *Nature* **458**(7235): 223-227.
- Hadjur, S., Williams, L. M., Ryan, N. K., Cobb, B. S., Sexton, T., Fraser, P., . . . Merckenschlager, M.** (2009). "Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus." *Nature* **460**(7253): 410-413.
- Hallacli, E. and Akhtar, A.** (2009). "X chromosomal regulation in flies: when less is more." *Chromosome research* **17**(5): 603-619.
- Han, J., Kim, D. and Morris, K. V.** (2007). "Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells." *Proc Natl Acad Sci U S A* **104**(30): 12422-12427.
- Hanahan, D.** (1983). "Studies on transformation of *Escherichia coli* with plasmids." *J Mol Biol* **166**(4): 557-580.
- Hartman, T., Stead, K., Koshland, D. and Guacci, V.** (2000). "Pds5p is an essential chromosomal protein required for both sister chromatid cohesion and condensation in *Saccharomyces cerevisiae*." *J Cell Biol* **151**(3): 613-626.
- Hassan, A. H., Prochasson, P., Neely, K. E., Galasinski, S. C., Chandy, M., Carrozza, M. J. and Workman, J. L.** (2002). "Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes." *Cell* **111**(3): 369-379.
- Hassig, C. A., Fleischer, T. C., Billin, A. N., Schreiber, S. L. and Ayer, D. E.** (1997). "Histone deacetylase activity is required for full transcriptional repression by mSin3A." *Cell* **89**(3): 341-347.
- Hastings, M. L. and Krainer, A. R.** (2001). "Pre-mRNA splicing in the new millennium." *Current Opinion in Cell Biology* **13**(3): 302-309.
- Hawkins, P. G., Santoso, S., Adams, C., Anest, V. and Morris, K. V.** (2009). "Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells." *Nucleic Acids Res* **37**(9): 2984-2995.
- Hayano, M., Kano, Y., Matsumoto, S., Renard-Guillet, C., Shirahige, K. and Masai, H.** (2012). "Rif1 is a global regulator of timing of replication origin firing in fission yeast." *Genes Dev* **26**(2): 137-150.

- Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, C. D. and Spector, D. L.** (2001). "Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation." *Cell* **107**(6): 727-738.
- Helbig, R. and Fackelmayer, F. O.** (2003). "Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain." *Chromosoma* **112**(4): 173-182.
- Henry, K. W., Wyce, A., Lo, W. S., Duggan, L. J., Emre, N. C., Kao, C. F., . . . Berger, S. L.** (2003). "Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8." *Genes Dev* **17**(21): 2648-2663.
- Hermann, A., Goyal, R. and Jeltsch, A.** (2004). "The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites." *J Biol Chem* **279**(46): 48350-48359.
- Hernandez-Munoz, I., Lund, A. H., van der Stoop, P., Boutsma, E., Muijers, I., Verhoeven, E., . . . van Lohuizen, M.** (2005). "Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin E3 ligase." *Proceedings of the National Academy of Sciences of the United States of America* **102**(21): 7635-7640.
- Hodgkin, J.** (1987). "Sex determination and dosage compensation in *Caenorhabditis elegans*." *Annual review of genetics* **21**: 133-154.
- Hoki, Y., Kimura, N., Kanbayashi, M., Amakawa, Y., Ohhata, T., Sasaki, H. and Sado, T.** (2009). "A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse." *Development (Cambridge, England)* **136**(1): 139-146.
- Hong, Y. K., Ontiveros, S. D., Chen, C. and Strauss, W. M.** (1999). "A new structure for the murine Xist gene and its relationship to chromosome choice/counting during X-chromosome inactivation." *Proceedings of the National Academy of Sciences of the United States of America* **96**(12): 6829-6834.
- Hong, Y. K., Ontiveros, S. D. and Strauss, W. M.** (2000). "A revision of the human XIST gene organization and structural comparison with mouse Xist." *Mamm Genome* **11**(3): 220-224.
- Hsu, J.-m., Huang, J., Meluh, P. B. and Laurent, B. C.** (2003). "The Yeast RSC Chromatin-Remodeling Complex Is Required for Kinetochores Function in Chromosome Segregation." *Society* **23**(9): 3202-3215.
- Huang, J. and Laurent, B. C.** (2004). "A role for the RSC chromatin remodeler in regulating cohesion of sister chromatid arms." *Cell Cycle* **3**(August): 973-975.
- Huynh, K. D. and Lee, J. T.** (2003). "Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos." *Nature*: 857-862.
- Ideraabdullah, F. Y., Vigneau, S. and Bartolomei, M. S.** (2008). "Genomic imprinting mechanisms in mammals." *Mutation research* **647**(1-2): 77-85.
- Jang, S. K. and Wimmer, E.** (1990). "Cap-independent translation of encephalomyocarditis virus RNA: structural elements of the internal ribosomal entry site and involvement of a cellular 57-kD RNA-binding protein." *Genes & Development* **4**(9): 1560-1572.
- Janowski, B. A., Huffman, K. E., Schwartz, J. C., Ram, R., Nordsell, R., Shames, D. S., . . . Corey, D. R.** (2006). "Involvement of AGO1 and AGO2 in mammalian transcriptional silencing." *Nat Struct Mol Biol* **13**(9): 787-792.

