



**HAL**  
open science

**Détermination de la structure secondaire d'une région de l'ARN Xist nécessaire à l'inactivation du chromosome X, la région des A-repeats, et identification de ses partenaires protéiques ayant un rôle structural ou fonctionnel dans l'inactivation**

Sylvain Maenner

► **To cite this version:**

Sylvain Maenner. Détermination de la structure secondaire d'une région de l'ARN Xist nécessaire à l'inactivation du chromosome X, la région des A-repeats, et identification de ses partenaires protéiques ayant un rôle structural ou fonctionnel dans l'inactivation. Médecine humaine et pathologie. Université Henri Poincaré - Nancy 1, 2009. Français. NNT : 2009NAN10092 . tel-01775836

**HAL Id: tel-01775836**

**<https://hal.univ-lorraine.fr/tel-01775836>**

Submitted on 24 Apr 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>

**FACULTE DES SCIENCES & TECHNIQUES**  
**U.F.R Sciences & Techniques Biologiques**

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

Thèse

présentée et soutenue publiquement pour l'obtention du titre de

**DOCTEUR DE L'UNIVERSITE HENRI POINCARÉ**

Mention : ARN, Biogenèse et Structure / Fonction

par **Sylvain MAENNER**

**Détermination de la structure secondaire d'une  
région de l'ARN Xist nécessaire à l'inactivation du  
chromosome X, la région des A-repeats,  
et identification de ses partenaires protéiques ayant  
un rôle structural ou fonctionnel dans l'inactivation**

**Soutenance publique prévue le 10 novembre 2009**

**Membres du jury :**

Rapporteurs :	Mme Luisa Dandolo	Directeur de recherche CNRS, Paris
	Mme Pascale Romby	Directeur de recherche CNRS, Strasbourg
Examineurs :	M. Philip Avner	Directeur de recherche PASTEUR, Paris
	M. Athanase Visvikis	Professeur UHP, Nancy I, directeur de thèse
	Mme Christiane Branlant	Directeur de recherche CNRS, Nancy I, co-directeur de thèse









Je commence tout d'abord par adresser mes sincères remerciements à Madame Christiane Branlant, Directeur de l'UMR CNRS-UHP 7214 pour m'avoir accueilli dans son équipe, pour m'avoir fait confiance à l'issue de mon DEA et pour m'avoir permis de réaliser ce travail. Je lui suis particulièrement reconnaissant pour les nombreuses discussions scientifiques que nous avons eues et les conseils qu'elle m'a prodigué tout au long de cette thèse et la rigueur qu'elle m'a transmise. Je la remercie aussi pour son aide très importante dans la rédaction de ce mémoire, ceci toujours sans compter les heures.

Je remercie également Athanase Visvikis, mon directeur de thèse, pour m'avoir fait partager son expérience concernant la culture de cellules et le génie génétique mais aussi pour sa capacité à communiquer son optimisme et ceci même dans les situations les plus compliquées.

J'exprime toute ma reconnaissance à Luisa Dandolo et Pascale Romby pour avoir accepté d'évaluer le travail que j'ai effectué au cours de ma thèse. Je remercie aussi très vivement Philip Avner qui a accepté de juger mon travail et pour m'avoir accueilli de nombreuses fois dans son laboratoire et également pour les discussions enrichissantes que nous avons eues lors de ces séjours.

Merci à Virginie pour m'avoir initié aux techniques d'analyses des structures secondaires d'ARN en début de thèse et pour m'avoir initié à la recherche au cours de mes premiers pas dans le domaine lors de mon stage de maîtrise. Ton enthousiasme m'a convaincu et je t'en remercie. Un grand merci à Magali Blaud pour m'avoir initié aux techniques de FRET et de dichroïsme circulaire sans lesquelles mon travail n'aurait pas été aussi convaincant. J'exprime un grand merci et toute ma reconnaissance à Laetitia et Sarah qui ont effectué un travail de titan pour le projet que nous avons mené en commun. J'adresse un grand merci à P Clerc qui m'a initié aux techniques d'imagerie et pour les conversations que j'ai eu l'occasion d'entretenir avec lui. Notre collaboration a été pour moi d'une grande importance. Merci également à Véro pour ses conseils avisés concernant les expériences d'immuno-précipitation.

Je remercie tous les membres du laboratoire AREMS qui ont participé de près ou de loin à l'élaboration de ce travail.

Un énorme MERCI à Florence et à la p'tite Nath, mes chères co-locatrices du Box 4 qui m'ont supporté toutes ces années de dur labeur. Merci pour les fou-rires, les blagues potaches et les adages pourris. Une page se tourne mais vous resterez dans mes pensées.

Un autre énorme MERCI aux résidentes du Box5. Merci à vous, Christelle, Gaby et Mag (résidente matinale du Box5) pour vos oreilles attentives et pour tous les bons moments passés ensemble. L'instant café me manquera à n'en pas douter et vive la verveine.

Un petit clin d'œil à Ben, avec qui j'ai eu quelques conversations existentielles tout à fait agréables....

Je tiens également à exprimer toute ma gratitude à l'équipe de P Avner en l'occurrence Andrew, Agnès, Céline, Corinne, Chenxia, Claire, Edith, Michael, Pablo, Philippe, Séb, Rachel et Ute. Merci à tous pour votre gentillesse, votre disponibilité sans faille et votre bonne humeur. A chacun de mes séjours parisiens, le sens du verbe accueillir a pris toute sa dimension.

Je veux également adresser un énorme merci à toute ma famille en particulier à mes parents qui m'ont soutenu dans les moments les plus pénibles (car il ne faut pas se le cacher, il y en a.....).

Et pour finir, je veux remercier ma petite Marlène.... Ton soutien et tes relectures m'ont été plus que salutaires, tu m'as été essentielle....MERCI

***Liste des Abréviations***

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  
ARNr : ARN ribosomique  
ARNt : ARN de transfert  
ATP : adénosine tri-phosphate  
ATRX : alpha thalassemia/mental retardation X-linked  
BAZ-1 : BWSCR2-associated zinc finger protein 1  
BF : blocking factor  
Bmi-1 : B cell-specific Mo-MLV integration site 1  
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  
Cbx2 : chromobox protein homolog 2  
CE : counting element  
ChIP : immunoprécipitation de chromatine  
CIP : calf intestine phosphatase  
cM : centi-morgan  
CMCT : 1-cyclohexyl-3-(2-morpholinoéthyl) carbodiimide méthap-toluène  
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  
dpy : dumpy  
DRM : région méthylée de façon différentielle  
DTT : dithiothréitol  
DXPas34 : mini-satellite constitué d'un motif de 34 paires de bases répété,  
EDTA : éthylène diamine tétraacétate  
Eed : protéine de développement de l'ectoderme embryonnaire  
EPI : épiblaste  
ES : cellules embryonnaires souches  
ESE : exonic splicing enhancer

ESS : exonic splicing silencer  
FBP : fuse binding protein  
Fig. : figure  
FISH : fluorescence in-situ hybridization  
FLP : flip-flop  
FRET : Förster resonance energy transfer  
FRT : FLP recognition target  
GFP : green fluorescent protein  
GGT : gamma-glutamyltranspeptidase 1  
H3K14 : lysine 14 de l'histone H3  
H3K27 : lysine 27 de l'histone H3  
H3K4 : lysine 4 de l'histone H3  
H3K9 : lysine 9 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  
HEK : human embryonic kidney  
HeLa : Henrietta Lacks  
hES : cellules embryonnaires humaine  
HMT : histone méthyl transférase  
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 : récepteur de Igf2  
IRES : Internal Ribosome Entry Site  
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  
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  
miRNA : micro ARN  
mix : mitosis and X associated  
MLE : maleless  
MOF : males absent on the first  
MSL : male specific lethal  
MT : marqueur de taille  
NCL : nucléoline  
NHP2 : high mobility group-like nuclear protein 2  
NHR : non histone region  
nm : nanomètre  
NMD : nonsense mediated mRNA decay  
Nop56 : nucleolar protein 56  
nt : nucléotides  
NURF : facteur de remodelage des nucléosomes  
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  
PE : endoderme primitif  
Peg3 : paternally-expressed gene 3 protein  
PGC : primordial germ cells  
Ph : polyméiotique  
pH : potentiel hydrogène  
PHO : protéine Pléiohoméotique  
PNK : polynucléotide kinase  
PRC1 : complexe polycomb 1  
PRC2 : complexe polycomb 2  
PRE : polycomb response element  
PRMT : enzyme de méthylation des arginines  
PSC : posterior Sex Combs  
PSF : splicing factor, proline- and glutamine-rich  
PTB : polypyrimidine tract-binding protein  
PTK : puromycine-thymidine kinase  
RbAp46-48 : retinoblastoma-binding protein p46-p48  
RGG : boîte riche en arginines et glycines  
RHA : RNA helicase A  
RIF1 : rap1-interacting factor 1 homolog  
Ring : ring-finger protein  
RIP : RNA immuno-precipitation  
RISC : RNA induced silencing complex  
RITS : RNA induced transcriptional silencing



RMCE : échange de cassette médiée par recombinaison  
RMN : résonance magnétique nucléaire  
RNase : ribonucléase  
RNasin : inhibiteur de ribonucléases  
RNP : riboNucléoParticule/complexe ribonucléoprotéique  
Rod1 : resistance to o-dinitrobenzene protein 1  
roX : RNA on the X chromosome  
RRM : RNA recognition motif  
RT : transcription inverse  
SAF-A : scaffold attachment factor A  
Sam68 : src-associated in mitosis 68 kDa protein  
SAP18 : sin3-associated polypeptide 18 kDa  
SATB : special AT-rich sequence-binding protein  
SD : simplet-doublet  
sdc : sex determination and dosage compensation defect  
SDS : sodium dodécylsulfate  
SET : domaine catalytique de méthylation des histones retrouvé dans les protéines  
SIAR : SD signals that are independant of asynchronous DNA replication  
Sir : NAD-dependent deacetylase sirtuin  
siRNA : short 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)  
SPOP : speckle-type POZ protein  
SuVar : enzyme de méthylation de H3K9  
SWI/SNF : mating type Switching-Sucose Non-Fertile  
SXL : protein sex-lethal  
TBE : tampon Tris-borate-EDTA  
TE : trophoctoderme  
TIAR : TIA-1-related protein  
TIF1b : transcription intermediary factor 1-beta  
Tris : tris (hydroxyméthyl) aminométhane  
TS : cellules souches du trophoctoderme  
TSA : trichostatine A  
Tsix : antisens à Xist  
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  
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), ni (b) **les abréviations internationales des unités de mesure** (ml, mg, s, h, etc.), ni (c) **celles de certains composants de la substance vivante** (ATP, RNA, DNA, mRNA, etc.) pour lesquels l'abréviation internationalement acceptée a été gardée.

**Listes des figures****INTRODUCTION**

<b>Figure 1</b> : Représentation d'un nucléosome	p 4
<b>Figure 2</b> : Organisation de l'ADN dans la structure chromatinienne	p 5
<b>Figure 3</b> : Les différents types de modifications des histones H3 et H4	p 6
<b>Figure 4</b> : Représentation schématique du rôle de chaque Dnmt, maintenance <i>versus</i> méthylation <i>de novo</i>	p 11
<b>Figure 5</b> : Modèle du réseau inter-chromosomique	p 15
<b>Figure 6</b> : Modèle de régulation du cluster génique <i>Igf2r</i> soumis à empreinte parentale et régulé par l'ARN non-codant Air	p 19
<b>Figure 7</b> : ARN <i>Kncq1ot1</i> et compaction de la chromatine du cluster <i>Kcnq1</i> par les protéines PcG	p 22
<b>Figure 8</b> : Formation d'hétérochromatine péri-centromérique médiée par les petits ARN chez la levure <i>S. pombe</i>	p 24
<b>Figure 9</b> : Schéma simplifié des relations entre évolution du déterminisme sexuel et compensation du dosage des gènes	p 27
<b>Figure 10</b> : Les diverses stratégies de compensation du dosage des gènes	p 28
<b>Figure 11</b> : Compensation de dosage des gènes chez la drosophile	p 30
<b>Figure 12</b> : Modèle proposé pour la compensation du dosage des gènes chez <i>C. elegans</i>	p 31
<b>Figure 13</b> : Illustration phénotypique de l'inactivation du chromosome X	p 32
<b>Figure 14</b> : Morphologie du chromosome X inactif	p 33
<b>Figure 15</b> : Evolution de la quantité et de la localisation de l'ARN <i>Xist</i> au cours de la différenciation des cellules ES	p 35
<b>Figure 16</b> : Dynamique de l'expression du gène <i>Xist</i> au cours du développement	p 38
<b>Figure 17</b> : Modèle de régulation de l'expression des gènes <i>Hox</i> par les ARN non-codants (HOTAIR) <i>via</i> les enzymes de modifications des histones	p 42
<b>Figure 18</b> : Interactions des protéines des complexes PRC1 et PRC2 avec les ARN	p 44
<b>Figure 19</b> : Composition en exons des ARNm des variants de macroH2A et domaines protéiques correspondants	p 45
<b>Figure 20</b> : Foyers de concentration de la protéine macroH2A et co-localisation avec le chromosome X inactif	p 46
<b>Figure 21</b> : Les différents types de modifications présentes sur les chromosomes X	p 52
<b>Figure 22</b> : Centre d'inactivation du chromosome X chez la souris	p 56
<b>Figure 23</b> : Modèle mono-factoriel proposé pour les étapes de comptage et de sélection du chromosome X à inactiver	p 58
<b>Figure 24</b> : Modèle à deux facteurs proposé pour les étapes de comptage et de sélection du chromosome X à inactiver	p 59
<b>Figure 25</b> : Modèle de choix aléatoire basé sur un passage continu entre 2 états des 2 chromosomes X	p 61
<b>Figure 26</b> : Modèle de choix et de comptage par l'appariement des domaines chromosomiques	p 62
<b>Figure 27</b> : Fenêtre d'opportunité de l'initiation de l'inactivation du chromosome X	p 64
<b>Figure 28</b> : Modèle proposé pour la formation du domaine <i>Xist</i>	p 67
<b>Figure 29</b> : Cinétique des événements de mise en place de l'inactivation aléatoire lors de la différenciation des cellules ES de souris	p 68
<b>Figure 30</b> : Conservation de l'architecture et des séquences répétées du gène <i>Xist</i> chez les euthériens	p 71
<b>Figure 31</b> : Eléments fonctionnels impliqués dans la localisation et dans la répression transcriptionnelle du Xi médiée par l'ARN <i>Xist</i>	p 72
<b>Figure 32</b> : Interaction dynamique entre <i>Xist</i> et <i>Tsix</i> durant la différenciation des cellules ES	p 74
<b>Figure 33</b> : Régulation de l'expression du gène <i>Xist</i>	p 78

<b>Figure 34</b> : Intervention de la région des A-repeats dans la relocalisation des gènes à inactiver dans le domaine Xist	p 81
<b>Figure 35</b> : Analyse de la région des A-repeats	p 83
<b>Figure 36</b> : Modèle de dimérisation des éléments répétés de la région des A-repeats proposé par Duszczuk <i>et al.</i> (2008)	p 84
 <b>RESULTATS ET DISCUSSION</b>	
<b>Figure 37</b> : Représentation schématique des positions modifiées au sein des ARN par l'action des sondes chimiques et des liaisons phosphodiester coupées par l'action des sondes enzymatiques	p 112
<b>Figure 38</b> : Principe de cartographie chimique et enzymatique des ARN et de l'analyse par extension d'amorce	p 113
<b>Figure 39</b> : Séquences des régions des A-repeats des ARN Xist murin et humain	p 115
<b>Figure 40</b> : Schéma représentant pour les régions des A-repeats murine et humaine, la séquence, le numéro, la position d'hybridation de chaque oligonucléotide amorce, le segment analysé est aussi indiqué	p 116
<b>Figure 41</b> : Schéma représentant les ARN utilisés pour l'analyse de la structure 2D de la région des A-repeats murine par emploi de sondes enzymatiques et chimiques	p 118
<b>Figure 42</b> : Confrontation de nos données expérimentales sur la structure 2D de la région des A-repeats au modèle proposé par Wutz <i>et al.</i> , (2002) pour un des éléments répétés	p 119
<b>Figure 43</b> : Représentation schématique des données expérimentales obtenues pour la région des A-repeats de l'ARN Xist murin (ARN-AM) sur le modèle de structure secondaire 3 qui semblait être le plus vraisemblable	p 122
<b>Figure 44</b> : Schéma représentant l'ARN-AH utilisé pour l'analyse de la structure 2D de la région des A-repeats humaine	p 123
<b>Figure 45</b> : Représentation schématique des données expérimentales obtenues pour la région des A-repeats de l'ARN Xist humain (ARN-AH) sur le modèle de structure secondaire 3 qui semblait le plus vraisemblable	p 124
<b>Figure 46</b> : Schéma représentant l'ARN-SLS1H utilisé pour l'analyse de la structure 2D de la région des A-repeats humaine	p 125
<b>Figure 47</b> : Principe du FRET	p 127
<b>Figure 48</b> : Structures chimiques des fluorophores Cyanine3 (Cy3) et Cyanine5 (Cy5) utilisés dans cette étude	p 128
<b>Figure 49</b> : Stratégie utilisée pour évaluer la validité de 3 modèles possibles de la région des A-repeats murine par l'approche de FRET	p 130
<b>Figure 50</b> : Protocole utilisé pour les expériences de FRET	p 134
<b>Figure 51</b> : Analyse par retard sur gel des complexes formés après incubation de l'ARN-AM en présence d'extrait nucléaire de cellules HeLa	p 136
<b>Figure 52</b> : Analyse de la protection contre l'action des RNases générée sur l'ARN-AM par fixation des protéines d'un extrait nucléaire de cellules HeLa	p 137
<b>Figure 53</b> : Représentation des protections contre l'hydrolyse par les RNases T1 et T2 générées par fixation des protéines d'un extrait nucléaire de cellules HeLa sur la région des A-repeats	p 139
<b>Figure 54</b> : Principe de la purification des complexes ribonucléoprotéiques par chromatographie d'affinité basée sur l'emploi de la protéine de l'enveloppe du phage MS2 et de la protéine MBP	p 143
<b>Figure 55</b> : Analyse par retard sur gel des complexes formés avec l'ARN-AH-MS2 en présence de concentrations croissantes de protéines MS2/MBP	p 144
<b>Figure 56</b> : Constructions réalisées pour les expériences de purification par chromatographie d'affinité MS2/MBP des RNP formées par la région des A-repeats humaine entière ou ses sous-fragments lorsqu'ils sont incubés en extrait nucléaire de cellules HeLa	p 146