- Jans, J., Gladden, J. M., Ralston, E. J., Pickle, C. S., Michel, A. H., Pferdehirt, R. R., . . . Meyer, B. J.** (2009). "A condensin-like dosage compensation complex acts at a distance to control expression throughout the genome." *Genes & development* **23**(5): 602-618.
- Jenuwein, T. and Allis, C. D.** (2001). "Translating the histone code." *Science* **293**(5532): 1074-1080.
- Jeon, Y. and Lee, J. T.** (2011). "YY1 Tethers Xist RNA to the Inactive X Nucleation Center." *Cell* **146**(1): 119-133.
- Johnston, C. M., Nesterova, T. B., Formstone, E. J., Newall, A. E., Duthie, S. M., Sheardown, S. A. and Brockdorff, N.** (1998). "Developmentally regulated Xist promoter switch mediates initiation of X inactivation." *Cell* **94**(6): 809-817.
- Johnston, C. M., Newall, A. E., Brockdorff, N. and Nesterova, T. B.** (2002). "Enox, a novel gene that maps 10 kb upstream of Xist and partially escapes X inactivation." *Genomics* **80**(2): 236-244.
- Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., . . . Wolffe, A. P.** (1998). "Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription." *Nat Genet* **19**(2): 187-191.
- Jonkers, I., Barakat, T. S., Achame, E. M., Monkhorst, K., Kenter, A., Rentmeester, E., . . . Gribnau, J.** (2009). "RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation." *Cell* **139**(5): 999-1011.
- Jonkers, I., Monkhorst, K., Rentmeester, E., Grootegoed, J. A., Grosveld, F. and Gribnau, J.** (2008). "Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle." *Molecular and cellular biology* **28**(18): 5583-5594.
- Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N. and Moore, M. J.** (2002). "Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis." *RNA* **8**(4): 426-439.
- Jurica, M. S. and Moore, M. J.** (2002). "Capturing splicing complexes to study structure and mechanism." *Methods* **28**(3): 336-345.
- Kafasla, P., Mickleburgh, I., Llorian, M., Coelho, M., Gooding, C., Cherny, D., . . . Smith, C. W.** (2012). "Defining the roles and interactions of PTB." *Biochemical Society transactions* **40**(4): 815-820.
- Kafasla, P., Morgner, N., Pöyry, T. a. a., Curry, S., Robinson, C. V. and Jackson, R. J.** (2009). "Polypyrimidine tract binding protein stabilizes the encephalomyocarditis virus IRES structure via binding multiple sites in a unique orientation." *Molecular cell* **34**(5): 556-568.
- Kalantry, S. and Magnuson, T.** (2006). "The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation." *PLoS genetics* **2**(5): e66-e66.
- Kaludov, N. K. and Wolffe, A. P.** (2000). "MeCP2 driven transcriptional repression in vitro: selectivity for methylated DNA, action at a distance and contacts with the basal transcription machinery." *Nucleic Acids Res* **28**(9): 1921-1928.
- Kanopka, A., Muhlemann, O. and Akusjarvi, G.** (1996). "Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA." *Nature* **381**(6582): 535-538.
- Kao, C. F., Hillyer, C., Tsukuda, T., Henry, K., Berger, S. and Osley, M. A.** (2004). "Rad6 plays a role in transcriptional activation through ubiquitylation of histone H2B." *Genes Dev* **18**(2): 184-195.

- Keohane, A. M., O'Neill, L. P., Belyaev, N. D., Lavender, J. S. and Turner, B. M.** (1996). "X-Inactivation and histone H4 acetylation in embryonic stem cells." *Developmental biology* **180**(2): 618-630.
- Kessler, M. M., Henry, M. F., Shen, E., Zhao, J., Gross, S., Silver, P. A. and Moore, C. L.** (1997). "Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast." *Genes & Development* **11**(19): 2545-2556.
- Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., . . . Rinn, J. L.** (2009). "Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression." *Proceedings of the National Academy of Sciences of the United States of America* **106**(28): 11667-11672.
- Kim, D. H., Saetrom, P., Snove, O., Jr. and Rossi, J. J.** (2008). "MicroRNA-directed transcriptional gene silencing in mammalian cells." *Proc Natl Acad Sci U S A* **105**(42): 16230-16235.
- Kim, D. H., Villeneuve, L. M., Morris, K. V. and Rossi, J. J.** (2006). "Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells." *Nat Struct Mol Biol* **13**(9): 793-797.
- Kishore, S., Khanna, A. and Stamm, S.** (2008). "Rapid generation of splicing reporters with pSpliceExpress." *Gene* **427**(1-2): 104-110.
- Kishore, S., Khanna, A., Zhang, Z., Hui, J., Balwierz, P. J., Stefan, M., . . . Stamm, S.** (2010). "The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing." *Hum Mol Genet* **19**(7): 1153-1164.
- Klose, R. J. and Zhang, Y.** (2007). "Regulation of histone methylation by demethylination and demethylation." *Nature reviews. Molecular cell biology* **8**(4): 307-318.
- Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T. and Wutz, A.** (2004). "A chromosomal memory triggered by Xist regulates histone methylation in X inactivation." *PLoS biology* **2**(7): E171-E171.
- Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia-Blanco, M. A. and Manley, J. L.** (1994). "Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors." *Nature* **368**(6467): 119-124.
- Kosinski, P. A., Laughlin, J., Singh, K. and Covey, L. R.** (2003). "A complex containing polypyrimidine tract-binding protein is involved in regulating the stability of CD40 ligand (CD154) mRNA." *J Immunol* **170**(2): 979-988.
- Koziol, M. J. and Rinn, J. L.** (2010). "RNA traffic control of chromatin complexes." *Current opinion in genetics & development* **20**(2): 142-148.
- Krainer, A. R. and Maniatis, T.** (1985). "Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro." *Cell* **42**(3): 725-736.
- Kueng, S., Hegemann, B., Peters, B. H., Lipp, J. J., Schleiffer, A., Mechtler, K. and Peters, J. M.** (2006). "Wapl controls the dynamic association of cohesin with chromatin." *Cell* **127**(5): 955-967.
- Lachner, M.** (2002). "Epigenetics : SUPERMAN Dresses Up DNA and histone methylation have been implicated." *Science* **12**(02): 434-436.