<b>Figure 57</b> : Analyse par électrophorèse en SDS-PAGE des protéines contenues dans les complexes ribonucléoprotéiques formés avec les ARN-MS2 et AH-MS2 purifiés par chromatographie d'affinité reposant sur l'emploi de la protéine de fusion MS2/MBP	<b>p 148</b>
<b>Figure 58</b> : Analyse des protéines contenues dans les complexes ribonucléoprotéiques formés avec différents fragments de la région des A-repeats humaine et un extrait nucléaire de cellules HeLa et purifiés par chromatographie d'affinité MS2/MBP	<b>p 150</b>
<b>Figure 59</b> : Analyse des protéines contenues dans les complexes ribonucléoprotéiques purifiés par chromatographie d'affinité MBP/MS2 formés dans un extrait nucléaire de cellules ES non différenciées avec différents fragments de la région des A-repeats de l'ARN Xist murin	<b>p 153</b>
<b>Figure 60</b> : Analyse par SDS-PAGE des protéines associées à la région des A-repeats de l'ARN Xist murin dans les complexes formés dans des extraits nucléaires préparés à partir de cellules ES de souris femelles récoltées à différents stades de différenciation	<b>p 213</b>
<b>Figure 61</b> : Constructions réalisées pour les expériences de purification par chromatographie d'affinité MS2/MBP des RNP formées avec une région de l'ARN du VIH (positions 5538 à 5514) et la région 5' terminale de l'ARN Xist murin et après incubation en extrait nucléaire de cellules ES non différenciées	<b>p 216</b>
<b>Figure 62</b> : Fractionnement par SDS-PAGE des protéines contenues dans les RNP formées avec la région des A-repeats murine entière et avec la région 5' terminale de l'ARN Xist murin dans des extraits nucléaires de cellules ES non différenciées de souris femelles	<b>p 217</b>
<b>Figure 63</b> : Modèle de structure secondaire de l'ARN-5'M établi à partir d'une prédiction informatique par le logiciel M-Fold	<b>p 248</b>
<b>Figure 64</b> : Modèle d'extinction transcriptionnelle médiée par des ARN édités	<b>p 253</b>
<b>Figure 65</b> : Analyse par Western-blot de protéines détectées dans les RNP formées avec la région des A-repeats de l'ARN Xist murin en extraits nucléaires de cellules ES après purification par chromatographie d'affinité	<b>p 259</b>
<b>Figure 66</b> : Analyse par Western-blot de protéines contenues dans les RNP formées par la région des A-repeats, ses sous-fragments ou l'extrémité 5' terminale de l'ARN Xist murin dans un extrait nucléaire de cellules ES non différenciées, complexes qui ont été purifiés par chromatographie d'affinité MS2/MBP	<b>p 261</b>
<b>Figure 67</b> : Etude de localisation de 8 protéines présentes dans les RNP au sein de cellules ES de souris 4 jours après l'initiation de la différenciation	<b>p 263</b>
<b>Figure 68</b> : Modèle de structure de la protéine PTB et interaction avec les ARN	<b>p 264</b>
<b>Figure 69</b> : Structure de la protéine KSRP	<b>p 265</b>
<b>Figure 70</b> : Purification des RNP formées en extrait nucléaire de cellules HeLa avec l'ARN de la région des A-repeats de l'ARN Xist humain ou du snoRNA U3, contenant la protéine PTB	<b>p 266</b>
<b>Figure 71</b> : Mise en évidence par pontage aux UV d'une interaction directe entre PTB et la région des A-repeats au sein des RNP formées par la région des A-repeats humaine en extrait nucléaire de cellules HeLa	<b>p 268</b>
<b>Figure 72</b> : Production et purification de la protéine PTB	<b>p 269</b>
<b>Figure 73</b> : Spectres de dichroïsme circulaire attendus entre les longueurs d'onde de 180 à 240 nm pour des peptides renfermant soit des hélices $\alpha$ , soit des feuilletts $\beta$ , soit n'ayant pas de structure 2D caractéristique	<b>p 270</b>
<b>Figure 74</b> : Spectre de dichroïsme circulaire de la protéine PTB enregistré entre 200 et 320 nm	<b>p 272</b>
<b>Figure 75</b> : Analyse par retard sur gel des complexes formés par incubation de la région des A-repeats humaine et murine entière ou la grande structure tige-boucle SLS1M de l'ARN de souris en présence de la protéine PTB	<b>p 273</b>
<b>Figure 76</b> : Exemple d'analyse des protections contre l'action des RNases générée par fixation de la protéine PTB sur la région des positions 670 à 740 de l'ARN-AM	<b>p 275</b>
<b>Figure 77</b> : Représentation des protections contre l'hydrolyse par les RNases T1 et T2 générées par fixation de la protéine PTB sur la région des A-repeats murine	<b>p 277</b>
<b>Figure 78</b> : Production, purification et spectre dichroïque de la protéine KSRP-4KH	<b>p 278</b>

<b>Figure 79</b> : Analyse par retard sur gel des complexes formés par incubation de la région des A-repeats humaine ou murine entière ou de la grande structure tige-boucle SLS1M de l'ARN de souris en présence de la protéine KSRP-4KH	p 279
<b>Figure 80</b> : Représentation schématique des régions de la structure tige-boucle SLS1M protégées par la protéine KSRP-4KH	p 280
<b>Figure 81</b> : Représentation des zones spectrales des protéines et des ARN	p 281
<b>Figure 82</b> : Schéma représentant les différentes étapes de la mesure des spectres de dichroïsme circulaire visant à tester l'effet de l'interaction entre l'ARN-SLS1M et la protéine testée sur la structure 2D de l'ARN	p 282
<b>Figure 83</b> : Spectre de dichroïsme de l'ARN-SLS1M	p 283
<b>Figure 84</b> : Etude des complexes [SLS1M/PTB] et [SLS1M/KSRP-4KH] par dichroïsme circulaire	p 284
<b>Figure 85</b> : Influence de la présence des protéines PTB et KSRP-4KH sur la stabilité thermique de la structure 2D de l'ARN-SLS1M	p 286
<b>Figure 86</b> : Stratégie utilisée pour générer la lignée de cellules ES de souris femelles présentant une délétion de 4,2 kb au locus Xist	p 291
<b>Figure 87</b> : La réaction d'épissage des introns des ARN pré-messagers nucléaires	p 294
<b>Figure 88</b> : Stratégie utilisée pour l'insertion de sites d'épissage encadrant la région des A-repeats	p 296
<b>Figure 89</b> : Les séquences choisies pour être converties en sites 5' et 3' d'épissage conformes aux séquences consensus des sites d'épissage des ARN pré-messagers de vertébrés par mutagenèse d'un fragment du gène de l'ARN Xist	p 297
<b>Figure 90</b> : Génération du plasmide pUC-Xistm réalisé pour les expériences d'épissage <i>in vitro</i>	p 298
<b>Figure 91</b> : Représentation schématique de l'ARN Xist muté et des positions d'hybridation des oligonucléotides utilisés pour les expériences de RT-PCR	p 299
<b>Figure 92</b> : Analyse par RT-PCR du devenir des ARN-A2C sauvages et mutés après incubation dans un extrait nucléaire	p 300
<b>Figure 93</b> : Mise en évidence de l'utilisation de deux sites 5' d'épissage par séquençage	p 301
<b>Figure 94</b> : Schéma de la stratégie utilisée pour produire le plasmide pXistm	p 303
<b>Figure 95</b> : Analyse par RT-PCR des produits d'épissage <i>in cellulo</i> des fragments de l'ARN Xist sauvage et muté	p 304

## Listes des tableaux

<b>Tableau I</b> : Anticorps utilisés au cours de ce travail	p 92
<b>Tableau II</b> : Comparaison des résultats d'identification par spectrométrie de masse des protéines contenues dans les RNP formées avec la région des A-repeats humaine	p 219
<b>Tableau III</b> : Comparaison des résultats d'identification par spectrométrie de masse du contenu en protéines des RNP formées avec la région des A-repeats murine	p 227
<b>Tableau IV</b> : Comparaison des résultats d'identification par spectrométrie de masse du contenu en protéines des RNP formées avec la région des A-repeats murine, différents sous-fragments de cette région et la région 5' terminale de l'ARN Xist murin avec l'extrait nucléaire de cellules ES non différenciées	p 243
<b>Tableau V</b> : Liste des protéines dont il serait intéressant d'étudier l'effet de l'extinction par ARN interférence sur l'inactivation du chromosome X	p 257

**Listes des figures annexes**

<b>Figure 1 annexe :</b> Comparaison des accessibilités aux enzymes des ARN-AM et ARN-1137 lors de l'analyse de la structure 2D de la région des A-repeats murine	<b>p 201</b>
<b>Figure 2 annexe :</b> Influence de la concentration en Mg <sup>2+</sup> sur l'accessibilité de l'ARN-AM aux sondes enzymatiques	<b>p 204</b>
<b>Figure 3 annexe :</b> Représentation schématique des données expérimentales obtenues pour la région des A-repeats de l'ARN Xist humain (ARN-AH) sur le modèle de structure secondaire 2	<b>p 206</b>
<b>Figure 4 annexe :</b> Spectres d'émission de fluorescence enregistrés pour chacun des couples d'oligonucléotides utilisés lors des expériences de FRET qui ont permis de valider le modèle 3 de structure 2D de la région des A-repeats murine	<b>p 207</b>
<b>Figure 5 annexe :</b> Analyse de la protection contre l'action des RNases générée sur l'ARN-AM par fixation des protéines d'un extrait nucléaire de cellules HeLa	<b>p 209</b>
<b>Figure 6 annexe :</b> Représentation des protections contre l'hydrolyse par les RNases T1 et T2 générées par fixation des protéines d'un extrait nucléaire de cellules HeLa sur le modèle de structure 3 de la région des A-repeats	<b>p 211</b>

<b>Remerciements</b>	
<b>Abréviations</b>	
<b>Listes des Figures et Tableaux</b>	
<b>Sommaire</b>	

<b>Avant-propos</b>	<b>p 1</b>
<b>I. INTRODUCTION</b>	<b>p 2</b>
<b>A. Généralités</b>	<b>p 2</b>
<b>1. Chromatine et épigénétique</b>	<b>p 2</b>
1.1 Les racines de l'épigénétique : Histoire et Evolution	p 2
1.2 Complexité de la chromatine et éléments régulateurs	p 3
1.2.1 Nucléosome et modifications covalentes des histones	p 3
1.2.2 Régulation en <i>cis</i> : physico-chimie de modifications	p 7
1.2.3 Régulation en <i>trans</i> : le code histone	p 7
• La diversité des modifications des histones	p 8
• La combinatoire des modifications des histones	p 8
• Lecteurs et effecteurs des modifications des histones	p 8
1.2.4 Méthylation de l'ADN	p 10
1.2.5 Compartiments nucléaires et répression génique	p 14
1.3. ARN non-codant et régulations épigénétiques	p 16
1.3.1 ARN non-codant et complexité du vivant	p 16
1.3.2 ARN non-codant et expression des gènes soumis à l'empreinte	p 17
• ARN non-codant Air et Cluster Igf2r	p 18
• ARN non-codant et locus <i>Kcnq1</i>	p 19
1.3.3 ARN interférence et hétérochromatine constitutive	p 23
<b>2. Notion de compensation du dosage des gènes</b>	<b>p 25</b>
<b>3. Evolution du déterminisme sexuel et compensation du dosage des gènes</b>	<b>p 25</b>
<b>4. Compensation du dosage des gènes chez <i>Drosophila melanogaster</i> et <i>Caenorhabditis elegans</i></b>	<b>p 28</b>
4.1 Compensation du dosage des gènes chez <i>Drosophila Melanogaster</i>	p 29
4.2 Compensation du dosage des gènes chez <i>C. elegans</i>	p 31
<b>B. Inactivation du chromosome X chez les mammifères</b>	<b>p 32</b>
<b>1. Principes généraux</b>	<b>p 32</b>
<b>2. Inactivation du chromosome X au cours de l'embryogenèse</b>	<b>p 34</b>
2.1 Les étapes de l'embryogenèse chez la souris	p 34
2.2 L'ARN Xist	p 35
2.3 Dynamique d'expression du gène <i>Xist</i> : inactivation soumise à empreinte parentale et inactivation aléatoire	p 36
2.4 Les systèmes d'étude de l'inactivation du chromosome X	p 39
2.4.1 Cellules somatiques (humaines, murines et hybrides)	p 39
2.4.2 Embryon et cellules ES	p 39
<b>3. Les acteurs protéiques de l'inactivation du chromosome X</b>	<b>p 40</b>
3.1 Les protéines du groupe polycomb	p 40
3.1.1 PRC2	
3.1.2 PRC1	p 42
3.2 macroH2A	p 44
3.3 Les autres facteurs protéiques associés au Xi	p 47
3.3.1 SAF-A	p 47



3.3.2	ATRX	p 48
3.3.3	BRCA1	p 48
3.3.4	PARP-1	p 49
3.3.5	smcHD1	p 50
3.3.6	Ctcf	p 50
<b>4.</b>	<b>Les caractéristiques du Xi</b>	<b>p 51</b>
4.1	Les marques épigénétiques du Xi	p 51
4.2	Réplication tardive	p 52
4.3	Méthylation de l'ADN	p 53
4.4	Gènes échappant à l'inactivation	p 53

<b>C.</b>	<b>Mécanistique de l'inactivation aléatoire : caractérisation du centre d'inactivation du chromosome X : le XIC</b>	<b>p 54</b>
1.	<b>Le centre de l'inactivation XIC : un locus complexe</b>	<b>p 54</b>
2.	<b>Les étapes de l'inactivation aléatoire</b>	<b>p 56</b>
2.1	Etapes de comptage et choix	p 56
2.1.1	Modèle mono-factoriel proposé pour les étapes de choix et comptage du chromosome X à inactiver	p 57
2.1.2	Modèle à deux facteurs proposé pour les étapes de comptage et choix du chromosome X à inactiver	p 58
2.1.3	Modèle de choix aléatoire basé sur un passage continu entre 2 états des 2 chromosomes X	p 60
2.1.4	Rôle potentiel de l'appariement des 2 chromosomes X dans le processus d'inactivation	p 61
2.2.	L'initiation	p 63
2.2.1	Fenêtre d'opportunité de l'inactivation aléatoire	p 63
2.2.2	Propagation de l'ARN <i>Xist</i>	p 64
2.2.3	Le domaine <i>Xist</i>	p 66
2.3	La mise en place des marques épigénétiques et maintenance de l'état réprimé	p 67
<b>3.</b>	<b>Régulation par le gène <i>Xist</i></b>	<b>p 69</b>
3.1	Structure du gène <i>Xist</i>	p 69
3.2	Caractérisation fonctionnelle du gène <i>Xist</i>	p 71
<b>4.</b>	<b>Régulation de l'expression du gène <i>Xist</i></b>	<b>p 73</b>
4.1	Régulation de l'expression du gène <i>Xist</i> par <i>Tsix</i>	p 73
4.1.1	Structure du gène <i>Tsix</i>	p 73
	• Sites alternatifs de transcription	p 73
	• Les variants d'épissage	p 73
4.1.2	Caractérisation fonctionnelle de <i>Tsix</i>	p 73
4.1.3	Comment <i>Tsix</i> régule l'expression du gène <i>Xist</i>	p 74
4.1.4	Régulation de l'expression de <i>Tsix</i>	p 75
4.2	Répression de <i>Xist</i> et pluripotence	p 76
4.3	Méthylation du promoteur du gène <i>Xist</i>	p 87
4.4	Vision globale actuelle du mécanisme de répression de l'expression du gène <i>Xist</i>	p 77
<b>5.</b>	<b>La région des A-repeats : l'ultime acteur</b>	<b>p 78</b>
5.1	Caractérisation fonctionnelle de la région des A-repeats	p 78
5.2	Régulation de l'expression de la région des A-repeats	p 79
5.3	Connaissances actuelles sur le(s) rôle(s) de la région des A-repeats	p 79
5.3.1	Importance de la région des A-repeats sous sa forme ADN	p 79

5.3.2	Importance de la région des A-repeats sous sa forme ARN	p 80
5.4	Composition et structure secondaire de la région des A-repeats	p 82
<b>Objectifs</b>		<b>p 85</b>
<b>II. MATERIEL ET METHODES</b>		<b>p 87</b>
<b>A. Matériel</b>		<b>p 87</b>
<b>1. Souches et plasmides utilisés</b>		<b>p 87</b>
1.1	Les souches d' <i>E. Coli</i>	p 87
1.2	Les cellules d'organismes eucaryotes	p 87
1.3	Les plasmides utilisés	p 87
1.3.1	Plasmides recombinants pour la production <i>in vitro</i> d'ARN	p 87
1.3.2	Les vecteurs d'expression eucaryote	p 88
1.3.3	Les vecteurs utilisés pour la production des protéines recombinantes	p 89
<b>2. Milieux de culture, tampons et solution d'usage courant</b>		<b>p 91</b>
<b>3. Anticorps</b>		<b>p 92</b>
<b>B. Méthodes</b>		<b>p 93</b>
<b>1. Méthodes utilisées pour l'analyse de la structure secondaire d'un ARN</b>		<b>p 93</b>
1.1	Techniques de cartographie des ARN par utilisation de sondes enzymatiques et chimiques	p 93
1.1.1	Préparation des ARN	p 93
1.1.2	Analyse de la structure secondaire d'un ARN en solution par utilisation de sondes enzymatiques	p 93
1.1.3	Analyse de la structure secondaire d'un ARN en solution par utilisation de sondes chimiques	p 94
1.1.4	Analyse des nucléotides protégés contre l'action des RNases par les protéines d'un extrait nucléaire ou par des protéines recombinantes	p 95
1.1.5	Analyse par extension d'amorce	p 95
	• Marquage de la sonde	p 95
	• Extension d'amorce	p 95
1.1.6	Recherche de modèles de structure secondaire : utilisation du logiciel MFold	p 96
1.2.	Technique de FRET utilisée pour apprécier la distance relative séparant 2 zones ARN d'intérêt	p 96
<b>2. Méthodes utilisées pour la culture cellulaire</b>		<b>p 97</b>
<b>2.1 Culture cellulaire</b>		<b>p 97</b>
2.1.1	Lignées HEK293 et HeLa	p 97
2.1.2	Lignée cellulaire ES XX de type HP3-10	p 97
2.1.3	Entretien des cellules	p 97
2.1.4	Technique de transfection	p 98
<b>2.2. Technique d'immunofluorescence</b>		<b>p 98</b>
2.2.1.	Culture sur lames, fixation et perméabilisation des cellules	p 98
2.2.2	RNA-FISH et immunodétection de protéines	p 98
	• RNA-FISH	p 98
	• Immunodétection de protéines	p 99
2.2.3	Montage des lames et acquisition des images	p 100

2.3	Préparation d'extrait nucléaire	p 100
<b>3.</b>	<b>Méthodes utilisées pour purifier des complexes RNP</b>	<b>p 100</b>
3.1	Techniques de chromatographie d'affinité MBP-MS2	p 100
3.2	Techniques d'immunosélection des complexes RNP	p 101
3.2.1	Principe	p 101
3.2.2	Préparation des ARN	p 101
3.2.3	Préparation des billes	p 102
3.2.4	Formation des complexes RNP, immunosélection et analyse	p 102
3.3	Analyses des complexes ARN/Protéines par pontage covalent aux UV suivi d'immunosélection	p 103
3.3.1	Préparation des billes G-sépharose	p 103
3.3.2	Préparation des puits	p 103
3.3.3	Préparation du mélange réactionnel	p 103
3.3.4	Pontage covalent aux UV entre les protéines et les ARN et immunosélection	p 103
<b>4.</b>	<b>Production de protéines recombinantes chez <i>E. Coli</i></b>	<b>p 104</b>
4.1	Purification de la protéine PTB	p 104
4.1.1	Transformation, culture bactérienne et auto-induction	p 104
4.1.2	Préparation des extraits cellulaires	p 104
4.1.3	Chromatographie d'affinité sur colonne His-Trap couplée à un système FPLC	p 104
4.1.4	Estimation de la quantité et de la qualité de l'échantillon protéique	p 105
4.2	Purification des protéines KSRP-4KH et KSRP-34	p 105
4.2.1	Transformation et culture bactérienne	p 105
4.2.2	Préparation des extraits cellulaires et purification des protéines KSRP-4KH et KSRP-34 en batch	p 105
<b>5.</b>	<b>Analyses des interactions ARN/Protéines <i>in vitro</i></b>	<b>p 106</b>
5.1	Analyses des complexes ARN/Protéines par expériences de retard sur gel et empreinte aux RNAses	p 106
5.1.1	Préparation des ARN	p 106
5.1.2	Formation des complexes ARN/Protéines	p 106
5.2	Analyses des complexes ARN/Protéines par dichroïsme circulaire	p 107
<b>6.</b>	<b>Techniques utilisées pour étudier l'épissage de la région des A-repeats</b>	<b>p 107</b>
6.1	Méthodes spécifiques de l'étude de l'épissage <i>in vitro</i>	p 107
6.1.1	Préparation de la matrice de transcription	p 107
6.1.2	Transcription <i>in vitro</i>	p 108
6.1.3	Epissage <i>in vitro</i>	p 108
6.1.4	Conversion des ARN en ADN complémentaires par RT-PCR	p 108
6.2	Méthodes spécifiques de l'étude de l'épissage <i>in vivo</i>	p 109
6.2.1	Dosage de l'activité luciférase	p 109
6.2.2	Extraction des ARN totaux et RT-PCR	p 110
<b>III.</b>	<b>RESULTATS ET DISCUSSION</b>	<b>p 111</b>
<b>A.</b>	<b>Détermination expérimentale de la structure 2D de la région des A-repeats des ARN Xist murin et humain</b>	<b>p 111</b>

<b>1. Introduction</b>	<b>p 111</b>
<b>2. Analyse de la structure 2D des régions des A-repeats murine et humaine par emploi de sondes enzymatiques et chimiques</b>	<b>p 111</b>
2.1 Principes de l'emploi de sondes chimiques et enzymatiques pour l'analyse de la structure des ARN en solution	<b>p 111</b>
2.1.1 Les sondes enzymatiques et chimiques	<b>p 111</b>
2.1.2 Identification des positions de clivage et modification	<b>p 113</b>
2.1.3 Le choix des oligonucléotides amorce	<b>p 114</b>
2.2 Le choix des ARN utilisés	<b>p 117</b>
2.3 Recherche d'un modèle de structure 2D en accord avec les données obtenues	<b>p 119</b>
2.3.1 Confrontation des données expérimentales avec le modèle déjà proposé	<b>p 119</b>
2.3.2 Utilisation du logiciel MFold pour la recherche d'un nouveau modèle de structure 2D compatible avec nos données	<b>p 120</b>
2.4 Etude phylogénétique de la région des A-repeats	<b>p 123</b>
2.4.1 Analyse de la structure secondaire de la région des A-repeats humaine	<b>p 123</b>
2.4.2 Comparaison de séquences de la région des A-repeats chez les différents mammifères dont le gène <i>Xist</i> a été séquencé	<b>p 125</b>
2.5 Analyse physico-chimique par FRET de la structure 2D de la région des A-repeats murine	<b>p 126</b>
2.5.1 Stratégie expérimentale	<b>p 126</b>
2.5.2 Mise au point du protocole expérimental	<b>p 128</b>
• Choix des fluorophores	<b>p 128</b>
• Choix des oligonucléotides	<b>p 129</b>
• Préparation des hybrides ARN/oligonucléotides	<b>p 131</b>
• Paramètres de mesures	<b>p 131</b>
• Protocole adopté pour les mesures	<b>p 132</b>
<b>3. Expérience d'empreinte en extrait nucléaire</b>	<b>p 135</b>
3.1 Objectifs et stratégie expérimentale	<b>p 135</b>
3.2 Résultats obtenus	<b>p 136</b>
<b>B. Stratégie de purification et d'identification des complexes ribonucléoprotéiques formés avec la région des A-repeats</b>	<b>p 141</b>
1. Principe de la méthode de chromatographie d'affinité basée sur le système MS2/MBP	<b>p 142</b>
2. Mise au point des conditions de purification des RNP basée sur l'emploi d'extrait nucléaire de cellules HeLa	<b>p 143</b>
3. Les différents ARN utilisés pour la purification des RNP formées en extrait nucléaire	<b>p 145</b>
4. Des RNP ayant des compositions en protéines différentes sont obtenues en utilisant la région des A-repeats entière ou seulement 1, 2 ou 4 répétitions	<b>p 147</b>
5. Purification des RNP formées avec la région des A-repeats entière et des sous-fragments dans un extrait nucléaire de cellules ES non différenciées (J0) de souris	<b>p 151</b>
<b>C. 2D structure of the A region of Xist RNA and its implication for PRC2 association</b>	
Article soumis à PLoS Biology	<b>p 154</b>

<b>Annexe des chapitres A, B et C, Figures supplémentaires non contenues dans l'article concernant les analyses structurales</b>	<b>p201</b>
<b>D. Bilan des résultats obtenus par purification des RNP formées avec la région des A-repeats et ses sous-fragments</b>	<b>p212</b>
<b>1. Purification des RNP formées avec la région des A-repeats entière dans des extraits nucléaires de cellules ES de souris femelles prélevées à différents stades de la différenciation</b>	<b>p212</b>
<b>2. Stratégie utilisée pour essayer d'identifier les protéines des RNP pouvant avoir une importance fonctionnelle pour l'inactivation du chromosome X</b>	<b>p 214</b>
2.1 Les protéines des RNP qui pourraient être impliquées dans l'activité de la région des A-repeats sur la base des données bibliographiques	<b>p 217</b>
2.1.1 Les protéines qui pourraient être intéressantes et qui sont retrouvées avec tous les extraits que nous avons utilisés ou seulement avec l'extrait nucléaire de cellules HeLa	<b>p 217</b>
2.1.2 Les protéines qui pourraient avoir un lien avec l'inactivation du chromosome X et qui sont spécifiques aux RNP formées dans les extraits nucléaires de cellules ES de souris	<b>p 254</b>
2.2 Analyses complémentaires sur les protéines identifiées	<b>p 258</b>
2.2.1 Etude des taux relatifs de certaines protéines dans les différentes RNP étudiées par Western-blot	<b>p 258</b>
2.2.2 Etude de la localisation intra-cellulaire de certaines des protéines identifiées par immunofluorescence	<b>p 261</b>
<b>3. Etude de l'interaction possible de PTB et KSRP avec la région des A-repeats</b>	<b>p 263</b>
3.1 PTB	<b>p 266</b>
3.1.1 Vérification par l'approche d'immuno-sélection de l'interaction de PTB avec la région des A-repeats	<b>p 266</b>
3.1.2 Vérification de l'existence d'une interaction directe entre PTB et la région des A-repeats au sein d'un extrait nucléaire par pontage aux UV	<b>p 266</b>
3.1.3 Clonage, production et purification d'une protéine PTB recombinante	<b>p 268</b>
• Clonage, production et purification de la protéine PTB	<b>p 268</b>
• Contrôle de la qualité de l'échantillon protéique par dichroïsme circulaire	<b>p 269</b>
i. Principe de la méthode	<b>p 269</b>
ii. Résultats obtenus	<b>p 271</b>
3.1.4 Expériences de retard sur gel	<b>p 272</b>
3.1.5 Etude par empreinte aux RNases T1 et T2 des sites de fixation de la protéine PTB sur la région des A-repeats	<b>p 274</b>
3.2 KSRP	<b>p 277</b>
3.2.1 Clonage, production et purification de la protéine KSRP-4KH recombinante	<b>p 277</b>
3.2.2 Expériences de retard sur gel	<b>p 278</b>
3.3 Etude des complexes SLS1M/PTB et SLS1M/KSRP-4KH par dichroïsme circulaire	<b>p 280</b>
3.3.1 Principe de la méthode	<b>p 280</b>
3.3.2 Résultats	<b>p 282</b>

---

<b>E. Constructions d'outils génétiques en vue d'étudier ultérieurement la relation structure/fonction de la région des A-repeats</b>	<b>p 288</b>
<b>1. Objectifs et stratégie envisagée</b>	<b>p 288</b>
<b>2. Génération de la lignée cellulaire</b>	<b>p 289</b>
2.1 Principe et stratégie de la méthode	<b>p 289</b>
2.2 Résultats	<b>p 290</b>
<b>3. Génération de sites d'épissage bordant la région des A-repeats</b>	<b>p 292</b>
3.1 Présentation de l'épissage	<b>p 292</b>
3.2 Choix des séquences à muter et mutagenèse	<b>p 294</b>
3.3 Tests d'épissage <i>in vitro</i>	<b>p 297</b>
3.3.1 Production du transcrit à épisser	<b>p 297</b>
3.3.2 Epissage et mise en évidence des produits d'épissage	<b>p 298</b>
3.3.3 Résultats	<b>p 299</b>
3.4 Tests d'épissage <i>in cellulo</i>	<b>p 302</b>
3.4.1 Stratégie utilisée	<b>p 302</b>
3.4.2 Préparation des plasmides contenant ou non les sites 5' et 3' d'épissage qui puissent être transfectés dans des cellules HeLa	<b>p 302</b>
3.4.3 Analyse de l'épissage <i>in cellulo</i>	<b>p 303</b>
<b>III. CONCLUSION GENERALE ET PERSPECTIVES</b>	<b>p 306</b>
<b>Annexe du chapitre D</b>	

## ***Avant-propos***

Pendant de nombreuses années, les ARN ont été considérés comme de simples intermédiaires entre l'ADN qui contient les gènes et les protéines (ARNm). Néanmoins, la mise en évidence de différentes classes d'ARN non-codants tels que les UsnRNA (*small nuclear RNA*), les grands ARNnc (*no codant RNA*), les snoRNA (*small nucleolar RNA*) et les microARN a poussé la communauté scientifique à envisager de nouvelles fonctions encore insoupçonnées des ARN. L'un d'entre eux, découvert dès les années 1990, est un exemple majeur de grands ARN non-codants. Il s'appelle l'ARN Xist et est impliqué dans la compensation de dose de l'expression des gènes liés au chromosome X chez les mammifères. En effet, chez ces derniers, les génotypes XY et XX déterminent respectivement le sexe mâle et femelle. Il y a donc, chez la femelle, un chromosome X supplémentaire qui provoque un déséquilibre potentiel d'expression des gènes liés aux chromosomes X entre les 2 sexes. Ce dernier est alors compensé par l'extinction transcriptionnelle de l'un des deux chromosomes X chez les femelles. Ce processus est très important car une défaillance dans sa mise en place provoque la mort de l'embryon très tôt lors du développement. Toutefois, bien que l'existence d'une telle « inactivation du chromosome X » ait été établie par Mary Lyon dès 1963 (Lyon, 1963), la compréhension de ce mécanisme essentiel et hautement complexe est encore très partielle.

Exemple type de régulation épigénétique médiée par un ARN non-codant, l'inactivation du chromosome X est à plusieurs égards tout à fait particulière. Tout d'abord, elle induit un comportement radicalement opposé des deux chromosomes X qui pourtant sont génétiquement identiques. D'autres part, cette régulation est régie par un locus unique appelé le centre d'inactivation du chromosome X (*Xic*) qui produit l'ARN Xist (*X inactive specific transcript*). L'expression de cet ARN induit la recomposition des marques épigénétiques du futur chromosome X inactif, parmi lesquelles plusieurs modifications des histones et la méthylation de l'ADN. De façon intéressante, cette cascade de réactions concourt à l'extinction transcriptionnelle d'un chromosome entier. Du fait de la nécessité de comprendre quelles modifications des histones (« code histone ») sont induites par les ARN non-codants, l'importance de l'ARN Xist est grandissante. Ainsi, son étude entre parfaitement dans la dynamique de recherche de ces dernières années concernant cette classe de molécule et la compréhension du mode d'action de l'ARN Xist apportera des informations sans doute d'intérêt général pour la compréhension des mécanismes d'action des ARN non-codants intervenant dans les phénomènes épigénétiques.

# **INTRODUCTION**



# I. Introduction

## A. Généralités

### 1. Chromatine et épigénétique

#### 1.1 L'épigénétique : Histoire et Evolution

Contrairement à ce que l'on pourrait penser, l'épigénétique n'est pas une notion récente.

Dès le IV<sup>ème</sup> siècle avant J.-C, Aristote proposa deux concepts, ceci en privilégiant le second. Selon le premier, l'individu est préformé dès le commencement et ne résulte que d'un accroissement de taille à partir d'une forme miniature originelle (notion d' « *homunculus* » dans le cas de l'homme, logé dans la tête du spermatozoïde). Cette hypothèse a été reprise par la suite par les courants préformistes qui suggéraient que l'adulte est déjà pleinement constitué à l'état de germe dans l'oeuf. Une simple croissance des tissus et organes est nécessaire au cours du développement. Cette théorie impliquait que dans un des gamètes (les gamètes mâles de préférence...), on pouvait trouver autant de germes que de générations futures, l'autre gamète ne servant que de source de nutriments (... les gamètes femelles).

Selon le second concept plus particulièrement défendu par Aristote, l'individu est constitué grâce à la « formation par dessus » (c'est-à-dire l'épigenèse) de nouvelles structures, qui peuvent être comparée à la « construction d'un filet ». Ainsi, l'adulte se forme par étapes successives de différenciation de l'œuf dont les constituants sont apportés par les gamètes mâles et femelles. Ces gamètes ont la capacité d'établir des interactions qui mènent progressivement à la création d'un nouvel individu.

Ces deux hypothèses se sont opposées pendant de nombreuses années, jusqu'à ce que la théorie du préformisme soit définitivement abandonnée et ceci suite à la découverte au XIX<sup>ème</sup> siècle de deux éléments majeurs de la biologie. Tout d'abord, celle des cellules en tant qu'élément de base de tous les êtres vivants avec la découverte que les gamètes sont elles-mêmes des cellules uniques. Puis, peu après furent posées les bases de la génétique moderne mendélienne, décrivant l'égalité de contribution des deux parents à l'identité génétique de leurs enfants.

Waddington fut le premier à définir l'épigénétique et ceci avant même les découvertes de Watson et Crick. En effet, dès 1942, Waddington définit l'épigénétique comme la « branche de la biologie qui étudie les relations de cause à effet entre les gènes et leurs produits, faisant apparaître le phénotype ».

En 1953, Watson et Crick identifièrent la structure de l'ADN et l'essor de la notion de support moléculaire de l'information génétique permit le développement de la génétique moléculaire. Le code génétique commença ensuite à être étudié. Néanmoins, celui-ci se révéla plutôt instructif qu'explicatif, car il ne procure qu'une vision parcellaire de ce qu'est un

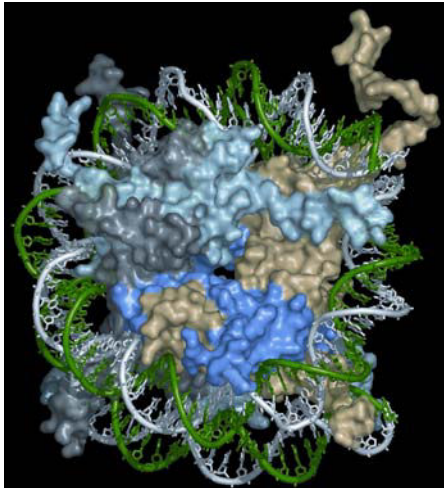
individu adulte. En effet, le code génétique correspond à une liste d'un grand nombre d'instructions régissant le développement d'un individu. L'exemple des vrais jumeaux n'ayant pas, malgré un support génétique identique, le même développement (par exemple le développement cérébral) suggère que l'information génétique dépend des conditions environnementales. De plus, il est essentiel que les cellules gardent en mémoire les instructions déjà mises en application, sans quoi les mêmes mécanismes cellulaires recommenceraient sans fin et aucun organisme multicellulaire correctement formé ne verrait le jour. La description du matériel génétique en tant que séquence d'ADN ne suffit donc plus à expliquer sa fonction. Ainsi, l'épigénétique apparaît comme un niveau d'information génique supplémentaire dictée d'une part par le matériel génétique et d'autre part par les interactions environnementales.

L'épigénétique est donc une notion complexe et c'est pourquoi, Denise Barlow, spécialiste reconnue de l'étude des gènes soumis à l'empreinte parentale, proposa dans les années 1990 cette définition : « L'épigénétique a toujours été l'ensemble de ces choses bizarres et merveilleuses que la génétique ne sait pas expliquer ».

## 1.2 Complexité de la chromatine et éléments régulateurs

### 1.2.1 Nucléosome et modifications covalentes des histones

Chaque noyau de cellules humaines compte approximativement 6 billions de paires de bases d'acide désoxyribonucléique (ADN) qui constituent les 23 paires de chromosomes (Nye *et al.*, 2005). L'ADN est donc fortement compacté et ordonné pour qu'il soit à la fois contenu dans un volume réduit tout en étant fonctionnel. Cet ADN compacté à l'intérieur du noyau s'appelle la chromatine. Le premier niveau d'organisation de la fibre d'ADN dans le noyau des cellules eucaryotes repose sur son empaquetage sous forme de nucléosomes. C'est en 1974 que Kornberg et Thomas proposent que le nucléosome constitue l'unité de base de la chromatine (Kornberg, 1974 ; Kornberg et Thomas, 1974). Ils décrivent cette particule dont le coeur est constitué par un octamère d'histones autour duquel s'enroule un segment d'ADN de 174 paires de bases. Cet octamère protéique comprend deux exemplaires de chaque histone H3, H4, H2A et H2B. Le corps des histones ou domaine C-terminal est globulaire alors que leurs extrémités N-terminales ne sont pas structurées (pour revue, Kouzarides, 2007). La première structure tridimensionnelle à haute résolution de ce complexe a été résolue en 1997 par Luger et ses collaborateurs (Luger *et al.*, 1997) (**Fig. 1**). Elle montre que le domaine C-terminal permet l'interaction avec l'ADN et avec les autres histones. L'extrémité N-terminale, quant à elle, se projette à l'extérieur du nucléosome et est donc accessible à d'autres protéines régulatrices.



**Figure 1 : Représentation d'un nucléosome**

L'histone H2A est représentée en bleu clair, H2B en gris, H3 en jaune et H4 en bleu foncé. Le double brin d'ADN est représenté en blanc et vert. Cette représentation met en évidence l'enroulement de la molécule d'ADN autour du corps protéique et la projection de l'extrémité N-terminale à l'extérieur du nucléosome.

D'après Luger *et al.*, 1997

Image obtenue à l'adresse suivante :

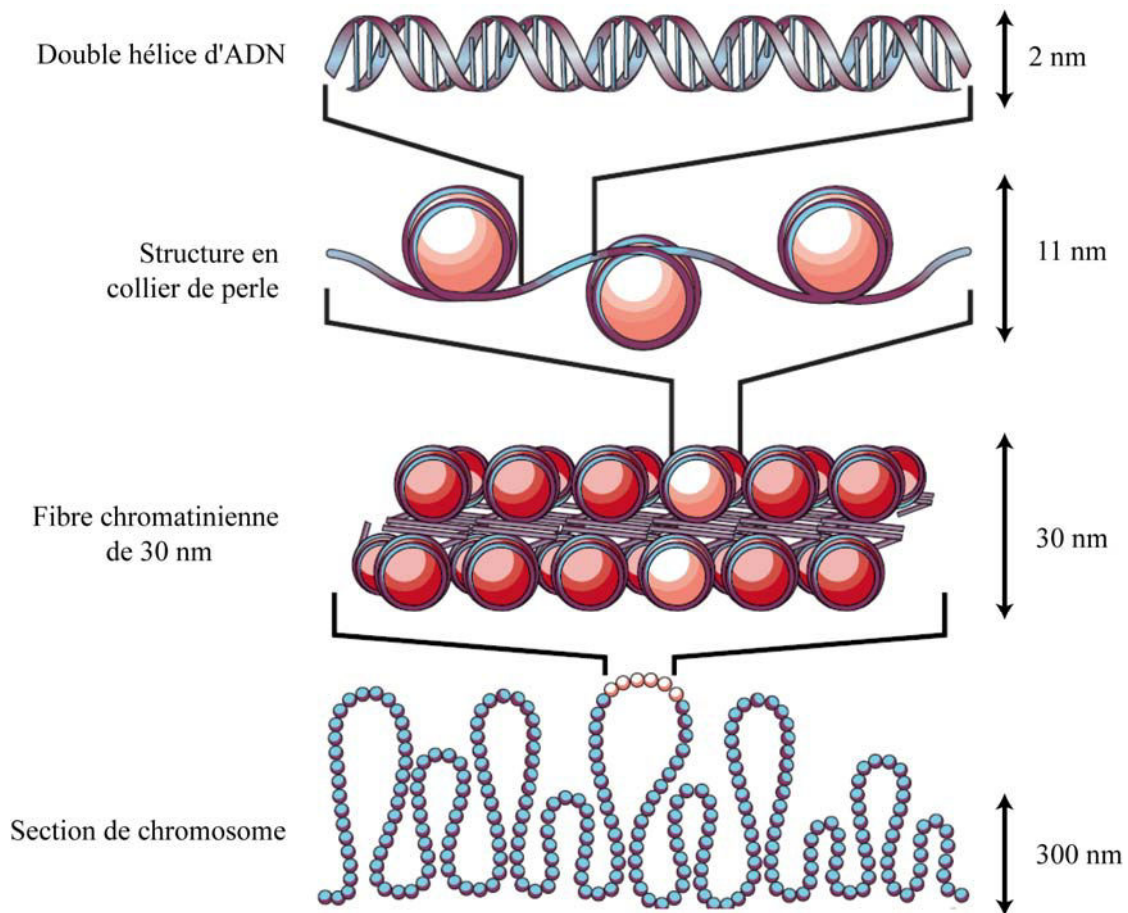
[http://www.chem.ucsb.edu/~molvisual/Img/142A/nucleosome\\_surf.png](http://www.chem.ucsb.edu/~molvisual/Img/142A/nucleosome_surf.png)

Les nucléosomes sont connectés entre eux par une séquence d'ADN appelée « ADN linker » de taille variable, pour former une structure chromatinienne de 11 nm comparée parfois à celle d'un collier de perles (Olins et Olins, 1974). L'ADN linker est le plus souvent associé à une cinquième protéine histone, l'histone H1 également appelée histone linker, impliquée dans l'espacement des nucléosomes et dans leur compaction (pour revue, Zlatanova et Van Holde, 1998 ; pour revue, Thomas, 1999). Ces histones stabilisent et modulent la distance entre les nucléosomes au niveau supérieur de compaction de la chromatine qu'est la fibre de 30 nm dans laquelle une hélice est formée à partir de l'ADN lié aux histones (Koop *et al.*, 2003 ; Fan *et al.*, 2005 ; pour revue, Felsenfeld et Groupine, 2003) (**Fig. 2**).