- Laherty, C. D., Yang, W. M., Sun, J. M., Davie, J. R., Seto, E. and Eisenman, R. N.** (1997). "Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression." *Cell* **89**(3): 349-356.
- Law, J. A. and Jacobsen, S. E.** (2010). "Establishing, maintaining and modifying DNA methylation patterns in plants and animals." *Nat Rev Genet* **11**(3): 204-220.
- Lechner, M. S., Begg, G. E., Speicher, D. W. and Rauscher, F. J., 3rd** (2000). "Molecular determinants for targeting heterochromatin protein 1-mediated gene silencing: direct chromoshadow domain-KAP-1 corepressor interaction is essential." *Mol Cell Biol* **20**(17): 6449-6465.
- Lee, J. T.** (2005). "Regulation of X-chromosome counting by Tsix and Xite sequences." *Science* **309**(5735): 768-771.
- Lee, J. T., Davidow, L. S. and Warshawsky, D.** (1999). "Tsix, a gene antisense to Xist at the X-inactivation centre." *Nature genetics* **21**(4): 400-404.
- Lee, J. T. and Jaenisch, R.** (1997). "Long-range cis effects of ectopic X-inactivation centres on a mouse autosome." *Nature* **386**(6622): 275-279.
- Lee, J. T., Strauss, W. M., Dausman, J. A. and Jaenisch, R.** (1996). "A 450 kb Transgene Displays Properties of the Mammalian X-Inactivation Center." *Cell* **86**(1): 83-94.
- Lemaire, R., Winne, A., Sarkissian, M. and Lafyatis, R.** (1999). "SF2 and SRp55 regulation of CD45 exon 4 skipping during T cell activation." *Eur J Immunol* **29**(3): 823-837.
- Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F. and Bird, A.** (1992). "Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA." *Cell* **69**(6): 905-914.
- Libri, D., Piseri, A. and Fiszman, M. Y.** (1991). "Tissue-specific splicing in vivo of the beta-tropomyosin gene: dependence on an RNA secondary structure." *Science* **252**(5014): 1842-1845.
- Libri, D., Stutz, F., McCarthy, T. and Rosbash, M.** (1995). "RNA structural patterns and splicing: molecular basis for an RNA-based enhancer." *Rna* **1**(4): 425-436.
- Lieb, J. D., Albrecht, M. R., Chuang, P. T. and Meyer, B. J.** (1998). "MIX-1: an essential component of the *C. elegans* mitotic machinery executes X chromosome dosage compensation." *Cell* **92**(2): 265-277.
- Lieb, J. D., Capowski, E. E., Meneely, P. and Meyer, B. J.** (1996). "DPY-26, a link between dosage compensation and meiotic chromosome segregation in the nematode." *Science* **274**(5293): 1732-1736.
- Liu, H. X., Zhang, M. and Krainer, A. R.** (1998). "Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins." *Genes & Development* **12**(13): 1998-2012.
- Llorian, M., Schwartz, S., Clark, T. a., Hollander, D., Tan, L.-Y., Spellman, R., . . . Smith, C. W. J.** (2010). "Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB." *Nature structural & molecular biology* **17**(9): 1114-1123.
- Long, J. C. and Caceres, J. F.** (2009). "The SR protein family of splicing factors: master regulators of gene expression." *Biochem J* **417**(1): 15-27.

- Lopes, A. M., Arnold-Croop, S. E., Amorim, A. and Carrel, L.** (2011). "Clustered transcripts that escape X inactivation at mouse XqD." *Mamm Genome* **22**(9-10): 572-582.
- Lorch, Y., Zhang, M. and Kornberg, R. D.** (1999). "Histone Octamer Transfer by a Chromatin-Remodeling Complex." *Cell* **96**: 389-392.
- 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." *Molecular and cellular biology* **19**(1): 78-85.
- Lyon, M. F.** (1961). "Gene action in the X-chromosome of the mouse (*Mus musculus* L.)." *Nature* **190**: 372-373.
- Lyon, M. F.** (1998). "X-chromosome inactivation spreads itself: effects in autosomes." *American journal of human genetics* **63**(1): 17-19.
- Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., . . . Branlant, C.** (2010). "2-D structure of the A region of Xist RNA and its implication for PRC2 association." *PLoS biology* **8**(1): e1000276-e1000276.
- Maison, C. and Almouzni, G.** (2004). "HP1 and the dynamics of heterochromatin maintenance." *Nat Rev Mol Cell Biol* **5**(4): 296-304.
- Mak, W., Baxter, J., Silva, J., Newall, A. E., Otte, A. P. and Brockdorff, N.** (2002). "Mitotically stable association of polycomb group proteins *eed* and *enx1* with the inactive x chromosome in trophoblast stem cells." *Current biology : CB* **12**(12): 1016-1020.
- Mak, W., Nesterova, T. B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A. P. and Brockdorff, N.** (2004). "Reactivation of the paternal X chromosome in early mouse embryos." *Science (New York, N.Y.)* **303**(5658): 666-669.
- Mancini-Dinardo, D., Steele, S. J. S., Levorse, J. M., Ingram, R. S. and Tilghman, S. M.** (2006). "Elongation of the *Kcnqlot1* transcript is required for genomic imprinting of neighboring genes." *Genes & development* **20**(10): 1268-1282.
- Marahrens, Y., Loring, J. and Jaenisch, R.** (1998). "Role of the Xist Gene in X Chromosome Choosing." *Cell* **92**(5): 657-664.
- Marahrens, Y., Panning, B., Dausman, J., Strauss, W. and Jaenisch, R.** (1997). "Xist-deficient mice are defective in dosage compensation but not spermatogenesis." *Genes & Development* **11**(2): 156-166.
- Marchand, V., Mereau, A., Jacquenet, S., Thomas, D., Mougin, A., Gattoni, R., . . . Branlant, C.** (2002). "A Janus splicing regulatory element modulates HIV-1 tat and rev mRNA production by coordination of hnRNP A1 cooperative binding." *J Mol Biol* **323**(4): 629-652.
- Marin, I., Siegal, M. L. and Baker, B. S.** (2000). "The evolution of dosage-compensation mechanisms." *Bioessays* **22**(12): 1106-1114.
- 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." *Molecular and cellular biology* **20**(20): 7463-7479.
- Masui, S., Ohtsuka, S., Yagi, R., Takahashi, K., Ko, M. S. and Niwa, H.** (2008). "Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells." *BMC Dev Biol* **8**: 45.