Ces structurations peuvent être labiles ce qui permet des régulations fines des mécanismes liés à la chromatine. On distingue deux types de chromatine en fonction de son état de compaction : l'euchromatine qui représente la portion du génome riche en gènes, souvent transcriptionnellement active, et qui est répliquée précocément au cours de la phase S du cycle cellulaire ; et l'hétérochromatine qui correspond à une structure condensée qui est pauvre en gènes exprimés et qui est à son tour divisée en deux catégories :

- la première appelée hétérochromatine constitutive, dont les exemples types sont l'hétérochromatine péricentrique et télomérique. Elle maintient sa structure compacte au cours de la mitose et est indépendante du type cellulaire.
- la seconde, l'hétérochromatine facultative est une structure plus dynamique et est mise en place à différents moments du développement pour réprimer certains ensembles de gènes dans des lignages spécifiques.

L'existence de ces différents types de chromatine participe largement à la régulation de l'expression des gènes. Pour comprendre ces régulations, il faut se focaliser sur la constitution même de la chromatine.



### Figure 2 : Organisation de l'ADN dans la structure chromatinienne

Le plus bas niveau d'organisation de l'ADN est le nucléosome constitué par l'octamère d'histones autour duquel la double hélice d'ADN s'enroule 2 fois. Cette organisation est décrite comme une structure rappelant celle d'un collier de perle.

Les nucléosomes sont reliés par un court fragment d'ADN linker dont la longueur relative est régulée par l'histone H1 qui compacte la chromatine sous forme de fibre chromatinienne de 30 nm de diamètre. Les détails régissant le niveau de structuration supérieure à celle-ci sont encore mal connus mais il est admis que ce haut degré de structuration atteint son maximum en phase métaphase lorsque les chromosomes sont observables.

D'après la revue Felsenfeld et Groupine, 2003.

En effet, la molécule d'ADN n'est pas inerte et elle est le siège de modifications structurales. Celles-ci témoignent de l'activité de protéines régulatrices qui, en agissant à divers niveaux conduisent à des variations de l'expression génique.

Premièrement, il existe de nombreux complexes protéiques ATP dépendants comme SWI/SNF, qui sont impliqués dans l'activation transcriptionnelle par « décompaction » partielle de l'ADN lié à l'octamère d'histones. Il en résulte un « écartement » des nucléosomes les uns par rapport aux autres, ce qui permet de moduler l'accessibilité des promoteurs (et autres séquences régulatrices) à la machinerie de transcription (pour revue, Mellor, 2005).

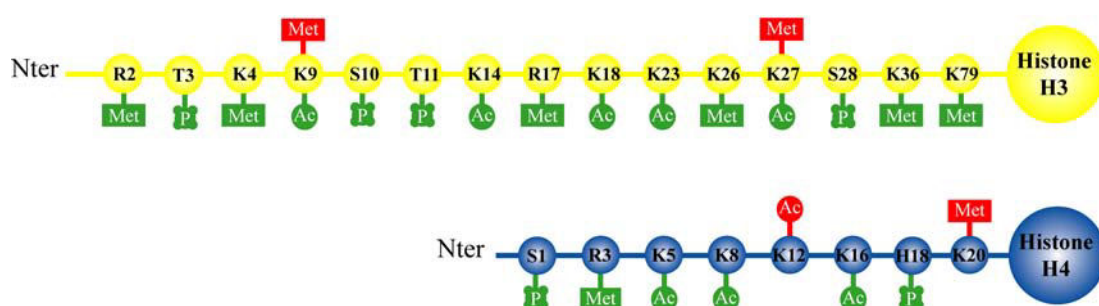
Deuxièmement, les histones sont le siège de nombreuses modifications covalentes. En effet, du fait de leur accessibilité, les queues N-terminal des histones sont la cible de plusieurs enzymes de modification. Ainsi, les résidus sérine et thréonine peuvent être modifiés par

phosphorylation, les lysines par acétylation, méthylation, ubiquitination et sumoylation. Les arginines peuvent être méthylées. Ces modifications qui concernent principalement les histones H3 et H4, ont été, entre autres, associées, soit à l'activation, soit à la répression transcriptionnelle (pour revue, Jenuwein et Allis, 2001 ; pour revue, Margueron *et al.*, 2005 ; Strahl et Allis, 2000). En effet, l'acétylation de plusieurs lysines des histones H3 (H3K9 et K14), l'hyper-acétylation de H4, ainsi que la méthylation de H3K4, sont plus particulièrement observées dans les régions d'euchromatine active c'est-à-dire associée à des régions transcriptionnellement actives (Bernstein *et al.*, 2005 ; Pokholok *et al.*, 2005 ; Schubeler *et al.*, 2004). En revanche, répression génique et états hétérochromatiniens sont généralement accompagnés de l'apposition de méthylations au niveau des lysines 9 et 27 de H3 (H3K9 et H3K27), ainsi que de la lysine 20 de H4 (pour revue, Sims *et al.*, 2003) (**Fig. 3**).

Comment de telles différences de modification peuvent-elles influencer l'activité transcriptionnelle ? Deux mécanismes sont envisagés :

Le premier, dit en *cis*, implique le caractère physico-chimique des modifications apposées et leurs influences sur l'enroulement de l'ADN au voisinage des promoteurs. En effet, ces modifications covalentes permettraient une structuration ou une déstructuration du nucléosome, permettant d'augmenter ou de diminuer l'état de tension de l'ADN enroulé autour de l'octamère d'histones. Elles faciliteraient ainsi la fixation de la machinerie de transcription sur des séquences régulatrices (pour revue, Turner, 1993).

Le deuxième, dit en *trans*, implique qu'une combinatoire complexe de modifications des histones permette le recrutement de complexes protéiques régulateurs (activateurs ou répresseurs) influençant l'état d'activité transcriptionnelle. Ce second mécanisme a permis de formuler le concept de « code histone » (Strahl et Allis, 2000 ; pour revue, Turner, 2000).



**Figure 3 : Les différents types de modifications des histones H3 et H4**

L'identité et la position des résidus modifiés sont indiquées par le code à 1 lettre des acides aminés, suivi du chiffre de référence à partir de l'extrémité N-terminale.

Les modifications sont indiquées de la façon suivante : Ac pour acétylation, Met pour méthylation et P pour phosphorylation.

Les résidus lysine et arginine peuvent être méthylés, les lysines peuvent être acétylées et les sérines et acides glutamiques peuvent être phosphorylés. Les modifications associées à la répression génique sont indiquées en rouge et celles associées à la chromatine active en vert.

D'après la revue Margueron *et al.*, (2005) ; Lavrov et Kibanov, (2007).

### 1.2.2 Régulation en *cis* : physico-chimie de modifications

Les queues N-terminales des histones ont la caractéristique d'être exposées à l'extérieur du nucléosome. Etant donné leur caractère flexible, elles ont la capacité d'explorer l'environnement proche du nucléosome. De plus, comme nous l'avons déjà mentionné, elles sont le siège de modifications covalentes. Ainsi, des groupements chimiques greffés à des positions particulières peuvent changer leurs propriétés physico-chimiques, ce qui influence leur environnement proche.

Sans même qu'il y ait modification covalente, la présence de résidus basiques, comme les lysines ou arginines chargées positivement au pH cellulaire (1 résidu  $\text{NH}_3^+$  pour les lysines et 2 pour les arginines), permet l'établissement d'interactions ioniques avec le squelette ribose-phosphate de l'ADN (chargé négativement par la présence de groupements phosphates). De même, grâce à leurs groupements hydroxyle, les sérines et les thréonines peuvent établir des liaisons hydrogène avec les groupements phosphate du squelette de l'ADN. Il est donc supposé que, dans leur état non-modifié, les queues des histones établissent des liaisons faibles avec l'ADN qui masquent des séquences régulatrices potentielles.

En revanche, la modification des histones pourrait permettre de rompre ces interactions. En effet, l'ajout de groupements acétyl sur les lysines et/ou arginines neutraliserait la charge positive. La phosphorylation des sérines et thréonines induirait une charge négative responsable d'une répulsion électrostatique avec les groupements phosphate des acides aminés et de l'ADN.

Ce système d'interaction nucléosome/ADN serait donc à la base des processus d'activation transcriptionnelle par le démasquage de séquences régulatrices. Toutefois, les effets des modifications des histones ne se réduisent pas à des changements électrostatiques modifiant les interactions ADN/histones, puisque l'effet des méthylations dépend de la position modifiée (exemple de di et tri-méthylation des H3K4 abordé ci-dessus), l'influence de ces modifications est donc bien plus complexe.

### 1.2.3 Régulation en *trans* : le code histone

La découverte d'une grande diversité de modification des histones, le fait qu'une même modification peut avoir différentes fonctions, et les différentes combinaisons de modifications mises en évidence ont poussé Brian Strahl et David Allis à suggérer l'existence du code des histones (Strahl et Allis, 2000). Selon ce modèle, toutes les modifications et leurs combinaisons constituent un « code » capable de réguler le recrutement spécifique de facteurs impliqués dans la régulation de la structure et de la fonction de la chromatine.

- La diversité des modifications des histones

Tout d'abord, il est intéressant de noter que le large ensemble de modifications covalentes que nous avons mentionné ci-dessus permet de générer une multitude de combinaisons de modifications. Par ailleurs, une modification donnée peut en induire ou en bloquer d'autres (pour revue, Margueron *et al.*, 2005). Par exemple, la méthylation de H3K4 qui marque les régions euchromatiques, facilite l'acétylation des lysines des histones H3 et H4 par la protéine p300 à activité acétyltransférase. De même, la méthylation de H4R3 par la protéine PRMT1 entraîne l'euchromatinisation de grandes régions d'ADN par acétylation de H3 et H4, et l'inhibition des méthylations des résidus H3K9 et H3K27 (Huang *et al.*, 2005).

- La combinatoire des modifications des histones

En plus de cette coordination, un deuxième niveau de complexité émane de la possibilité de former 1, 2, ou 3 groupements méthyle sur un résidu donné (H3K4, K9, et K27, ainsi que sur H4K20). Les régions du génome peuvent être classées en fonction des modifications des histones qu'elles contiennent. Chez les mammifères, la combinaison de triméthylation de H3K9 et de H4K20, avec la mono-méthylation de H3K27 et une hypo-acétylation globale, correspond à des marques épigénétiques caractéristiques de l'hétérochromatine constitutive. La di-méthylation de H3K9, la triméthylation de H3K27 et la mono-méthylation de H4K20 caractérisent l'hétérochromatine facultative, qui est mise en place au cours du développement embryonnaire (pour revue, Martin et Zhang, 2005). Toujours dans ce contexte de complexité grandissant, au niveau de la méthylation H3K4, une différence fonctionnelle très marquée est observée entre l'ajout de 2 ou de 3 groupements méthyl. En effet, H3K4me2 peut être associée aussi bien aux promoteurs inactifs qu'aux promoteurs et régions internes des gènes activement transcrits (Santos-Rosa *et al.*, 2002 ; Schneider *et al.*, 2004), alors que H3K4me3 semble être associée uniquement aux régions promotrices actives.

- Lecteurs et effecteurs des modifications des histones

Le code histone implique la présence d'enzymes et de protéines capables de le décrypter et d'agir en conséquence. La mise en évidence de domaines de reconnaissance des modifications covalentes des histones sur de nombreuses protéines régulatrices de la structure et de la fonction de la chromatine est en accord avec l'existence de ce code. Deux catégories de domaines protéiques ont été bien décrites. Les bromodomains capables de reconnaître spécifiquement les lysines acétylées des extrémités N-terminales des histones. De tels domaines sont présents dans les protéines impliquées dans la régulation de la structure chromatinienne liée aux processus d'activation, comme P/CAF, CBP/p300, ou SWI/SNF.

Les chromodomains sont au contraire capables de reconnaître des lysines méthylées (pour revue, de la Cruz *et al.*, 2005a). Par exemple, le chromodomaine de HP1 reconnaît les résidus H3K9 di/tri-méthylés (pour revue, Lachner et Jenuwein, 2002), celui des protéines

polycomb est capable de fixer la tri-méthylation des H3K27 (pour revue, Simon et Tamkun, 2002). De façon intéressante, la spécificité de reconnaissance des modifications par ces domaines est extrêmement élevée, puisqu'il suffit de remplacer le chromodomaine des protéines polycomb par celui de HP1 pour relocaliser ces protéines au site de localisation de HP1, les centromères (Platero *et al.*, 1995). D'autres domaines, comme les domaines WD-40 sont capables d'interagir avec les résidus H3K4 di-méthylés. Le domaine PHD de NURF (un co-activateur transcriptionnel associé à de nombreuses protéines à activité acétyl-transférase et de remodelage ATP-dépendant), est capable de reconnaître la triméthylation de H3K4 (Li *et al.*, 2006 ; Wysocka *et al.*, 2006).

L'existence de différentes protéines capables de reconnaître spécifiquement une ou plusieurs modifications portées par les histones a permis d'expliquer, au moins en partie, l'existence de grands domaines chromatinien caractérisés par un ensemble de marques épigénétiques précises. Un exemple caractéristique est celui de l'hétérochromatine péri-centromérique, où la tri-méthylation de H3K9 induite par l'enzyme SuVar39h est reconnue par HP1. HP1 est capable de dimériser et d'interagir avec SuVar39h, ce qui permet, la propagation de la tri-méthylation à partir d'une seule région modifiée et l'établissement de domaines compacts de répression. La tri-méthylation péri-centromérique de H4K20, ainsi que la méthylation des CpG (voir section suivante), semblent dépendre de celle de H3K9, puisqu'elles sont perdues dans des mutants de SuVar39h (Lehnertz *et al.*, 2003 ; Schotta *et al.*, 2004). Ceci s'explique par le fait que l'enzyme responsable de la tri-méthylation de H4K20, SuVar420h, (Schotta *et al.*, 2004), et Dnmt1 et 3b ("*DNA (cytosine 5) methyltransferase*" 1 et 3b) (Lehnertz *et al.*, 2003 ; Schotta *et al.*, 2004), enzymes méthylant les îlots CpG, interagissent avec HP1. Ainsi, une fois que la tri-méthylation de H3K9 est établie sur une région localisée, le recrutement de HP1 permet la propagation de cette marque, ainsi que de la tri-méthylation de H4K20 et la méthylation des CpG, sur l'ensemble d'un domaine hétérochromatinien.

Par un mécanisme similaire, la tri-méthylation de H3K27 induite par le complexe polycomb PRC2 est reconnue par le chromodomaine des membres du complexe polycomb PRC1, lequel bloque l'accès du complexe de remodelage de la chromatine ATP-dépendant SWI-SNF (Shao *et al.*, 1999). Cela maintient une structure chromatinienne condensée, propre aux régions inactives du génome. Notons aussi que le complexe PRC2 reconnaît les H3K27me3 qu'il génère, ce qui suggère qu'il est aussi l'initiateur de la propagation de cette marque épigénétique (Hansen *et al.*, 2008). Il a aussi été démontré que la protéine Ezh2 (la méthyltransférase agissant sur H3K27) est cruciale pour la méthylation de l'ADN. En effet, Ezh2 interagit avec les Dnmt et la fixation de ces ADN méthylases dépend de la présence de Ezh2 sur les gènes à réprimer (Viré *et al.*, 2006).



La méthylation de H4R3 médiée par la méthylase PRMT5 permet le recrutement de l'ADN méthylase Dnmt3a. Cette enzyme permet l'apposition de groupements méthyl sur l'ADN. C'était la première mise en évidence d'un couplage direct entre les « Arginine-méthyltransférases » et les ADN méthyltransférases (Zhao *et al.*, 2009).

L'ensemble de ces données a permis de mettre en évidence un couplage et une régulation complexe entre modification des histones et méthylation de l'ADN.

En conclusion, la multitude des marques d'histones, de domaines protéiques de reconnaissance de ces marques, d'interactions protéine-protéine entre modificateurs d'histone et remodeleurs des nucléosomes, confère aux modifications covalentes des histones un potentiel régulateur d'une très grande complexité. Les modifications des histones peuvent par combinaison des marques épigénétiques qu'elles portent, conférer à une/des région(s) donnée(s) une information supplémentaire par rapport à la séquence nucléotidique, ce qui permet le traitement différentiel des gènes au cours du développement.