- Mermoud, J. E., Popova, B., Peters, A. H., Jenuwein, T. and Brockdorff, N.** (2002). "Histone H3 lysine 9 methylation occurs rapidly at the onset of random X chromosome inactivation." *Current biology : CB* **12**(3): 247-251.
- Meyer, B. J.** (2010). "Targeting X chromosomes for repression." *Curr Opin Genet Dev* **20**(2): 179-189.
- Min, H., Turck, C. W., Nikolic, J. M. and Black, D. L.** (1997). "A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer." *Genes & Development* **11**(8): 1023-1036.
- Mitchell, S. A., Spriggs, K. A., Coldwell, M. J., Jackson, R. J. and Willis, A. E.** (2003). "The Apaf-1 internal ribosome entry segment attains the correct structural conformation for function via interactions with PTB and unr." *Molecular Cell* **11**(3): 757-771.
- Modrek, B. and Lee, C.** (2002). "A genomic view of alternative splicing." *Nature Genetics* **30**(1): 13-19.
- Mohrmann, L. and Verrijzer, C. P.** (2005). "Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes." *Biochimica et biophysica acta* **1681**(2-3): 59-73.
- Monkhorst, K., Jonkers, I., Rentmeester, E., Grosveld, F. and Gribnau, J.** (2008). "X inactivation counting and choice is a stochastic process: evidence for involvement of an X-linked activator." *Cell* **132**(3): 410-421.
- Moore, K. L.** (1966). "The sex chromatin." *WB Saunders, Philadelphia*.
- Moore, M. J. and Sharp, P. A.** (1993). "Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing." *Nature* **365**(6444): 364-368.
- Moreira de Mello, J. C., de Araújo, E. S. S., Stabellini, R., Fraga, A. M., de Souza, J. E. S., Sumita, D. R., . . . Pereira, L. V.** (2010). "Random X inactivation and extensive mosaicism in human placenta revealed by analysis of allele-specific gene expression along the X chromosome." *PloS one* **5**(6): e10947-e10947.
- Morillon, A., Karabetsov, N., O'Sullivan, J., Kent, N., Proudfoot, N. and Mellor, J.** (2003). "Isw1 chromatin remodeling ATPase coordinates transcription elongation and termination by RNA polymerase II." *Cell* **115**(4): 425-435.
- Morillon, A., O'Sullivan, J., A, A., Proudfoot, N. and Mellor, J.** (2003). "Regulation of elongating RNA polymerase II by forkhead transcription factors in yeast." *Science* **300**(5618): 492-495.
- Morris, K. V.** (2005). "siRNA-mediated transcriptional gene silencing: the potential mechanism and a possible role in the histone code." *Cell Mol Life Sci* **62**(24): 3057-3066.
- Nagano, T., Mitchell, J. A., Sanz, L. A., Pauler, F. M., Ferguson-Smith, A. C., Feil, R. and Fraser, P.** (2008). "The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin." *Science* **322**(5908): 1717-1717.
- Nagel, R. J., Lancaster, A. M. and Zahler, A. M.** (1998). "Specific binding of an exonic splicing enhancer by the pre-mRNA splicing factor SRp55." *Rna* **4**(1): 11-23.
- Nakielny, S. and Dreyfuss, G.** (1999). "Transport of proteins and RNAs in and out of the nucleus." *Cell* **99**(7): 677-690.
- Nan, X., Campoy, F. J. and Bird, A.** (1997). "MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin." *Cell* **88**(4): 471-481.

- Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N. and Bird, A.** (1998). "Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex." *Nature* **393**(6683): 386-389.
- Navarro, P., Chambers, I., Karwacki-Neisius, V., Chureau, C., Morey, C., Rougeulle, C. and Avner, P.** (2008). "Molecular coupling of Xist regulation and pluripotency." *Science (New York, N.Y.)* **321**(5896): 1693-1695.
- Navarro, P., Moffat, M., Mullin, N. P. and Chambers, I.** (2011). "The X-inactivation trans-activator Rnf12 is negatively regulated by pluripotency factors in embryonic stem cells." *Human Genetics* **130**(2): 255-264.
- Navarro, P., Oldfield, A., Legoupi, J., Festuccia, N., Dubois, A., Attia, M., . . . Avner, P.** (2010). "Molecular coupling of Tsix regulation and pluripotency." *Nature* **468**(7322): 457-460.
- Navarro, P., Page, D. R., Avner, P. and Rougeulle, C.** (2006). "Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program." *Genes & Development* **20**(20): 2787-2792.
- Navarro, P., Pichard, S., Ciaudo, C., Avner, P. and Rougeulle, C.** (2005). "Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation." *Genes & Development* **19**(12): 1474-1484.
- Nesterova, T. B., Slobodyanyuk, S. Y., Elisaphenko, E. a., Shevchenko, a. I., Johnston, C., Pavlova, M. E., . . . Zakian, S. M.** (2001). "Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence." *Genome research* **11**(5): 833-849.
- Ng, H. H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., . . . Bird, A.** (1999). "MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex." *Nat Genet* **23**(1): 58-61.
- Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., . . . Heard, E.** (2012). "Spatial partitioning of the regulatory landscape of the X-inactivation centre." *Nature* **485**(7398): 381-385.
- Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., . . . Allain, F. H.** (2005). "Structure of PTB bound to RNA: specific binding and implications for splicing regulation." *Science* **309**(5743): 2054-2057.
- Ogawa, Y. and Lee, J. T.** (2003). "Xite, X-inactivation intergenic transcription elements that regulate the probability of choice." *Molecular cell* **11**(3): 731-743.
- Ohno, S., Becak, W. and Becak, M. L.** (1964). "X-autosome ratio and the behavior pattern of individual X-chromosomes in placental mammals." *Chromosoma* **15**: 14-30.
- Okamoto, I., Arnaud, D., Le Baccon, P., Otte, A. P., Disteche, C. M., Avner, P. and Heard, E.** (2005). "Evidence for de novo imprinted X-chromosome inactivation independent of meiotic inactivation in mice." *Nature* **438**(7066): 369-373.
- Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D. and Heard, E.** (2004). "Epigenetic dynamics of imprinted X inactivation during early mouse development." *Science (New York, N.Y.)* **303**(5658): 644-649.

- Okano, M., Bell, D. W., Haber, D. A. and Li, E.** (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." *Cell* **99**(3): 247-257.
- Okano, M., Xie, S. and Li, E.** (1998). "Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases." *Nat Genet* **19**(3): 219-220.
- Osley, M. A.** (2006). "Regulation of histone H2A and H2B ubiquitylation." *Brief Funct Genomic Proteomic* **5**(3): 179-189.
- Ostareck-Lederer, A. and Ostareck, D. H.** (2004). "Control of mRNA translation and stability in haematopoietic cells: the function of hnRNPs K and E1/E2." *Biol Cell* **96**(6): 407-411.
- Panizza, S., Tanaka, T., Hochwagen, A., Eisenhaber, F. and Nasmyth, K.** (2000). "Pds5 cooperates with cohesin in maintaining sister chromatid cohesion." *Curr Biol* **10**(24): 1557-1564.
- Panning, B., Dausman, J. and Jaenisch, R.** (1997). "X chromosome inactivation is mediated by Xist RNA stabilization." *Cell* **90**(5): 907-916.
- Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., . . . Merckenschlager, M.** (2008). "Cohesins functionally associate with CTCF on mammalian chromosome arms." *Cell* **132**(3): 422-433.
- Park, Y., Kelley, R. L., Oh, H., Kuroda, M. I. and Meller, V. H.** (2002). "Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins." *Science* **298**: 1620-1623.
- Park, Y., Oh, H., Meller, V. H. and Kuroda, M. I.** (2005). "Variable Splicing of Noncoding roX2 RNAs Influences Targeting of MSL Dosage Compensation Complexes in Drosophila." *RNA Biology* **2**(4): 157-164.
- Paronetto, M. P., Achsel, T., Massiello, A., Chalfant, C. E. and Sette, C.** (2007). "The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x." *The Journal of cell biology* **176**(7): 929-939.
- Peng, H., Begg, G. E., Schultz, D. C., Friedman, J. R., Jensen, D. E., Speicher, D. W. and Rauscher, F. J., 3rd** (2000). "Reconstitution of the KRAB-KAP-1 repressor complex: a model system for defining the molecular anatomy of RING-B box-coiled-coil domain-mediated protein-protein interactions." *J Mol Biol* **295**(5): 1139-1162.
- Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. and Brockdorff, N.** (1996). "Requirement for Xist in X chromosome inactivation." *Nature* **379**(6561): 131-137.
- Perez, I., McAfee, J. G. and Patton, J. G.** (1997). "Multiple RRM domains contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein." *Biochemistry* **36**(39): 11881-11890.
- Peters, A. H., Mermoud, J. E., O'Carroll, D., Pagani, M., Schweizer, D., Brockdorff, N. and Jenuwein, T.** (2002). "Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin." *Nature Genetics* **30**(1): 77-80.
- Peters, J. M., Tedeschi, A. and Schmitz, J.** (2008). "The cohesin complex and its roles in chromosome biology." *Genes Dev* **22**(22): 3089-3114.
- Petoukhov, M. V., Monie, T. P., Allain, F. H., Matthews, S., Curry, S. and Svergun, D. I.** (2006). "Conformation of polypyrimidine tract binding protein in solution." *Structure* **14**(6): 1021-1027.

- Pham, A. D. and Sauer, F.** (2000). "Ubiquitin-activating/conjugating activity of TAFII250, a mediator of activation of gene expression in *Drosophila*." *Science* **289**(5488): 2357-2360.
- Pinol-Roma, S. and Dreyfuss, G.** (1992). "Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm." *Nature* **355**(6362): 730-732.
- Plenefisch, J. D., DeLong, L. and Meyer, B. J.** (1989). "Genes that implement the hermaphrodite mode of dosage compensation in *Caenorhabditis elegans*." *Genetics* **121**(1): 57-76.
- Popova, B. C., Tada, T., Takagi, N., Brockdorff, N. and Nesterova, T. B.** (2006). "Attenuated spread of X-inactivation in an X;autosome translocation." *Proceedings of the National Academy of Sciences* **103**(20): 7706-7711.
- Pozzoli, U. and Sironi, M.** (2005). "Silencers regulate both constitutive and alternative splicing events in mammals." *Cell Mol Life Sci* **62**(14): 1579-1604.
- Preuss, U., Landsberg, G. and Scheidtmann, K. H.** (2003). "Novel mitosis-specific phosphorylation of histone H3 at Thr11 mediated by Dlk/ZIP kinase." *Nucleic Acids Res* **31**(3): 878-885.
- Pugacheva, E. M., Tiwari, V. K., Abdullaev, Z., Vostrov, A. A., Flanagan, P. T., Quitschke, W. W., . . . Lobanenko, V. V.** (2005). "Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation." *Hum Mol Genet* **14**(7): 953-965.
- Rastan, S.** (1983). "Non-random X-chromosome inactivation in mouse X-autosome translocation embryos--location of the inactivation centre." *Journal of embryology and experimental morphology* **78**: 1-22.
- Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., . . . Jenuwein, T.** (2000). "Regulation of chromatin structure by site-specific histone H3 methyltransferases." *Nature* **406**(6796): 593-599.
- Regha, K., Sloane, M. a., Huang, R., Pauler, F. M., Warczok, K. E., Melikant, B., . . . Barlow, D. P.** (2007). "Active and repressive chromatin are interspersed without spreading in an imprinted gene cluster in the mammalian genome." *Molecular cell* **27**(3): 353-366.
- Robertson, K. D. and Wolffe, A. P.** (2000). "DNA methylation in health and disease." *Nat Rev Genet* **1**(1): 11-19.
- Ropers, D., Ayadi, L., Gattoni, R., Jacquenet, S., Damier, L., Branlant, C. and Stevenin, J.** (2004). "Differential effects of the SR proteins 9G8, SC35, ASF/SF2, and SRp40 on the utilization of the A1 to A5 splicing sites of HIV-1 RNA." *J Biol Chem* **279**(29): 29963-29973.
- Rougeulle, C., Chaumeil, J., Sarma, K., David, C., Reinberg, D., Avner, P., . . . Allis, C. D.** (2004). "Differential Histone H3 Lys-9 and Lys-27 Methylation Profiles on the X Chromosome Differential Histone H3 Lys-9 and Lys-27 Methylation Profiles on the X Chromosome †." **24**(12).
- Royce-Tolland, M. E., Andersen, A. a., Koyfman, H. R., Talbot, D. J., Wutz, A., Tonks, I. D., . . . Panning, B.** (2010). "The A-repeat links ASF/SF2-dependent Xist RNA processing with random choice during X inactivation." *Nature structural & molecular biology* **17**(8): 948-954.
- Sado, T., Fenner, M. H., Tan, S. S., Tam, P., Shioda, T. and Li, E.** (2000). "X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation." *Developmental biology* **225**(2): 294-303.