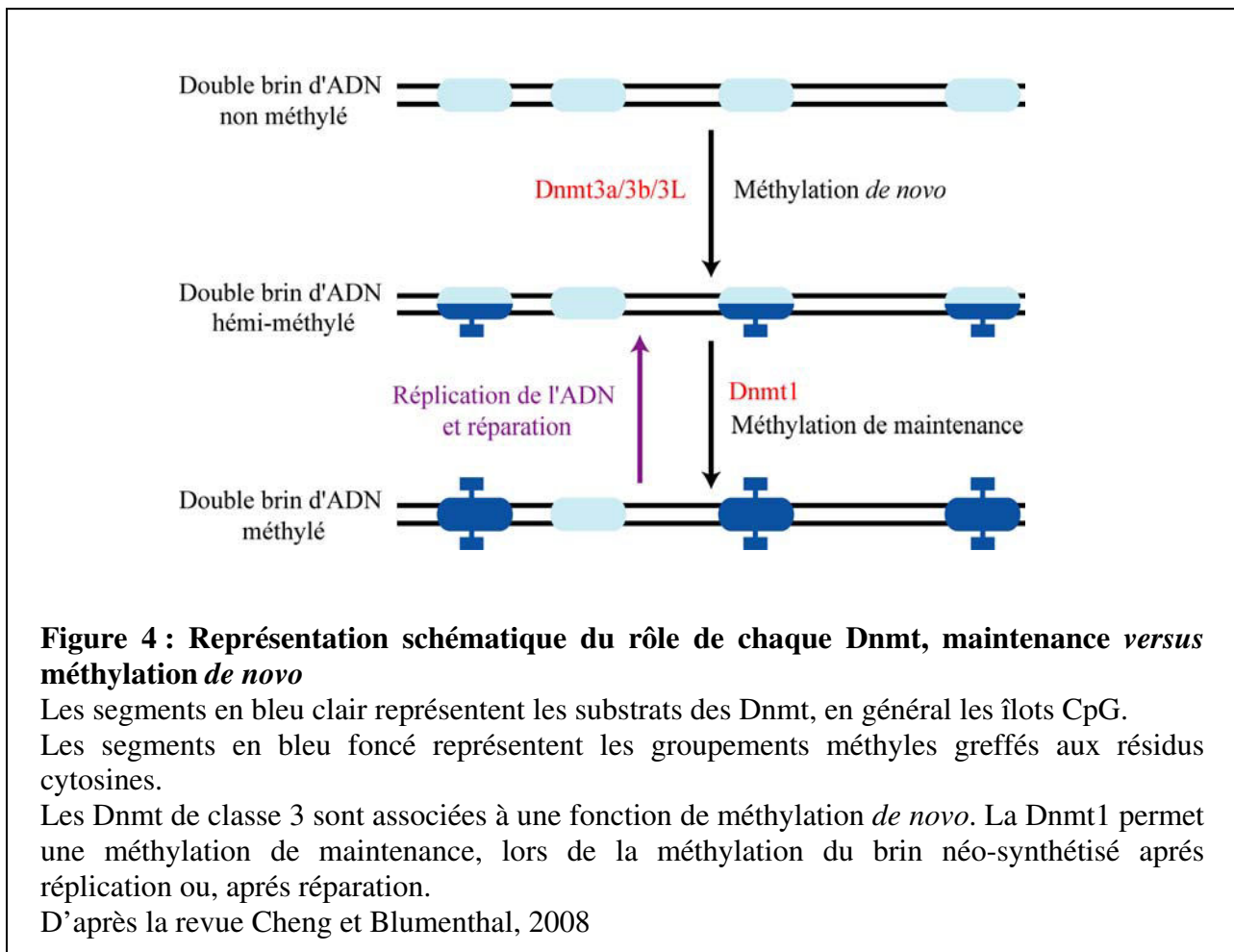
#### 1.2.4 Méthylation de l'ADN

La modification covalente des cytosines par méthylation déjà mentionnée ci-dessus, est une marque épigénétique importante pour de nombreux processus cellulaires. Le profil de méthylation de l'ADN est stable et transmis à chaque division cellulaire (Wigler *et al.*, 1981). Il s'agit de la seule modification de la chromatine dont l'héritabilité a été démontrée (pour revue, Bestor, 2000). La méthylation de l'ADN existe chez des espèces phylogénétiquement très éloignées, depuis les plantes jusqu'à l'homme, mais semble absente chez certains organismes comme le nématode *C.elegans* et la levure *S.cerevisiae* (pour revue, Bird, 2002). Chez les mammifères, la seule marque épigénétique connue, touchant directement la molécule d'ADN, est la méthylation des cytosines, suivies d'une guanine. Ces dinucléotides CpG sont souvent retrouvés sous forme d'îlots autour des régions promotrices des gènes, où une grande densité de CpG est présente par rapport au reste du génome. Ces îlots sont très souvent le siège de méthylation qui sont associées à la répression génique. D'une manière générale, 60 à 90% des dinucléotides CpG du génome de mammifères sont méthylés (pour revue, Ng et Bird, 1999).

La méthylation de l'ADN joue un rôle fondamental dans la répression transcriptionnelle (pour revue, Bestor, 2000). Elle est en particulier essentielle pour la mise en place de l'empreinte génomique parentale comme nous le verrons dans la partie suivante, l'inactivation du chromosome X, la répression des rétrotransposons et le développement embryonnaire (pour revue, Bird, 2002).

Comme nous l'avons déjà mentionné, la méthylation des dinucléotides CpG est catalysée par un groupe conservé de protéines appelées "*DNA (cytosine 5) methyltransferase*" (Dnmt). Chez les mammifères, il en existe 4 qui fonctionnent de façon coordonnée et

permettent de réguler finement la méthylation du génome : Dnmt1, Dnmt3a/3b/3L (pour revue, Bestor, 2000). Il a été démontré que le profil de méthylation de l'ADN d'une lignée cellulaire résulte de la combinaison de deux voies du système de méthylation (pour revue, Bestor, 2000 ; pour revue, Bird, 2002) : la méthylation *de novo* au cours du développement embryonnaire assurée par Dnmt3a et Dnmt3b et la méthylation de maintenance catalysée par Dnmt1, qui permet la méthylation d'une section d'ADN nouvellement réparée et la transmission du profil de méthylation entre les divisions cellulaires (**Fig. 4**).



Nous consacrerons la suite de cette partie de l'introduction à la présentation des différentes Dnmt. Nous commencerons par la Dnmt1. L'inactivation de son gène chez la souris est létale. L'absence de Dnmt1 induit une déméthylation massive du génome, provoquant l'expression bi-allélique des gènes soumis à l'empreinte génomique parentale (Li *et al.*, 1992). Les rétrotransposons se réactivent chez les mutants  $Dnmt1^{-/-}$ , dont l'ADN n'est pas méthylé, démontrant que la méthylation de l'ADN est nécessaire à leur répression (Bourc'his et Bestor, 2004 ; Walsh *et al.*, 1998). *In vitro*, Dnmt1 est 5 à 30 fois plus active sur les substrats héli-méthylés que sur les substrats non-méthylés, ce qui est en accord avec la fonction de maintenance de cette enzyme (Yoder *et al.*, 1997). Remarquons que dans certaines cellules cancéreuses, la méthylation de l'ADN peut être maintenue en absence de

Dnmt1, alors qu'elle est l'unique ADN méthyltransférase de maintien connue à ce jour (Rhee *et al.*, 2000). Il doit donc exister une voie alternative du maintien de la méthylation de l'ADN.

En ce qui concerne les enzymes Dnmt3a et Dnmt3b, les doubles mutants Dnmt3a<sup>-/-</sup>, Dnmt3b<sup>-/-</sup> meurent pendant la gestation et présentent une déméthylation globale du génome (Okano *et al.*, 1999). Les cellules ES Dnmt3a<sup>-/-</sup>, Dnmt3b<sup>-/-</sup>, sont incapables de méthyler un ADN rétroviral nouvellement intégré ce qui démontre la nécessité de Dnmt3a/b pour établir la méthylation *de novo* lors du développement (Okano *et al.*, 1999).

Le mécanisme par lequel la méthylation de l'ADN participe à la répression transcriptionnelle n'est pas connu. Deux hypothèses ont été proposées. La première repose sur l'idée que la méthylation de l'ADN altère la liaison ADN-Protéines, en particulier de facteurs de transcription dont les séquences de fixation contiennent des CpG (pour revue, Tate et Bird, 1993 ; Watt et Molloy, 1988). C'est le cas des facteurs de transcription Ets1 ou CTCF dont la fixation à l'ADN est inhibée par méthylation (Bell et Felsenfeld, 2000 ; Maier *et al.*, 2003). Le deuxième mécanisme repose sur l'existence de protéines qui fixeraient spécifiquement sur les CpG méthylés, cela de façon indépendante de la séquence nucléotidique dans certains cas. Chez les mammifères, seulement quelques protéines se sont avérées avoir cette propriété. C'est le cas de MeCP2, MBD1, 2, et 3, qui partagent un domaine protéique commun de fixation à l'ADN (Hendrich et Bird, 1998). En plus, un facteur de transcription nommé Kaiso est aussi capable de se fixer à l'ADN méthylé lorsque les CpG sont méthylés (Prokhortchouk *et al.*, 2001). Ces facteurs de transcription sont capables de recruter différentes enzymes de remodelage de la chromatine responsables, en particulier, de la désacétylation des histones et de la méthylation du résidu H3K9 (Fujita *et al.*, 2003). D'une manière générale, une relation étroite existe entre méthylation de l'ADN et modifications répressives des histones aussi bien chez les plantes, que chez les métazoaires, ce qui établit une boucle d'auto-propagation et de stabilisation de l'hétérochromatine comme nous l'avons déjà explicité plus haut (pour revue, Fuks, 2005).

L'activité de méthylation de l'ADN est aussi étroitement couplée à la réplication, puisque Dnmt1 et PCNA, un facteur d'élongation de l'ADN polymérase, interagissent ensemble (Chuang *et al.*, 1997). Plus récemment, il a été montré qu'une protéine de fixation aux CpG méthylés, MBD1, permet le recrutement de l'enzyme de di-méthylation de H3K9 (SETDB1), ceci grâce au complexe d'assemblage de la chromatine réplication dépendant CAF1 (Sarraf et Stancheva, 2004). Par ce mécanisme, le maintien de la méthylation par Dnmt1 permettrait à MBD1 d'induire la modification des histones lorsqu'elles sont incorporées à la chromatine nouvellement synthétisée grâce à CAF1.

Comme nous l'avons déjà mentionné, l'enzyme de tri-méthylation de H3K27 du complexe polycomb PRC2, Ezh2, contrôle la méthylation des CpG *via* le recrutement des trois enzymes de méthylation Dnmt1, 3a, et 3b (Vire *et al.*, 2006). De plus, le complexe PRC1

semble aussi être impliqué car la protéine Bmi1 (membre du complexe PRC1) interagit directement avec la protéine Dmap1, associée elle-même avec Dnmt1 (Negishi *et al.*, 2007). Ces dernières données impliquent que les trois grands systèmes de répression épigénétique, c'est à dire (H3K9 méthylée / HP1 / SuVar39h), (H3K27 méthylée / PRC2 / PRC1) et (CpG méthylé / Dnmt) coopèrent fonctionnellement pour initier et/ou maintenir l'état répressif. Certaines protéines ayant la capacité de se lier spécifiquement aux CpG méthylés dont MeCP2 et MBD2 pourraient être impliquées dans la répression des gènes dont le promoteur est méthylé (Hendrich et Bird, 1998 ; Nan *et al.*, 1997).

On est sans doute loin de les avoir toutes répertoriées. En effet, le phénotype de souris déficientes pour MeCP2 et MBD2 est différent de celui de souris déficientes pour les Dnmt (Hendrich *et al.*, 2001 ; Hendrich et Bird, 1998). De plus, la délétion des gènes de MeCP2 et MBD2 n'induit pas d'activation des séquences réprimées par la méthylation, telles que les rétrotransposons et les gènes soumis à empreinte parentale (Hendrich *et al.*, 2001 ; Hendrich et Bird, 1998). Dans ce cadre, il a été proposé que la méthylation de l'ADN entraîne la répression des rétrotransposons et des gènes soumis à l'empreinte par le recrutement du variant d'histone macroH2A1 (Choo *et al.*, 2006 ; Damelin et Bestor, 2007).

Dnmt3L est la moins décrite à ce jour. Son rôle est encore mal connu. Elle ne participe pas à l'activité méthylase, car il lui manque les résidus nécessaires au niveau de son site actif. Par contre, cette protéine interagit avec l'extrémité N-terminale de l'histone H3 non méthylé en position K4 (Ooi *et al.*, 2007) et avec les Dnmt de classe 3 (Margot *et al.*, 2003 ; Gowher *et al.*, 2005). Ce qui suggère l'existence de liens entre les protéines histones, leur modification et l'ADN.

Comme nous l'avons cité ci-dessus, la méthylation de l'ADN est essentielle au développement embryonnaire. En effet, durant les étapes de divisions cellulaires pré-implantatoires, une vague de déméthylation a lieu. Le génome paternel et le génome maternel sont déméthylés (Mayer *et al.*, 2000 ; Oswald *et al.*, 2000 ; pour revue, Li, 2002 ; Rougier *et al.*, 1998). Malgré cette déméthylation, les cellules de la gastrula présentent un degré de méthylation globale important, indiquant que l'implantation de l'embryon est accompagnée d'évènements de méthylation *de novo* (pour revue, Jaenisch, 1997). L'entrée des cellules dans les processus de différenciation induit, dans certaines lignées, de nouvelles vagues de déméthylation (pour revue, Jaenisch et Bird, 2003). En particulier, les cellules de la lignée germinale qui portent un profil de méthylation des CpG probablement identique à celui des cellules somatiques environnantes, subissent une déméthylation globale à un stade précoce du développement des gonades (Hajkova *et al.*, 2002). Cette déméthylation semble affecter l'ensemble du génome, à l'exception des séquences répétées, et permet que chaque *locus*, chaque chromosome, soit compétent pour recevoir les nouvelles marques de méthylation nécessaires à la mise en place du profil d'expression génique spécifique des gamètes. Cette

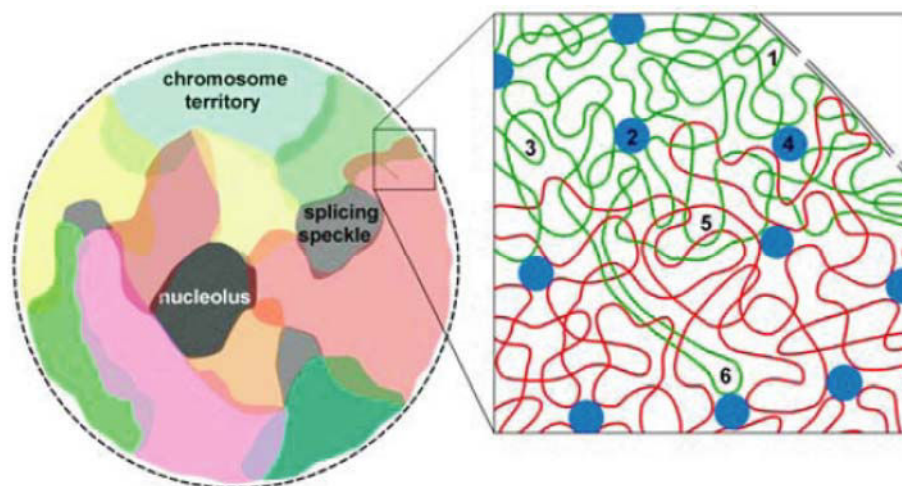
dynamique de déméthylation globale/méthylation/déméthylation sélective, suggère fortement que la méthylation de l'ADN joue un rôle important dans la spéciation des différentes lignées. C'est en effet le moyen d'éteindre à long terme de nombreux gènes à expression tissu-spécifique (Higgs *et al.*, 2006).

### 1.2.5 Compartiments nucléaires et répression génique

L'organisation génomique du noyau apparaît de plus en plus essentielle à la régulation de l'expression des gènes et la compartimentation spatiale d'un chromosome dans un territoire chromosomique nucléaire donné semble avoir une signification biologique importante (pour revue, Taddei, 2007 ; pour revue, Heard et Bickmore, 2007). Ainsi au cours des dernières années, des démonstrations expérimentales de l'existence de Territoires Chromosomiques (CT pour Chromosomal Territories) ont pu être obtenues par des analyses de fluorescence *in situ* d'ARN et d'ADN et de microscopie. De nombreux modèles proposent l'existence de relations étroites entre l'organisation des CT et la machinerie de transcription dans le noyau (pour revue, Heard et Bickmore, 2007). Néanmoins, le lien entre activité transcriptionnelle des gènes et positionnement dans des territoires chromosomiques définis est encore débattue (pour revue, Cremer et Cremer, 2001 ; pour revue, Williams, 2003). Ceci vraisemblablement du fait de l'imprécision de la définition des territoires chromosomiques. Les gènes sont souvent localisés à la périphérie de ces derniers (Moen *et al.*, 2004 ; Nogami *et al.*, 2000 ; Scheuermann *et al.*, 2004). Les modèles initiaux suggéraient qu'il pourrait y avoir une limite assez distincte entre la surface d'un CT et d'un domaine interchromosomique (ICD) contenant la machinerie transcriptionnelle (**Fig. 5**). Dans ces modèles, le CT est un domaine impénétrable par la machinerie transcriptionnelle et les gènes sont localisés à sa surface (Zirbel *et al.*, 1993). Les expériences montrent néanmoins que pris individuellement, les gènes peuvent se retrouver aussi bien à l'intérieur, à la périphérie ou à l'extérieur des territoires chromosomiques lorsqu'ils sont exprimés (Mahy *et al.*, 2002 ; Osborne *et al.*, 2004 ; pour revue, Parada et Misteli, 2002). De plus, les sites de transcription sont dispersés au sein du CT (Sadoni et Zink, 2004 ; Verschure *et al.*, 1999 ; Mahy *et al.*, 2002). Ces données sont incompatibles avec les modèles initialement envisagés. La situation prête moins à controverse lorsqu'on s'intéresse à des régions riches en gènes, prises dans leur ensemble, car leur relocalisation vers l'extérieur des territoires chromosomiques est bien corrélée avec leur expression (Dietzel *et al.*, 1999 ; Mahy *et al.*, 2002 ; Volpi *et al.*, 2000 ; pour revue, Chambeyron et Bickmore, 2004). Il a donc été ensuite suggéré que le CT a beaucoup plus de structures invaginées que ce que l'on imaginait. Ainsi, il existerait un compartiment interchromatinien de "chromatine-libre" (IC) qui forme des tunnels et ces derniers pénétreraient dans le CT (pour revue, Cremer *et al.*, 2006). L'analyse des propriétés dynamiques de la chromatine, de la localisation des gènes à l'extérieur des CT et un réexamen des interactions

s'établissant entre les gènes et les CT (Branco et Pombo, 2006 ; Simonis *et al.*, 2006) ont conduit à avoir une vue beaucoup plus plastique de cette organisation. Ces études ont effectivement démontré que les CT s'entremêlent de façon significative dans le noyau interphasique des cellules humaines. De plus, le fait de bloquer la transcription change le profil d'entremêlement tout en conservant l'état de compaction de la chromatine et le positionnement du chromosome. Ces données indiquent que les liens existant entre la transcription et les CT peuvent influencer l'organisation fine au sein des chromosomes. En conclusion, tout porte à croire que les territoires chromosomiques sont des entités versatiles, dont les changements d'organisation participent à la régulation de l'expression des gènes. Le modèle d'organisation de la chromatine dans les noyaux de mammifères le plus largement admis et le plus en accord avec les données expérimentales est le modèle appelé "Interchromosomal Network" (Branco et Pombo, 2006). Selon ce modèle, les chromosomes occupent des territoires qui s'entremêlent et ceci de façon plus importante à leurs extrémités. L'événement de transcription influence le degré d'entremêlement entre des CT spécifiques, probablement du fait de l'association entre *locus* particuliers.

Nous venons donc de voir que la structure et l'organisation du territoire chromosomique (CT) représente l'environnement nucléaire immédiat d'un gène et est important pour la régulation de son expression. Nous verrons par la suite les caractéristiques de l'organisation nucléaire du chromosome X devenu inactif.



### Figure 5 : Modèle du réseau inter-chromosomique

Dans ce modèle, les chromosomes occupent des territoires (CT) qui s'entremêlent. L'ADN qui compose ces CT a des propriétés qui déterminent sa compaction, sa mobilité et son affinité pour les composants nucléaires.

1. Attachements à la membrane nucléaire
2. Contacts intrachromosomiques maintenus par des points d'ancrage
3. Contacts intrachromosomiques par diffusion
4. Contacts interchromosomiques maintenus par des points d'ancrage
5. Contacts interchromosomiques par diffusion
6. Boucle de chromatine s'étendant dans d'autres CT

D'après Branco et Pombo, 2006

### 1.3 ARN non-codant et régulations épigénétiques

#### 1.3.1 ARN non-codant et complexité du vivant

Pendant de nombreuses années, les ARN ont été considérés comme des intermédiaires entre l'ADN codant les gènes et les protéines. Cette restriction a conduit à limiter les premières études des génomes à la simple recherche des gènes codants des protéines, et ceci au départ, par des stratégies de génétique inverse. Suite à la découverte de chaque protéine, leurs gènes étaient recherchés. Le séquençage de différents génomes a révélé, d'une part, que le nombre de gènes identifiés est bien inférieur à celui des protéines. Cette observation fut expliquée par la mise en évidence des mécanismes d'épissage alternatif des ARN pré-messagers, qui renforce la complexité du protéome sans augmentation du nombre de gènes. D'autre part, il s'est avéré qu'au moins 98% de la production d'ARN chez les mammifères supérieurs n'a aucun potentiel codant (pour revue, Mattick, 2003). Ces données montraient que le nombre de gènes codant des protéines ne pouvait pas expliquer à lui seul la complexité du vivant (pour revue, Mattick, 2001). Ainsi les ARN ne sont pas que des intermédiaires entre le support de l'information génétique et les protéines. Ils interviennent aussi dans de nombreuses autres fonctions et à différents niveaux de la régulation de l'expression génique.