- Sado, T., Hoki, Y. and Sasaki, H.** (2005). "Tsix silences Xist through modification of chromatin structure." *Developmental Cell* **9**(1): 159-165.
- Sado, T., Hoki, Y. and Sasaki, H.** (2006). "Tsix defective in splicing is competent to establish Xist silencing." *Development* **133**(24): 4925-4931.
- Sarma, K. and Reinberg, D.** (2005). "Histone variants meet their match." *Nat Rev Mol Cell Biol* **6**(2): 139-149.
- Sassone-Corsi, P., Mizzen, C. A., Cheung, P., Crosio, C., Monaco, L., Jacquot, S., . . . Allis, C. D.** (1999). "Requirement of Rsk-2 for epidermal growth factor-activated phosphorylation of histone H3." *Science* **285**(5429): 886-891.
- Scherer, S. F.** (1953). "The utilization of a pure strain of mammalian cells (Earle) for the cultivation of viruses in vitro. I. Multiplication of pseudorabies and herpes simplex viruses." *Am J Pathol* **29**(1): 113-137.
- Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. and Rauscher, F. J., 3rd** (2002). "SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins." *Genes Dev* **16**(8): 919-932.
- Schultz, D. C., Friedman, J. R. and Rauscher, F. J.** (2001). "Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD." *Genes & development* **15**(4): 428-443.
- Schwartz, J. C., Younger, S. T., Nguyen, N. B., Hardy, D. B., Monia, B. P., Corey, D. R. and Janowski, B. A.** (2008). "Antisense transcripts are targets for activating small RNAs." *Nat Struct Mol Biol* **15**(8): 842-848.
- Sheardown, S. a., Duthie, S. M., Johnston, C. M., Newall, a. E., Formstone, E. J., Arkell, R. M., . . . Brockdorff, N.** (1997). "Stabilization of Xist RNA mediates initiation of X chromosome inactivation." *Cell* **91**(1): 99-107.
- Sheardown, S. A., Newall, A. E., Norris, D. P., Rastan, S. and Brockdorff, N.** (1997). "Regulatory elements in the minimal promoter region of the mouse Xist gene." *Gene* **203**(2): 159-168.
- Shen, H. and Green, M. R.** (2004). "A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly." *Molecular Cell* **16**(3): 363-373.
- Shen, H., Kan, J. L., Ghigna, C., Biamonti, G. and Green, M. R.** (2004). "A single polypyrimidine tract binding protein (PTB) binding site mediates splicing inhibition at mouse IgM exons M1 and M2." *RNA* **10**(5): 787-794.
- Shi, H., Hoffman, B. E. and Lis, J. T.** (1997). "A specific RNA hairpin loop structure binds the RNA recognition motifs of the Drosophila SR protein B52." *Mol Cell Biol* **17**(5): 2649-2657.
- Shi, W., Wang, H., Pan, G., Geng, Y., Guo, Y. and Pei, D.** (2006). "Regulation of the pluripotency marker Rex-1 by Nanog and Sox2." *J Biol Chem* **281**(33): 23319-23325.
- Shibata, S. and Lee, J. T.** (2004). "Tsix transcription- versus RNA-based mechanisms in Xist repression and epigenetic choice." *Current biology : CB* **14**(19): 1747-1754.

- Shin, J.-Y., Fitzpatrick, G. V. and Higgins, M. J.** (2008). "Two distinct mechanisms of silencing by the KvDMR1 imprinting control region." *The EMBO journal* **27**(1): 168-178.
- Shin, J., Bossenz, M., Chung, Y., Ma, H., Byron, M., Taniguchi-Ishigaki, N., . . . Bach, I.** (2010). "Maternal Rnf12/RLIM is required for imprinted X-chromosome inactivation in mice." *Nature* **467**(7318): 977-981.
- Sif, S., Saurin, A. J., Imbalzano, A. N. and Kingston, R. E.** (2001). "Purification and characterization of mSin3A-containing Brg1 and hBrm chromatin remodeling complexes." *Genes Dev* **15**(5): 603-618.
- Simmler, M. C., Cunningham, D. B., Clerc, P., Vermat, T., Caudron, B., Cruaud, C., . . . Avner, P.** (1996). "A 94 kb genomic sequence 3' to the murine Xist gene reveals an AT rich region containing a new testis specific gene Tsx." *Human Molecular Genetics* **5**(11): 1713-1726.
- Sleutels, F., Zwart, R. and Barlow, D. P.** (2002). "The non-coding Air RNA is required for silencing autosomal imprinted genes." *Nature* **415**(6873): 810-813.
- Smith, C. W. and Valcarcel, J.** (2000). "Alternative pre-mRNA splicing: the logic of combinatorial control." *Trends in Biochemical Sciences* **25**(8): 381-388.
- Spellman, R. and Smith, C. W.** (2006). "Novel modes of splicing repression by PTB." *Trends in biochemical sciences* **31**(2): 73-76.
- Staknis, D. and Reed, R.** (1994). "SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex." *Molecular and cellular biology* **14**(11): 7670-7682.
- Starmer, J. and Magnuson, T.** (2009). "A new model for random X chromosome inactivation." *Development (Cambridge, England)* **136**(1): 1-10.
- Stavropoulos, N., Rowntree, R. K. and Lee, J. T.** (2005). "Identification of developmentally specific enhancers for Tsix in the regulation of X chromosome inactivation." *Molecular and cellular biology* **25**(7): 2757-2769.
- Stead, K., Aguilar, C., Hartman, T., Drexel, M., Meluh, P. and Guacci, V.** (2003). "Pds5p regulates the maintenance of sister chromatid cohesion and is sumoylated to promote the dissolution of cohesion." *J Cell Biol* **163**(4): 729-741.
- Strahl, B. D. and Allis, C. D.** (2000). "The language of covalent histone modifications." *Nature* **403**(6765): 41-45.
- Studier, F. W. and Moffatt, B. A.** (1986). "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes." *J Mol Biol* **189**(1): 113-130.
- Sugawara, O., Takagi, N. and Sasaki, M.** (1983). "Allocyclic early replicating X chromosome in mice: genetic inactivity and shift into a late replicator in early embryogenesis." *Chromosoma* **88**(2): 133-138.
- Sumanasekera, C., Kelemen, O., Beullens, M., Aubol, B. E., Adams, J. A., Sunkara, M., . . . Stamm, S.** (2012). "C6 pyridinium ceramide influences alternative pre-mRNA splicing by inhibiting protein phosphatase-1." *Nucleic Acids Res* **40**(9): 4025-4039.
- Sun, B. K., Deaton, A. M. and Lee, J. T.** (2006). "A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization." *Molecular Cell* **21**(5): 617-628.
- Sun, Z. W. and Allis, C. D.** (2002). "Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast." *Nature* **418**(6893): 104-108.