En plus des ARNr et ARNt essentiels au processus de traduction, un nombre croissant d'ARN non-codants à fonction structurale, catalytique ou régulatrice, ont été décrits au cours des dix dernières années. Ils participent à de nombreux processus comme la régulation des taux d'expression de gènes codants, au niveau de la transcription (*7SK*) et de la structure de la chromatine (*siRNA*), ou encore au niveau de la stabilité des ARNm ou de leur efficacité de traduction (*miRNA*). Ils participent aussi à d'autres phénomènes liés à l'expression des gènes comme l'épissage (*UsnRNA*) et les modifications post-transcriptionnelles des ARN (*snoRNA*, *scaRNA*). D'autres ARN non-codants, souvent de grande taille, ont un rôle déterminant dans des phénomènes épigénétiques complexes comme l'empreinte parentale (*Air*, *Kcnq1ot1*) ou la compensation de l'expression des gènes en fonction leur nombre (*Xist*, *roX1* et 2). Souvent, leur expression est spatio-temporellement régulée au cours du développement embryonnaire et de la vie de l'adulte (pour revue, Mattick et Makunin, 2006 ; pour revue, Morey et Avner, 2004 ; pour revue, O'Gorman *et al.*, 2006 ; pour revue, Zamore et Haley, 2005). De façon intéressante, plusieurs ARN non-codants ont été reliés à la régulation de la structure chromatinienne, notamment dans le cas de mécanismes épigénétiques comme le maintien de l'hétérochromatine péri-centromérique, l'empreinte parentale ou l'inactivation du chromosome X (pour revue, Bernstein et Allis, 2005 ; pour revue, Morey et Avner, 2004 ; pour revue, O'Neill, 2005). A titre d'exemple, nous allons brièvement décrire le rôle des ARN non-codants dans la régulation de deux *loci* soumis à l'empreinte génomique et dans la formation et le maintien de l'hétérochromatine centromérique. Les relations entre ARN non-codant et structure chromatinienne dans

l'inactivation du chromosome X, thématique qui a fait l'objet de mon travail de thèse, seront abordées ensuite en détail.

### 1.3.2 ARN non-codant et expression soumise à l'empreinte

En 1984, des embryons diploïdes de souris gynogéniques (homozygote pour le génome maternel) ou androgéniques (homozygote pour le génome paternel) ont été créés par des expériences de transplantation de noyaux au stade une cellule (MacGrath et Solter, 1984 ; Surani *et al.*, 1984). L'analyse de ces embryons a montré qu'ils étaient incapables de se développer correctement. Ainsi fut démontré que bien que les 2 sexes contribuent de façon équivalente au matériel informatif du zygote, l'information génétique venant de chaque parent n'est pas nécessairement équivalente d'un point de vue fonctionnel. Ainsi, émergea pour la première fois le concept d'empreinte parentale (MacGrath et Solter, 1984 ; Surani *et al.*, 1984), concept qui sous-entend que certains gènes diploïdes sont exprimés de façon mono-allélique et que cette expression est dépendante de son origine parentale. Depuis l'émergence de ce nouveau concept, diverses études ont permis de caractériser ces gènes soumis à l'empreinte parentale (pour revue, Edwards et Ferguson-Smith, 2007 ; pour revue, Verona *et al.*, 2003). Ces gènes sont épigénétiquement marqués par leurs origines sexuelles de manière à ce que l'un des allèles soit exprimé et l'autre soit réprimé. La nature de ces marques épigénétiques est en cours d'investigation, il est admis qu'elles apparaissent dans les lignées germinales. En effet, c'est à ce moment là que les matériels génétiques sont compartimentés et qu'ils peuvent être modifiés de façon différentielle (Stöger *et al.*, 1993 ; pour revue, Reik et Walter, 2001).

Approximativement 90 gènes ont été décrits comme étant soumis à l'empreinte parentale (pour revue, Ideraabdullah *et al.*, 2008, pour la liste complète regarder <http://igc.otago.ac.nz/home.html> et [http://www.har.mrc.ac.uk/research/genomic imprinting/maps.html](http://www.har.mrc.ac.uk/research/genomic_imprinting/maps.html)) bien qu'il soit tout à fait probable que tous les gènes concernés n'aient pas encore été identifiés. L'une des caractéristiques de ces gènes est qu'ils sont bien souvent regroupés dans des "clusters" pouvant atteindre 1 Mb (pour revue, Morison et Reeve, 1998). Ces gènes sont exprimés soit à partir de l'allèle maternel, soit à partir de l'allèle paternel et sont généralement conjointement régulés par une région appelée "Imprinting Control region" (ICR). Cette région porte les modifications épigénétiques parentales (méthylation de l'ADN et modification des histones) qui gouvernent l'activité des "clusters" et est sans doute reconnue par des facteurs régulant l'expression du "cluster". De façon intéressante, la délétion de cette région conduit à la perte de l'empreinte parentale des multiples gènes du "cluster" (Fitzpatrick *et al.*, 2002 ; Lin *et al.*, 2003 ; Thorvaldsen *et al.*, 1998 ; Wutz *et al.*, 1997 ; Yang *et al.*, 1998). Les ICR peuvent être subdivisées en 2 catégories : celles dont la fonction est régulée par un insulateur (H19/Igf2) et celles qui servent de promoteur à un ARN non-codant



régulateur (pour revue, Ideraabdullah *et al.*, 2008). Nous nous focaliserons sur 2 exemples du 2<sup>ème</sup> cas qui correspond à notre thématique de recherche. Le mécanisme mis en évidence semble d'ailleurs être utilisé par la majorité des *loci* soumis à l'empreinte parentale. Ainsi, dans les paragraphes suivants nous passerons en revue les exemples les plus décrits d'empreinte parentale que sont le "cluster" *Igf2r* et le *locus Kcnq1*.

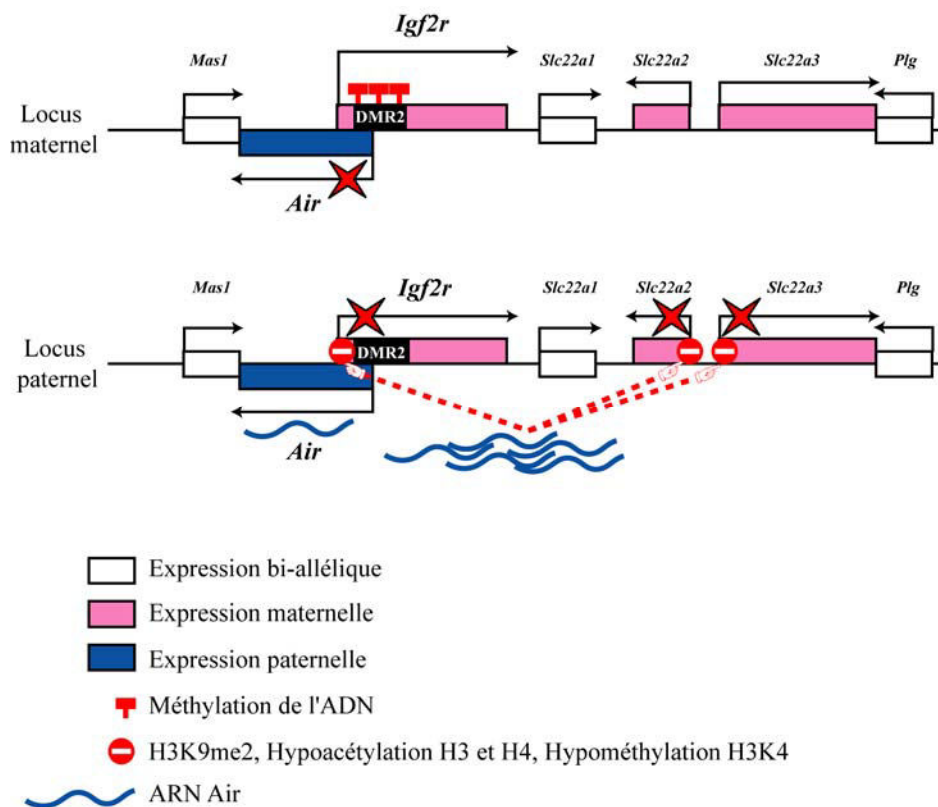
- ARN non-codant Air et "cluster" *Igf2r*

Le mannose 6-phosphate/insulin-like growth factor 2 receptor (M6P/*Igf2r*), appelé plus tardivement, simplement *Igf2r*, est une protéine multifonctionnelle codée par un grand gène d'environ 138 kb composé de 48 exons. Il est localisé sur le chromosome 17 chez la souris et 6 chez l'homme (pour revue, Vu *et al.*, 2006, pour revue, Ideraabdullah *et al.*, 2008). *Igf2r* fut le premier gène dont la régulation par empreinte parentale a été démontrée (Barlow *et al.*, 1991). Il fait partie d'un "cluster" de gènes comportant entre autres les gènes *slc22a2* et *slc22a3* situés au voisinage. Ces 3 gènes sont soumis à l'empreinte parentale dans tous les tissus adultes murins (exprimés à partir de l'allèle maternel), à l'exception du cerveau, dans lequel ils sont transcrits de façon bi-allélique (Wang *et al.*, 1994 ; Hu *et al.*, 1999). Leur expression dans les tissus humains est plus compliquée. En effet, ils ne seraient pas soumis à l'empreinte parentale dans les tissus adultes (Kalscheuer *et al.*, 1993), alors qu'il semble l'être dans certains tissus embryonnaires et extra-embryonnaires tels que le placenta (Xu *et al.*, 1993 ; Oudejans *et al.*, 2001).

L'expression de *Igf2r* murin soumise à l'empreinte maternelle dépend de la présence d'un îlot CpG et d'un ARN non-codant antisens appelé Air (Wutz *et al.*, 1997). En effet, la délétion de l'unité de transcription de cet ARN est suffisante pour altérer la régulation de l'empreinte sur l'ensemble du *locus* (Sleutels *et al.*, 2002). L'îlot CpG est situé dans l'intron 2 et comporte 113 pb (région 2). Il est méthylé sur l'allèle maternel et constitue la "région méthylée de façon différentielle 2" (DRM2) (**Fig. 6**). En d'autres termes, il s'agit d'une région dont le degré de méthylation diffère selon l'origine sexuelle, ce qui influence l'expression de l'ARN Air. Cet ARN antisens localisé dans l'intron 2 de *Igf2r* s'exprime exclusivement à partir de l'allèle paternel du fait de l'hypométhylation de la région DRM2 sur cet allèle (Lyle *et al.*, 2000 ; Wutz *et al.*, 1997 ; Stoger *et al.*, 1993). L'expression de l'ARN Air induit la répression en *cis* des gènes *Igf2r*, *Slc22a2* et *Slc22a3* avec la mise en place de modifications d'histones établissant la répression génique.

La présence de méthylations et d'acétylations a été mise en évidence par des expériences d'immunoprécipitation de la chromatine avec des anticorps dirigés contre ces modifications. Dans la plupart des tissus de souris à l'exception du cerveau, l'acétylation des histones caractéristique de la chromatine transcriptionnellement active est fortement enrichie dans la région promotrice du gène *IGF2R* d'origine maternel. La méthylation H3K4me

marque l'allèle actif et H3K9me2 marque l'allèle éteint. La tri-méthylation de H3K9 caractéristique de l'hétérochromatine constitutive est retrouvée spécifiquement au niveau de la DMR2 murine de l'allèle maternel à partir duquel *Air* est réprimé (Vu *et al.*, 2004) (**Fig. 6**). Le mécanisme par lequel les mises en place ciblées de marques épigénétiques qui régulent l'expression de *Air*, ainsi que le mode d'action de ce dernier dans la répression des gènes du "cluster" sont encore méconnus.



**Figure 6 : Modèle de régulation du "cluster" génique *Igf2r* soumis à empreinte parentale et régulé par l'ARN non-codant *Air***

L'ARN *Air* est exprimé à partir de l'allèle paternel et induit la répression en *cis* des gènes *Igf2r*, *Slc22a2* et *Slc22a3*. La présence d'*Air* est corrélée avec la mise en place de H3K9me2, l'hypométhylation de H3K4 et l'hypoacétylation de H3 et H4. Par contre, sur l'allèle maternel, la région ADN "DMR2" est méthylée et réprime l'expression de *Air*. Les gènes *Igf2r*, *Slc22a2* et *Slc22a3* sont exprimés.

- ARN non-codant et locus *Kcnq1*

Le locus *Kcnq1* contient un gène exprimé à partir de l'allèle paternel. Ce gène code un ARN non-codant appelé *Kcnq1ot1* (Mancini-Dinardo *et al.*, 2006 ; Mohammad *et al.*, 2008 ; Thakur *et al.*, 2004). Il comporte aussi plusieurs gènes, exprimés à partir de l'allèle maternel qui code des protéines comme *Cdkn1*, *Mash2*, *Phlda2* (Thorvaldsen et Bartolomei, 2007) (**Figure 7**). Ce locus est gouverné par la méthylation maternelle de l'ICR appelée *KvDMR1*.

Cette ICR est localisée dans l'intron 10 du gène *Kcnq1* (Lee *et al.*, 1999 ; Mitsuya *et al.*, 1999 ; Smilnich *et al.*, 1999) qui inclut aussi le promoteur du gène *Kcnq1ot1*. L'hypométhylation du promoteur de l'allèle paternel du gène *Kcnq1ot1* est associée à l'expression de l'ARN *Kcnq1ot1* et à la répression des gènes adjacents. Au contraire, l'hyperméthylation de KvDMR1 sur l'allèle maternel est associée à la répression de l'expression de l'ARN non-codant *Kcnq1ot1* et à l'expression des gènes adjacents soumis à l'empreinte parentale (Engemann *et al.*, 2000 ; Mancini-Dinardo *et al.*, 2006) (**Fig. 7**). Il a été démontré que la synthèse de cet ARN non-codant est corrélée avec le silence transcriptionnel de 8 des 10 gènes du "cluster" paternel, les gènes s'étendant sur une longueur de plus de 1 Mb (Fitzpatrick *et al.*, 2002 ; Mancini-Dinardo *et al.*, 2006). Malgré le fait que la délétion de *Kcnq1* ou de celle de son promoteur conduise à une dérégulation de l'empreinte parentelle (Fitzpatrick *et al.*, 2002 ; Mancini-Dinardo *et al.*, 2006 ; pour revue, Wu et Bernstein, 2008), le mode d'action précis de l'ARN *Kcnq1ot1* est encore mal connu.

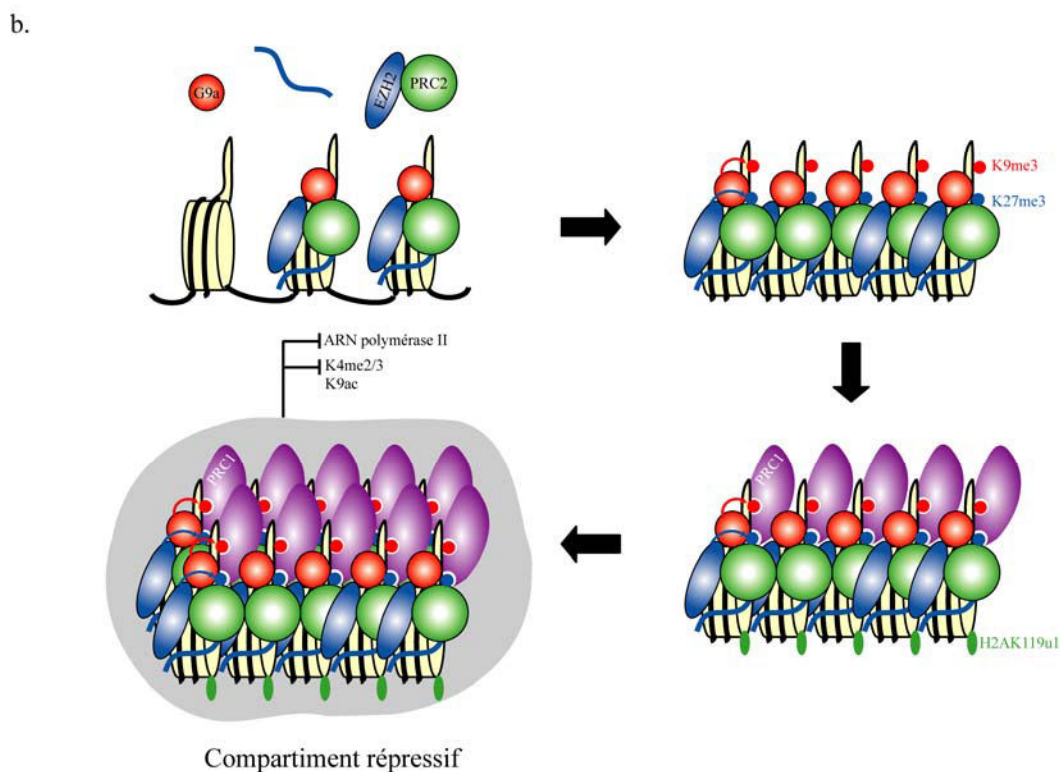
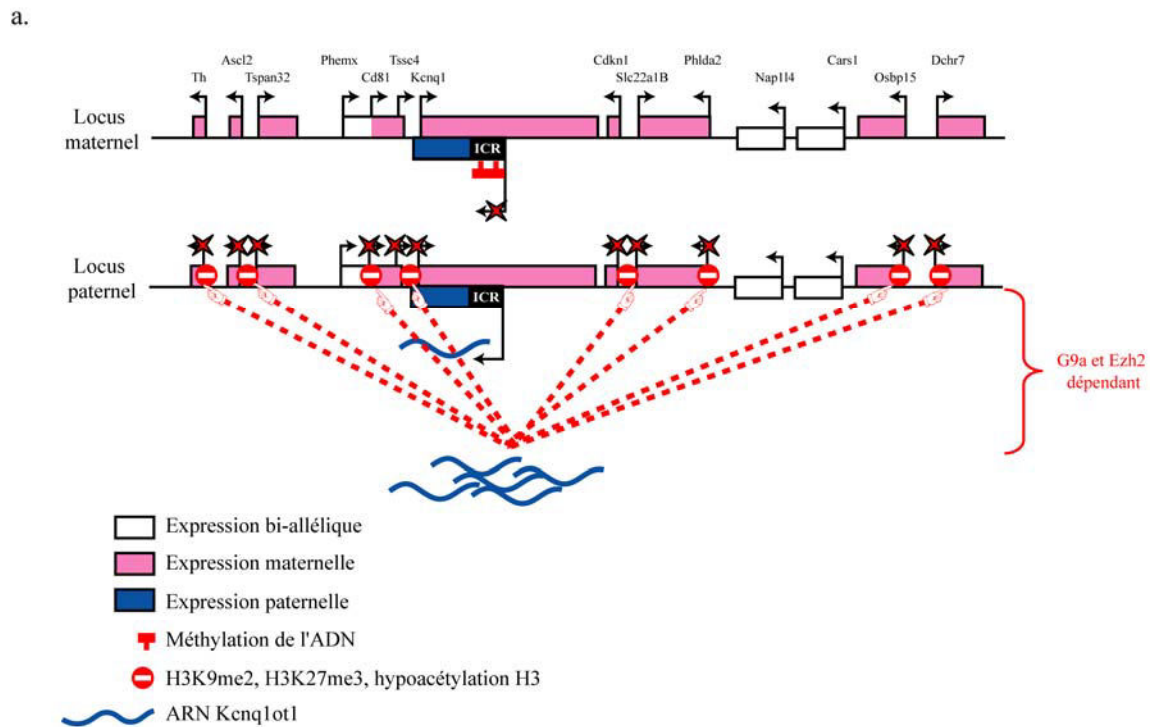
Durant l'année 2008, plusieurs travaux ont profondément modifié les connaissances sur cet ARN et ces nouvelles connaissances sont importantes pour l'interprétation des données sur l'inactivation du chromosome X. En effet, il est vraisemblable que les régulations épigénétiques liées à l'inactivation du chromosome X et celles impliquées dans la mise en place du silence transcriptionnel des "clusters" de gènes soumis à empreinte parentale par de longs ARN non-codants reposent sur des mécanismes similaires.