- Suzuki, K., Juelich, T., Lim, H., Ishida, T., Watanebe, T., Cooper, D. A., . . . Kelleher, A. D.** (2008). "Closed chromatin architecture is induced by an RNA duplex targeting the HIV-1 promoter region." *J Biol Chem* **283**(34): 23353-23363.
- Tacke, R. and Manley, J. L.** (1995). "The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities." *EMBO Journal* **14**(14): 3540-3551.
- Takagi, N. and Sasaki, M.** (1975). "Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse." *Nature* **256**(5519): 640-642.
- Tarn, W. Y. and Steitz, J. A.** (1995). "Modulation of 5' splice site choice in pre-messenger RNA by two distinct steps." *Proceedings of the National Academy of Sciences of the United States of America* **92**(7): 2504-2508.
- Tarn, W. Y. and Steitz, J. A.** (1997). "Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge." *Trends in biochemical sciences* **22**(4): 132-137.
- Taylor, S. J., Resnick, R. J. and Shalloway, D.** (2004). "Sam68 exerts separable effects on cell cycle progression and apoptosis." *BMC Cell Biol* **5**: 5.
- Tian, D., Sun, S. and Lee, J. T.** (2010). "The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation." *Cell* **143**(3): 390-403.
- Tillmar, L., Carlsson, C. and Welsh, N.** (2002). "Control of insulin mRNA stability in rat pancreatic islets. Regulatory role of a 3'-untranslated region pyrimidine-rich sequence." *J Biol Chem* **277**(2): 1099-1106.
- Tsai, C.-L., Rowntree, R. K., Cohen, D. E. and Lee, J. T.** (2008). "Higher order chromatin structure at the X-inactivation center via looping DNA." *Developmental biology* **319**(2): 416-425.
- Tsai, C. J., Mets, D. G., Albrecht, M. R., Nix, P., Chan, A. and Meyer, B. J.** (2008). "Meiotic crossover number and distribution are regulated by a dosage compensation protein that resembles a condensin subunit." *Genes & development* **22**(2): 194-211.
- Tse, C., Sera, T., Wolffe, A. P. and Hansen, J. C.** (1998). "Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III." *Mol Cell Biol* **18**(8): 4629-4638.
- Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., . . . Fuks, F.** (2006). "The Polycomb group protein EZH2 directly controls DNA methylation." *Nature* **439**(7078): 871-874.
- Wade, P. A., Geronne, A., Jones, P. L., Ballestar, E., Aubry, F. and Wolffe, A. P.** (1999). "Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation." *Nat Genet* **23**(1): 62-66.
- Wagner, E. J.** (2001). "Polypyrimidine Tract Binding Protein Antagonizes Exon Definition." *Society* **21**(10): 3281-3288.
- Wagner, E. J. and Garcia-Blanco, M. A.** (2001). "Polypyrimidine tract binding protein antagonizes exon definition." *Mol Cell Biol* **21**(10): 3281-3288.
- Wahl, M. C., Will, C. L. and Luhrmann, R.** (2009). "The spliceosome: design principles of a dynamic RNP machine." *Cell* **136**(4): 701-718.
- Wakefield, M. J., Keohane, A. M., Turner, B. M. and Graves, J. A.** (1997). "Histone underacetylation is an ancient component of mammalian X chromosome inactivation."

*Proceedings of the National Academy of Sciences of the United States of America* **94**(18): 9665-9668.

**Wakefield, R. I., Smith, B. O., Nan, X., Free, A., Soteriou, A., Uhrin, D., . . . Barlow, P. N.** (1999). "The solution structure of the domain from MeCP2 that binds to methylated DNA." *J Mol Biol* **291**(5): 1055-1065.

**Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S. and Zhang, Y.** (2004). "Role of histone H2A ubiquitination in Polycomb silencing." *Nature* **431**(7010): 873-878.

**Wang, J. and Manley, J. L.** (1995). "Overexpression of the SR proteins ASF/SF2 and SC35 influences alternative splicing in vivo in diverse ways." *RNA* **1**(3): 335-346.

**Warshawsky, D., Stavropoulos, N. and Lee, J. T.** (1999). "Further examination of the Xist promoter-switch hypothesis in X inactivation: evidence against the existence and function of a P(0) promoter." *Proc Natl Acad Sci U S A* **96**(25): 14424-14429.

**Weinberg, M. S., Villeneuve, L. M., Ehsani, A., Amarzguioui, M., Aagaard, L., Chen, Z. X., . . . Morris, K. V.** (2006). "The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells." *Rna* **12**(2): 256-262.

**West, M. H. and Bonner, W. M.** (1980). "Histone 2B can be modified by the attachment of ubiquitin." *Nucleic Acids Res* **8**(20): 4671-4680.

**Will, C. L. and Luhrmann, R.** (2001). "Spliceosomal UsnRNP biogenesis, structure and function." *Current Opinion in Cell Biology* **13**(3): 290-301.

**Wilson, B. G. and Roberts, C. W.** (2011). "SWI/SNF nucleosome remodellers and cancer." *Nat Rev Cancer* **11**(7): 481-492.

**Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. A. and Smith, C. W.** (2004). "Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay." *Molecular Cell* **13**(1): 91-100.

**Wu, J. Y. and Maniatis, T.** (1993). "Specific interactions between proteins implicated in splice site selection and regulated alternative splicing." *Cell* **75**(6): 1061-1070.

**Wu, S. C. and Zhang, Y.** (2010). "Active DNA demethylation: many roads lead to Rome." *Nature reviews. Molecular cell biology* **11**(9): 607-620.

**Wutz, A. and Jaenisch, R.** (2000). "A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation." *Molecular cell* **5**(4): 695-705.

**Wutz, A., Rasmussen, T. P. and Jaenisch, R.** (2002). "Chromosomal silencing and localization are mediated by different domains of Xist RNA." *Nature genetics* **30**(2): 167-174.