Récemment, Pandey *et al.* ont montré que l'ARN *Kcnq1ot1* est plus long que décrit précédemment (91,5 kb contre 60 kb, Mancini-Dinardo *et al.*, 2006). Il est transcrit par l'ARN pol II, retenu exclusivement dans le noyau et ne semble pas être épissé (Pandey *et al.*, 2008). Ces auteurs ont aussi mis en évidence le fait que l'ARN *Kcnq1ot1* lui-même, et non pas l'événement de transcription de cet ARN, joue un rôle dans l'extinction transcriptionnelle bidirectionnelle, ceci par son interaction avec la chromatine. De plus, cette interaction dépend de la lignée étudiée : elle est plus importante dans les tissus placentaires extra-embryonnaires, que dans les tissus hépatiques embryonnaires. Ces auteurs ont aussi montré que l'ARN *Kcnq1ot1* interagit avec certains membres du complexe PRC2 (Ezh2 et Suz12), ainsi qu'avec la protéine G9a HMT dans le placenta. Par des expériences d'hybridation *in situ* avec des sondes fluorescentes, ces auteurs ont montré que le domaine *Kcnq1ot1* d'un des deux chromosomes parentaux est fréquemment localisé au contact du territoire nucléolaire dans le placenta. Tous ces résultats concourent à envisager un mécanisme de répression transcriptionnelle spécifique de la lignée placentaire pour le domaine *Kcnq1ot1*, comme c'est le cas pour l'inactivation du chromosome X.

Les données que l'équipe de M Reik a obtenues par une analyse extrêmement fine de la localisation de certains gènes soumis à empreinte et de l'ARN *Kcnq1ot1*, renforcent l'idée de l'existence d'un domaine de répression nucléaire (Redrup *et al.*, 2009). Bien que l'absence

de la protéine Ezh2 ne semble avoir aucun effet sur la formation du complexe répressif nucléaire (Redrup *et al.*, 2009), des données obtenues par des expériences de CHIP montrent que l'extinction transcriptionnelle soumise à l'empreinte paternelle des gènes le long du "cluster" est bloquée en absence des protéines Ezh2 (PRC2) ou Rnf2 (PRC1) (Terranova *et al.*, 2008). Il a donc été suggéré que les protéines polycomb interviennent dans la propagation de l'ARN Kcnq1ot1 le long de la chromatine et/ou aient un rôle prédominant dans l'organisation chromatinienne (Terranova *et al.*, 2008).

En conclusion, il est pertinent de faire la comparaison avec les données concernant l'inactivation du chromosome X. En effet, dans les 2 cas, un ARN non-codant est essentiel à l'initiation de la répression transcriptionnelle. De multiples étapes sont impliquées dans l'établissement de la répression y compris le recrutement des complexes PRC1 et PRC2. Cette répression s'accompagne de la formation d'un compartiment nucléaire répressif excluant les facteurs de transcription (Chaumeil *et al.*, 2006).



### Figure 7 : ARN *Kcnq1ot1* et compaction de la chromatine du «cluster» *Kcnq1* par les protéines PcG

a. L'ARN *Kcnq1ot1* est exprimé à partir de l'allèle paternel. Il induit la répression en *cis* de la majorité des gènes du "cluster" sur cet allèle. Cette régulation requiert également les histones méthyl-transférases G9a et Ezh2. Par contre, sur l'allèle maternel, la région ADN "ICR" est méthylée et réprime l'expression de l'ARN *Kcnq1ot1*.

b. *Kcnq1ot1* recrute les protéines HMT G9a et Ezh2 ce qui permet respectivement la mise en place des modifications H3K9me2 et H3K27me3. La présence de ces méthylations permet le recrutement du complexe PRC1 catalysant l'ubiquitinylation des lysines 119 des histones H2A (H2AK119u1). Un domaine de *Kcnq1ot1* correspondant à un domaine répressif nucléaire s'organise, l'ARN Pol II en est ensuite exclue. Ce domaine est également globalement hypoacétylé et présente des H3K4 di- et tri-méthylées.

D'après la revue Wu et Bernstein, 2008

### 1.3.3 ARN interférence et hétérochromatine constitutive

Durant cette dernière décennie, le phénomène d'ARN interférence a été découvert et a été intensément étudié (pour revue, Carthew, 2001 ; pour revue, Tijsterman et Plasterk, 2004 ; pour revue, Meister et Tuschl, 2004 ; pour revue, Matzke et Birchler, 2005). Il fut détecté au départ chez les plantes et décrit comme un système de dégradation des ARNm par des petits ARN interférence. Chez tous les mammifères qui le possèdent, ce mécanisme repose sur la présence d'un grand ARN double-brin précurseur qui est découpé en fragments (siRNA pour Short-interference RNA) de 22 à 26 nucléotides par des ribonucléases de la famille *dicer*. Ces petits ARN db fixent les protéines du complexe RISC (RNA-induced silencing complex) et permettent le ciblage des ARNm, suivi de leur dégradation ou du blocage de leur traduction. Plus récemment, il a été montré que les siRNA peuvent aussi conduire à la répression génique par la formation d'hétérochromatine (pour revue, Meister et Tuschl, 2004). Un cas très bien décrit est celui de la formation d'hétérochromatine péri-centromérique chez la levure *S. Pombe*.

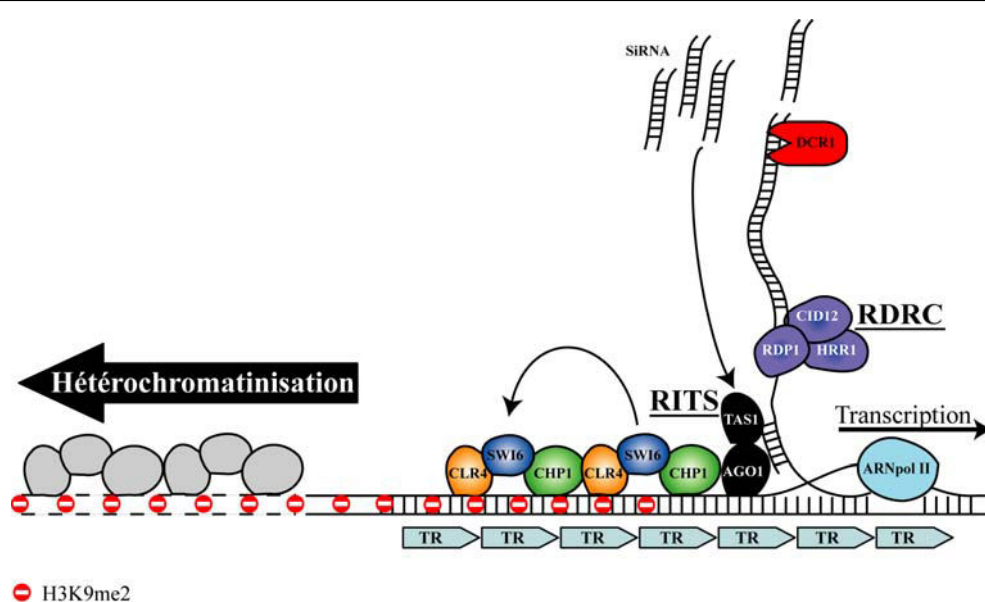
Pour faciliter la compréhension, nous débuterons par la présentation du processus permettant la formation d'hétérochromatine "classique" chez la levure. Tout d'abord, les histones sont désacétylés par les facteurs protéiques CLR3, CLR6 et SIR2. Les lysines des histones H3 sont ensuite méthylées en position 9 (H3K9me) par l'histone méthyltransférase CLR4 (Shankaranarayana *et al.*, 2003 ; pour revue, Grewal *et al.*, 1998 ; Nakayama *et al.*, 2001 ; Partridge *et al.*, 2000). La protéine SWI6 (équivalente à la protéine HP1 chez les mammifères) reconnaît les méthylations des histones et interagit avec la chromatine. Un complexe formé par les protéines SWI6, CHP1 et CRL4 est capable d'oligomériser et de se propager le long de la chromatine à partir de son point d'initiation. Ceci permet la formation d'hétérochromatine sur une région de plusieurs kb de long (Grewal et Moazed, 2003). Les gènes de cette région seront alors réprimés.

Qu'en est-il pour la formation d'hétérochromatine faisant appel aux composants de la machinerie du système d'ARN interférence et aux siRNA produits à partir des répétitions péri-centromériques. Par une approche biochimique, deux complexes protéiques ont été purifiés :

Le complexe RITS composé des protéines CHP (Chromodomain-containing protein), AGO1 (protéine argonaute 1) et TAS1 (dont la fonction n'est encore pas connue) (Verdel *et al.*, 2004). Ces composants sont associés aux régions péri-centromériques et cette localisation dépend du complexe DICER et des siRNA (**Fig. 8**).

Le complexe RDRC composé des protéines RDP1 (ARN polymérase ARN-dépendante), HRR1 (RNA hélicase), CID12 et la polyA polymérase (**Fig. 8**). Ce complexe présente une activité ARN polymérase ARN-dépendante et de façon similaire au complexe RISC est associé aux régions péri-centromériques. Cette interaction avec l'hétérochromatine dépend de

DICER, des composants du complexe RISC et de CLR4/SWI6. Et réciproquement, les mutations affectant le complexe RDRC bloquent la fixation du complexe RITS et la formation d'hétérochromatine. De plus, il a été suggéré que le complexe RISC n'interagit pas avec la molécule d'ADN mais avec des transcrits néosynthétisés (Motamedi *et al.*, 2004 ; Sugiyama *et al.*, 2005). D'autre part, il a été démontré qu'il existe un lien physique entre la production des siRNA et la machinerie transcriptionnelle (Schramke *et al.*, 2005). Des expériences de ciblage de gènes à inactiver (*ura4*) mettent en évidence que la répression n'a lieu que sur le gène à partir duquel les siRNA sont produits et n'a pas lieu sur un deuxième gène identique physiquement éloigné (Buhler *et al.*, 2006). Cette expérience a donc permis de confirmer le lien étroit entre les événements de transcription et de répression. Ainsi, le modèle actuellement admis est le suivant : l'ARN est produit à partir des répétitions péri-centromériques (TR), et il est converti sur place en ARN double-brin par le complexe RDRC (incluant les protéines HRR1, SID12 et RDP1). Les dsRNA sont clivés toujours sur place par DICER pour produire les siRNA. Les siRNA fixent ensuite le complexe RITS (incluant TAS1, CHP1 et AGO1) et les complexes RNP ainsi formés interagissent avec les transcrits des répétitions par complémentarité avec les siRNA. La formation d'hétérochromatine est alors engagée. La protéine CHP1 pourrait jouer un rôle important dans cette étape car elle est commune au complexe RITS et SWI6/CLR4 impliqués dans la formation d'hétérochromatine. L'hétérochromatine se propage à partir du point d'initiation le long du chromosome. CLR4 méthyle les lysines 9 des histones H3, qui seront alors reconnues par SWI6. Il semble que les composants des complexes RITS, RDRC et les facteurs d'hétérochromatinisation aient la capacité d'interagir pour former un complexe commun.



**Figure 8 : Formation d'hétérochromatine péri-centromérique médiée par les petits ARN chez la levure *S. pombe***

Les ARN sont produits à partir des répétitions en tandem centromériques (TR). Ils sont convertis en ARN double-brin par le complexe RDRC, puis sont clivés par DICER, ce qui aboutit à la production de siRNA. Ces petits ARN sont ensuite pris en charge par le complexe RITS qui induit la formation d'hétérochromatine. L'hétérochromatine se propage le long du chromosome à partir du point d'initiation.

D'après la revue Lavrov et Kibanov, 2007.

## 2. Notion de compensation du dosage des gènes

« Toute chose est poison et rien n'exclut le poison, seule la dose autorise quelque chose à ne pas être intoxicante ». Cette citation de Theophrastus Bombastus von Hohenheim (plus connu sous le nom de Paracelse, 1493-1541) décrit un principe fondamental valable non seulement dans son domaine qui était la toxicologie, mais également dans celui de la biologie. En effet, il est ainsi plus aisé de comprendre les raisons pour lesquelles une cellule doit garder un taux d'expression génique sous contrôle (pour revue, Bayer et Lee, 2008).

Chez de nombreuses espèces, les chromosomes sexuels présentent des différences morphologiques. Les individus homogamétiques (XX pour un organisme mammifère femelle) présentent deux chromosomes sexuels identiques, alors que les individus hétérogamétiques (XY pour un organisme mammifère mâle) présentent deux chromosomes sexuels distincts. La caractéristique chromosomique majeure d'un individu hétérogamétique est un nombre de gènes inférieur à celui présent dans un organisme homogamétique (pour revue, Marin *et al.*, 2008 ; pour revue, Charlesworth *et al.*, 2005). Donc, potentiellement, le taux d'expression des gènes liés aux chromosomes sexuels peut provoquer un déséquilibre entre d'une part les autosomes et les gonosomes et d'autre part entre les 2 sexes. Pour pallier ce problème, un mécanisme de compensation, appelé compensation de dose des gènes a été mis en place par sélection naturelle au cours d'une longue évolution. Nous décrirons comment et pourquoi les mécanismes de compensation du dosage des gènes ont été probablement établis durant l'évolution des chromosomes sexuels et la détermination sexuelle.

## 3. Evolution du déterminisme sexuel et compensation du dosage des gènes

La détermination du sexe dans le monde animal est très éclectique et emprunte différentes voies dépendantes de la constitution des chromosomes et des facteurs environnementaux. Chez les poissons et les reptiles, le sexe est déterminé par la température d'incubation de l'œuf après fertilisation (pour revue, Crews *et al.*, 2003). Les futurs individus sexuellement différents sont chromosomiquement identiques et cette forme de détermination n'a besoin d'aucun système de compensation du dosage des gènes. Cette détermination environnementale permet d'adapter au mieux la capacité reproductive d'une population. En contrepartie, celle-ci est particulièrement sensible aux changements environnementaux, tels qu'un réchauffement global. En accord avec cette théorie, les changements de climat auraient provoqué l'extinction des dinosaures, car ce système de détermination sexuelle aurait orienté

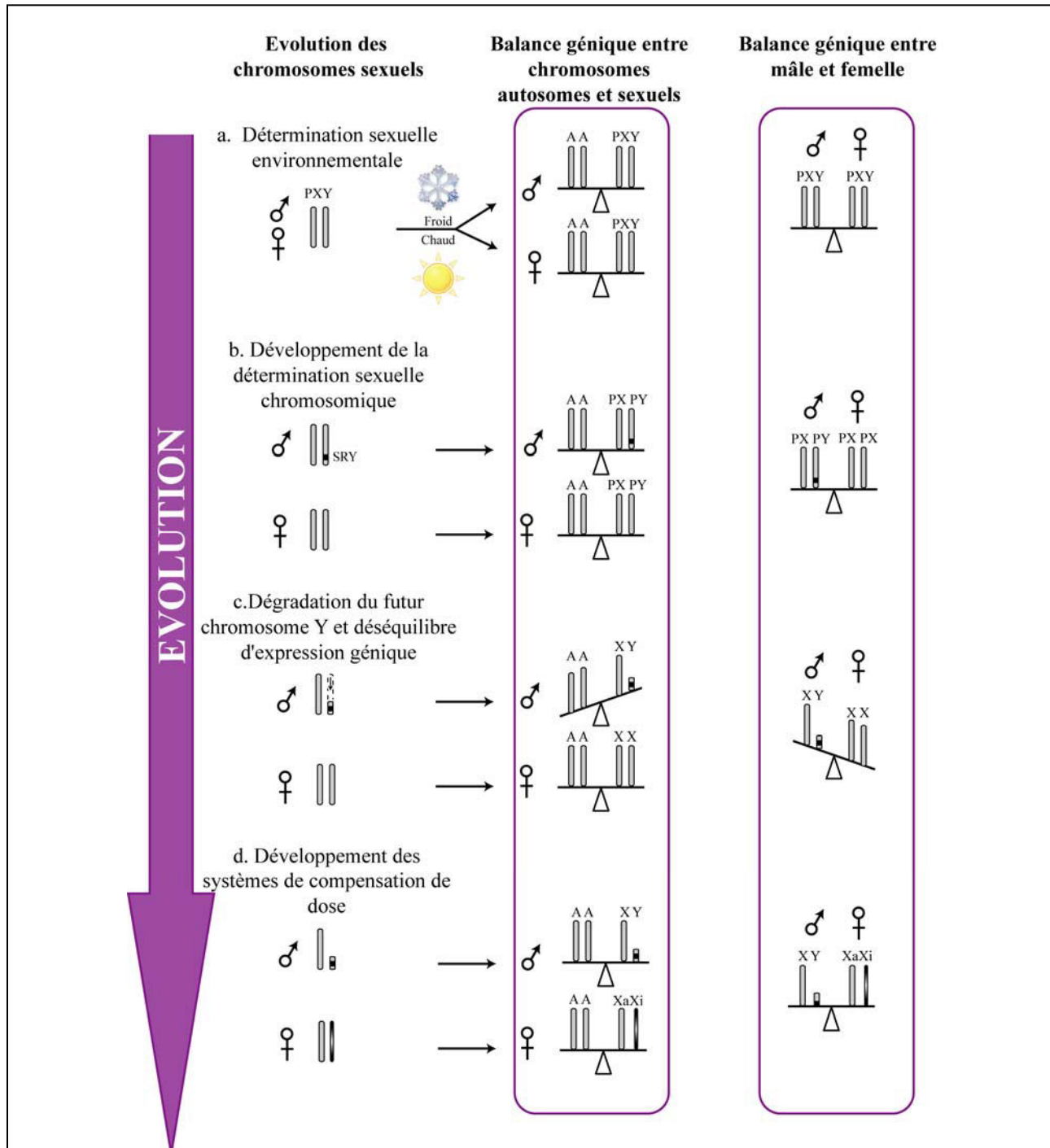


le typage sexuel vers une seule catégorie provoquant à terme l'incapacité de la population à se reproduire (pour revue, Miller *et al.*, 2004). En opposition avec cette détermination environnementale, il existe la détermination génétique ou chromosomique grâce à laquelle le caractère sexuel de l'individu adulte est prédéterminé lors de la fertilisation.

D'un point de vue évolutif, il a été proposé que les ancêtres des reptiles, des oiseaux et des mammifères utilisaient une détermination sexuelle basée sur les facteurs environnementaux comme énoncée précédemment (pour revue, Crews, 2003). Les chromosomes sexuels se seraient alors développés à partir des autosomes par acquisition de gènes déterminant le caractère sexuel ou par la mutation d'un des deux autosomes (Ohno, 1967, pour revue, Charlesworth, 1996 ; pour revue, Graves, 2006). Selon ce principe, deux systèmes de passage à un organisme hétérogamétique peuvent être recensés : le premier, adopté par les oiseaux et quelques reptiles comme les serpents, chez lesquels les femelles sont hétérogamétiques (ZW) alors que les mâles sont homogamétiques et ont 2 chromosomes Z (ZZ) ; le second, est basé sur les paires de chromosomes XY, système utilisé par les mammifères chez lesquels le mâle est hétérogamétique (XY) et la femelle homogamétique (XX). Au cours de l'évolution, un des deux chromosomes Z ou X a perdu un certain nombre de gènes. Ces chromosomes sont alors devenus respectivement, les chromosomes W et Y. Cette perte a eu pour conséquence l'apparition subite d'un problème de dosage de l'expression des gènes étant donné que leurs équivalents sont restés sur les chromosomes Z et X (**Fig. 9**). En effet, ces gènes sont présents seulement en une seule copie dans les organismes hétérogamétiques (ZW et XY). Leurs niveaux d'expression sont donc la moitié de ce qu'ils étaient avant la perte génique. Cela perturbe l'équilibre entre l'expression des gènes des chromosomes sexuels et l'expression des gènes autosomiaux, qui sont chacun présents en deux copies. De plus, les gènes des chromosomes sexuels sont présents en double chez les organismes homogamétiques (XX ou ZZ), ce qui provoque également un dosage déséquilibré de leur expression chez les deux sexes. Une des conséquences de l'évolution de la détermination sexuelle a été l'acquisition de mécanismes de compensation de dosage des gènes qui permettent de pallier le problème lié à la création des chromosomes sexuels hétéromorphiques (pour revue, Payer et Lee, 2008).