**Xiao, S. H. and Manley, J. L.** (1997). "Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing." *Genes & Development* **11**(3): 334-344.

**Xiao, T., Kao, C. F., Krogan, N. J., Sun, Z. W., Greenblatt, J. F., Osley, M. A. and Strahl, B. D.** (2005). "Histone H2B ubiquitylation is associated with elongating RNA polymerase II." *Mol Cell Biol* **25**(2): 637-651.

**Xu, N., Donohoe, M. E., Silva, S. S. and Lee, J. T.** (2007). "Evidence that homologous X-chromosome pairing requires transcription and Ctfc protein." *Nature genetics* **39**(11): 1390-1396.

- Xu, N., Tsai, C.-L. and Lee, J. T.** (2006). "Transient homologous chromosome pairing marks the onset of X inactivation." *Science (New York, N.Y.)* **311**(5764): 1149-1152.
- Yamazaki, S., Ishii, A., Kanoh, Y., Oda, M., Nishito, Y. and Masai, H.** (2012). "Rif1 regulates the replication timing domains on the human genome." *EMBO J.*
- Yang, L., Mei, Q., Zielinska-Kwiatkowska, A., Matsui, Y., Blackburn, M. L., Benedetti, D., . . . Chansky, H. A.** (2003). "An ERG (ets-related gene)-associated histone methyltransferase interacts with histone deacetylases 1/2 and transcription co-repressors mSin3A/B." *Biochem J* **369**(Pt 3): 651-657.
- Yen, Z. C., Meyer, I. M., Karalic, S. and Brown, C. J.** (2007). "A cross-species comparison of X-chromosome inactivation in Eutheria." *Genomics* **90**(4): 453-463.
- Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A. and Cui, H.** (2008). "Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA." *Nature* **451**(7175): 202-206.
- Zahler, A. M., Lane, W. S., Stolk, J. A. and Roth, M. B.** (1992). "SR proteins: a conserved family of pre-mRNA splicing factors." *Genes & Development* **6**(5): 837-847.
- Zhang, L.-F., Huynh, K. D. and Lee, J. T.** (2007). "Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing." *Cell* **129**(4): 693-706.
- Zhang, Y.** (2003). "Transcriptional regulation by histone ubiquitination and deubiquitination." *Genes Dev* **17**(22): 2733-2740.
- Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P. and Reinberg, D.** (1997). "Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex." *Cell* **89**(3): 357-364.
- Zhang, Y. and Reinberg, D.** (2001). "Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails." *Genes & development* **15**(18): 2343-2360.
- Zhao, J., Sun, B. K., Erwin, J. a., Song, J.-J. and Lee, J. T.** (2008). "Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome." *Science (New York, N.Y.)* **322**(5902): 750-756.
- Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K. A., Lozach, J., . . . Rosenfeld, M. G.** (2008). "Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation." *Mol Cell* **29**(1): 69-80.
- Zhu, J. and Krainer, A. R.** (2000). "Pre-mRNA splicing in the absence of an SR protein RS domain." *Genes & Development* **14**(24): 3166-3178.
- Zuo, P. and Manley, J. L.** (1993). "Functional domains of the human splicing factor ASF/SF2." *EMBO Journal* **12**(12): 4727-4737.



## **Analyse structurale et fonctionnelle de la région des A-repeats de l'ARN Xist impliqué dans l'inactivation du chromosome X chez les mammifères femelles**

L'inactivation du chromosome X correspond au silence transcriptionnel de l'un des deux chromosomes X dans les cellules des mammifères femelles. Il s'agit d'un mécanisme de compensation du dosage du chromosome X qui assure un taux d'expression des gènes liés aux chromosomes X équivalent entre organismes mâles (XY) et femelles (XX). Elle débute par une accumulation de l'ARN Xist (X inactive specific transcript) sur le chromosome X qui sera inactivé (Xi). Elle est suivie très rapidement par des modifications des histones qui assurent l'établissement, le maintien et la transmission de l'état transcriptionnel inactif de la chromatine. L'ARN Xist comprend plusieurs régions d'éléments répétés et notamment la région des A-repeats, essentielle pour la mise en place de l'inactivation. Mes recherches se sont portées sur l'étude de cette région singulière : sa structure et ses interactions protéiques. La technique de FRET (Fluorescence Resonance Energy Transfer) appliquée à l'ARN nous a permis de confirmer la structure de cette région parmi 3 modèles possibles. Elle se structure en deux tiges-boucles formée par l'appariement 2 à 2 de 4 répétitions successives. Dans une seconde partie, j'ai caractérisé l'interaction de cette région avec certains de ses partenaires protéiques *in vitro*. La région des A-repeats interagit notamment de manière directe avec les protéines PTB, KSRP et ASF/SF2. Les 2 premières protéines pourraient avoir un rôle dans la stabilité de l'ARN tandis qu'ASF/SF2 serait impliquée dans la maturation de l'ARN Xist.

**Mots-clés** : inactivation du X, ARN Xist, A-repeats, FRET, PTB, KSRP, ASF/SF2, épissage.

---

## **Structural and functional analysis of the A region of the Xist RNA involved in the X-chromosome inactivation in mammals female cells**

X-chromosome inactivation is the transcriptional silencing of one of the two X chromosomes in female mammal cells. This mechanism of dosage compensation ensures an equal level of the X-linked genes expression between males (XY) and females (XX). It initiates with the accumulation of the Xist RNA (X inactive specific transcript) on the future inactive X chromosome (Xi). It is followed by the apposition of epigenetic marks such as histone modifications, that ensure establishment, maintenance and transmission of the inactive state of the chromatin. Xist RNA comprises a number of repeated regions and, in particular to its 5' end the A region, absolutely necessary for the establishment of the X-inactivation. My research was focused on the study of this singular region: its structure and its protein interactions. The FRET method (Fluorescence Resonance Energy Transfer) applied to RNA allowed us to ascertain that the RNA is structured in two long stem-loop structures each including four repeats. In a second part, I characterized the *in vitro* interaction of this region with some of its protein partners. The A region interacts directly with PTB, KSRP and ASF/SF2 proteins. The first two proteins may have a role in RNA stability whereas ASF/SF2 could be involved in the splicing process.

**Keywords** : X inactivation, Xist RNA, A region, FRET, PTB, KSRP, ASF/SF2, splicing.