Généralement, la perte de fonction de certains gènes a été causée par la restriction du chromosome qui les contient à un seul sexe (souvent en raison de l'émergence d'un gène impliqué dans la détermination du sexe de façon dominante) et donc à l'impossibilité de recombinaison avec son homologue. Les mutations ne pouvant être éliminées par recombinaison, elles s'accumulent et conduisent à une perte progressive de la fonction génique. Ce processus est appelé dégénérescence chromosomique (pour revue, Charlesworth *et al.*, 2005 ; pour revue, Marin, 2008). Il conduit à une divergence rapide des chromosomes sexuels par rapport aux autres chromosomes. La part restant commune aux deux sexes est

appelée la région pseudoautosomale ou PAR. Elle peut donner lieu à des recombinaisons entre les différents chromosomes sexuels lors de la méiose. Par contre, les « crossing-over » entre les allèles sexe-spécifiques hors de la PAR n'étant pas possibles, les chromosomes W et Y ont accumulé des mutations et des délétions non réparables. Cela a conduit à une réduction de la taille de ces chromosomes (**Fig. 9**) (Graves, 2006 ; Ohno, 1967).



**Figure 9 : Schéma simplifié des relations entre évolution du déterminisme sexuel et compensation du dosage des gènes**

(a). Mâles et femelles ont deux chromosomes sexuels identiques et le sexe est déterminé par les conditions environnementales comme la température. Une paire d'autosomes est la forme ancestrale des futurs chromosomes sexuels (PXY ou Proto-XY).

(b). Acquisition d'un gène déterminant le sexe (SRY) en une seule copie sur la paire de proto-XY permettant de distinguer les deux chromosomes sexuels (PX ou Proto-X, PY ou Proto-Y). Cette acquisition conduit à un système de détermination chromosomique du sexe. Comme PX

et PY partagent toujours la plupart des gènes l'un avec l'autre, aucun déséquilibre de dosage de gène n'existe encore.

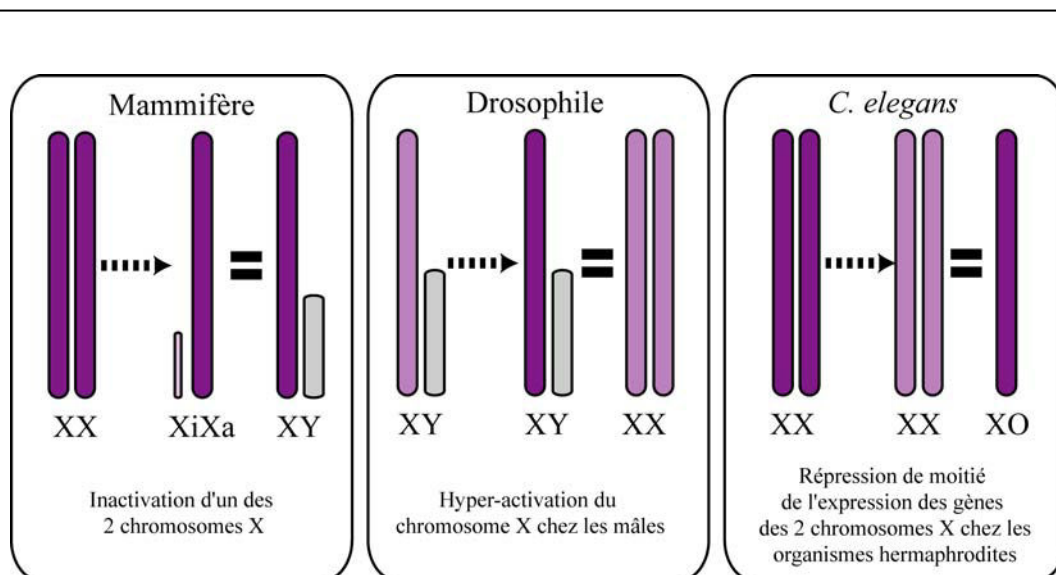
(c.) L'accumulation de gènes spécifiques aux mâles sur PY conduisant à une suppression de toutes recombinaisons méiotiques avec PX et à une dégénérescence du chromosome Y vers sa forme actuelle. Beaucoup de gènes liés au chromosome X sont présents en une seule copie chez les mâles, alors qu'ils étaient présents en deux copies sur les autosomes. (A.) Induction d'un déséquilibre de dosage des gènes entre le chromosome X et les autosomes. Un autre déséquilibre émerge entre les gènes liés au chromosome X chez les cellules mâles et femelles étant donné qu'elles comptent deux copies des gènes de l'X.

(d.) Pour contrecarrer ce déséquilibre, les mammifères ont développé 2 mécanismes distincts de compensation. Les gènes du Xa (X actif) sont surexprimés d'un facteur 2 par un système inconnu permettant de rétablir la balance entre le chromosome X et les autosomes chez les mâles. La surexpression des gènes du Xa chez les femelles est équilibrée par l'inactivation d'un des 2 chromosomes X (Xi)

D'après la revue Payer et Lee, 2008

#### 4. Compensation du dosage des gènes chez *Drosophila melanogaster* et *Caenorhabditis elegans*

Plusieurs systèmes de compensation du dosage des gènes ont été décrits chez plusieurs espèces. Alors que chez les mammifères, la compensation du dosage des gènes se fait par une inactivation d'un des 2 chromosomes X chez la femelle, chez les drosophiles mâles (*Drosophila melanogaster*), les gènes de l'unique chromosome X sont hyperactivés et chez le nématode hermaphrodite XX (*Caenorhabditis elegans*), une diminution de l'expression gènes des 2 chromosomes X est observée (**Fig. 10**).



**Figure 10 : Les diverses stratégies de compensation du dosage des gènes**

Suivant les organismes, différentes stratégies sont utilisées pour équilibrer l'expression des gènes liés au chromosome X entre mâles (XY ou XO) et femelles (ou hermaphrodites, XX).

La compensation du dosage se fait chez les mammifères par une inactivation d'un des 2 chromosomes X chez la femelle, par une hyperactivation des gènes de l'unique chromosome X dans les cellules de drosophile mâle et par une diminution de moitié de l'expression des gènes des 2 chromosomes X chez *C. elegans*.

L'intensité de la couleur est proportionnelle à l'expression génique.

D'après la revue Meyer, 2005

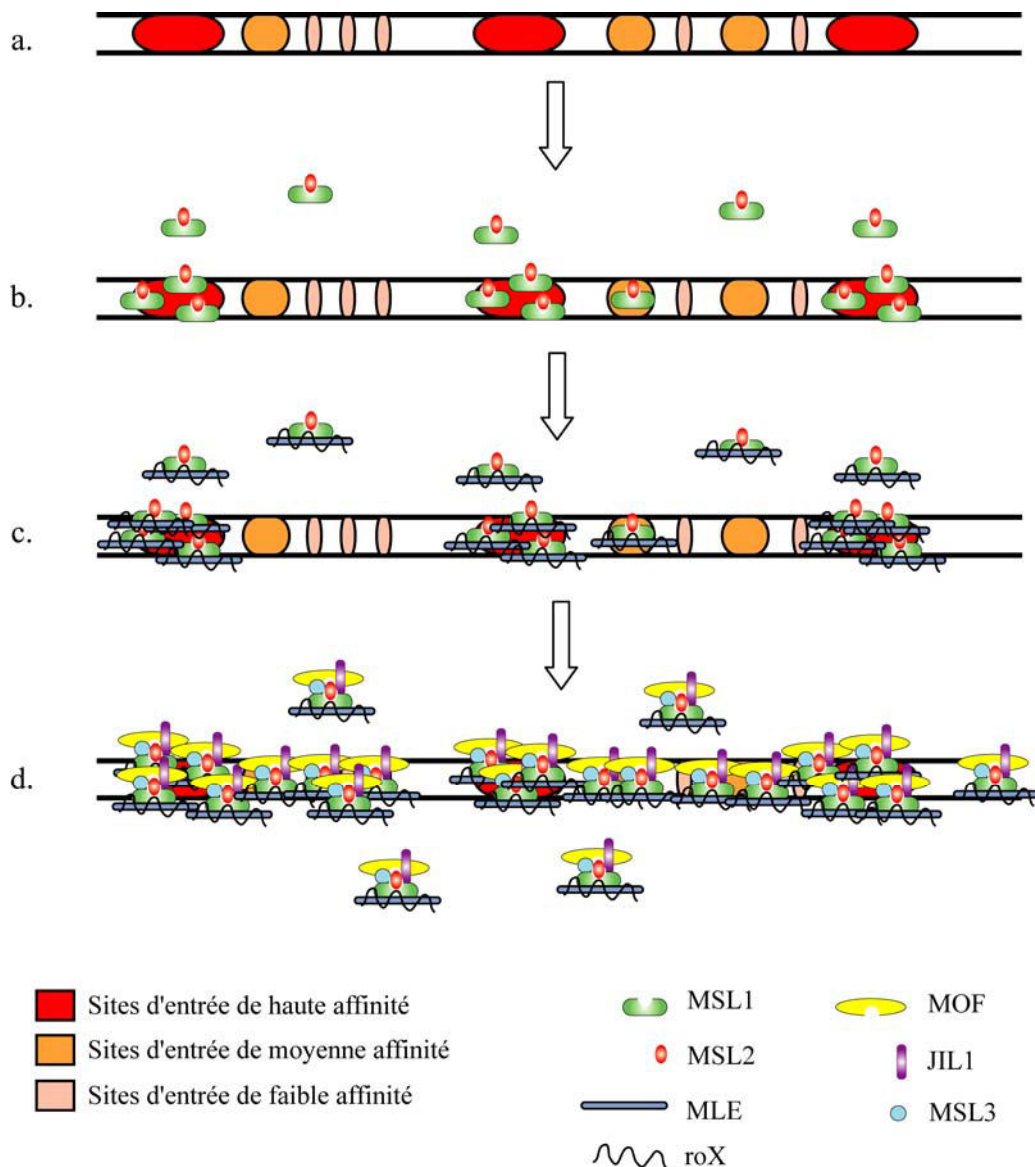
#### 4.1 Compensation de dosage des gènes chez *Drosophila Melanogaster*

Chez *D. melanogaster*, comme chez les mammifères, les chromosomes sexuels sont dimorphiques, et les femelles sont caractérisées par la présence de deux chromosomes X, tandis que les mâles n'ont qu'un seul chromosome X, accompagné d'un chromosome Y.

Par contre, le mécanisme de compensation de dosage chez la drosophile est différent de celui acquis par les mammifères. Les gènes portés par l'unique chromosome X des mâles ont une expression doublée chez les mâles grâce à l'expression spécifique dans les tissus somatiques mâles, de la protéine MSL2 (Hamada *et al.*, 2005 ; Straub *et al.*, 2005). Cette protéine est une composante essentielle du complexe de compensation de dosage (Dosage Compensation Complex, DCC ou compensasome). Ce complexe DCC est constitué de 6 protéines partenaires dont MLE (maleless) (Kuroda *et al.*, 1991), MSL1 (male-specific lethal) (Scott *et al.*, 2000), MSL2 (Copps *et al.*, 1998), MSL3, MOF (males absent on the first) (Gu *et al.*, 1998) et JIL1 (kinase). L'association de ces facteurs protéiques avec 2 ARN non codants appelés roX1 et roX2 (RNA on the X) (Kageyama *et al.*, 2001 ; Meller *et al.*, 1998 ; Jin *et al.*, 2000), conduit à la formation d'un complexe actif localisé sur le chromosome X. La fixation de ce complexe induit un enrichissement en acétylation de l'histone H4 en position 16 (H4Ac16) et un changement structural de la chromatine. L'expression des gènes liés au chromosome X chez le mâle est alors hyperactivée. Chez la femelle, la protéine SXL est exprimée. Elle dérégule la traduction des transcrits MSL2 ce qui empêche la formation d'un DCC fonctionnel (Palmer *et al.*, 1994 ; Beckmann *et al.*, 2005). La mise en place de la compensation de dosage chez les drosophiles mâles grâce au complexe DCC a été étudiée en détail. Il s'agit d'un mécanisme ordonné débutant par la fixation des protéines MSL1 et MSL2 (corps du DCC) sur le chromosome X au niveau de 35 sites appelés sites d'entrée de haute affinité. Les ARN roX sont ensuite recrutés sur le DCC naissant et les complexes ribonucléoprotéiques sont stabilisés par la protéine Mle (Kageyama *et al.*, 2001; Meller *et al.*, 2002). Le compensasome ne devient complet qu'après le recrutement des protéines MSL3, MOF et JIL1 kinase. MSL3 est connue pour être capable d'interagir avec la chromatine, alors que MOF porte une activité acétyltransférase spécifique des histones H4 (Smith *et al.*, 2000). Lorsque le complexe DCC est complet, il a la capacité d'interagir avec de nombreux sites de moindre affinité localisés sur l'ensemble du chromosome X et l'hyperactivation transcriptionnelle du chromosome devient homogène (**Fig. 11**).

Notons qu'une fois encore la compensation du dosage est médiée par des ARN, ce qui met en évidence leur rôle important dans la maintenance de la chromatine et l'expression des gènes. L'expression des ARN roX1 et roX2 est elle-même activée par le DCC chez les mâles, ce qui suggère une complexité supplémentaire liée à ce mécanisme de « feed-back » (Bai *et al.*, 2004 ; Rattner et Meller, 2004 ; Straub *et al.*, 2005). Il reste encore à savoir comment la

protéine SXL contrôle l'expression de MSL2 et comment son absence empêche la formation du DCC.



### Figure 11 : Compensation du dosage des gènes chez la drosophile

Chez la drosophile, le chromosome renferme plusieurs sites d'entrées ou site de fixation du complexe ribonucléoprotéique de compensation DCC

a. Lorsque les protéines corps MSL1 et MSL2 du DCC sont produites, elles interagissent avec les sites de plus haute affinité.

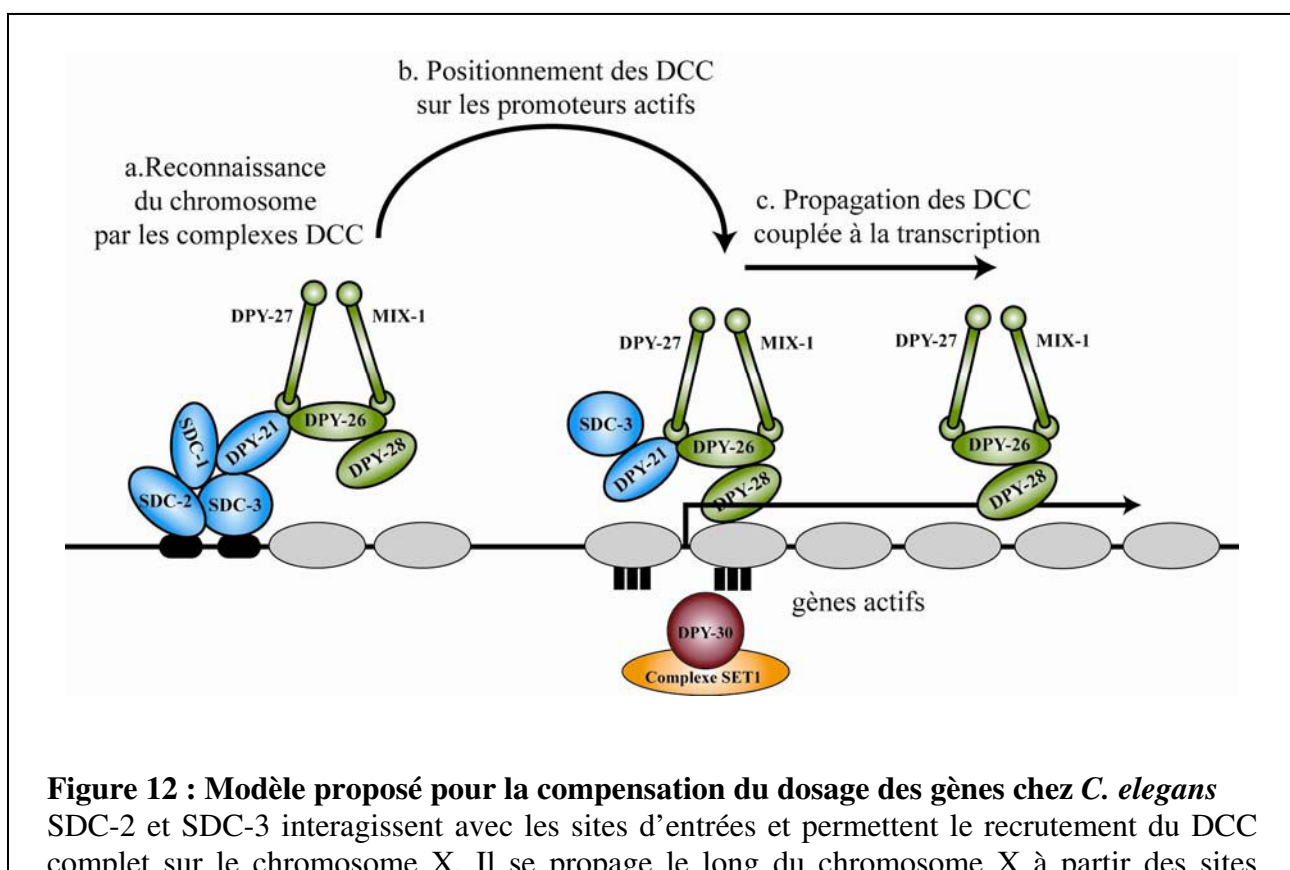
b. L'addition des facteurs MLE et des 2 ARN roXs permet l'amplification de la reconnaissance des sites.

c. et d. Seuls les complexes entiers (addition des facteurs MOF, JIL1 et MSL3) sont capables de se propager le long du chromosome entier, ceci par la fixation des sites de plus basse affinité.

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

4.2 Compensation de dosage des gènes chez *C. elegans*

*C. elegans* est un organisme particulier. Il existe deux types d'organismes : les hermaphrodites ayant un génotype XX et les mâles avec un génotype XO. Etant donné ce caractère pseudo-dimorphique, *C. elegans* fait appel à un système de compensation du dosage des gènes. À la différence de la drosophile, ce système correspond à la réduction de moitié de la transcription des gènes liés aux chromosomes X dans les organismes XX. Ce processus met en jeu 8 facteurs protéiques codés par les gènes *sdc-1*, *sdc-2*, *sdc-3* (sex determination et dosage compensation defect), *dpy-21*, *dpy-26*, *dpy-27*, *dpy-28* (dumpy) et *mix-1* (mitosis et X associated) (Hodgkin et Brenner, 1977 ; Hodgkin, 1980 ; Trent *et al.*, 1983 ; Meneely et Wood, 1987 ; Meyer et Casson, 1986 ; Villeneuve et Meyer, 1987 ; Nusbaum et Meyer, 1989 ; Plenefisch *et al.*, 1989 ; Lieb *et al.*, 1998) (**Fig. 12**). Ces protéines sont localisées sur les chromosomes X d'organismes hermaphrodites et leurs mutations augmentent leur létalité. Ce complexe protéique reconnaît 55 séquences d'ADN composées d'éléments répétés de 10 pb (McDonel *et al.*, 2006 ; Ercan *et al.*, 2007). A partir de ces sites d'entrée, les complexes DCC se propagent et induisent la méthylation de H3K4, ce qui réduit les capacités transcriptionnelles de la chromatine (Hsu et Meyer, 1994 ; Nagy *et al.*, 2002, pour revue, Ercan et Lieb, 2009). Chez les organismes mâles, le facteur SDC-2 n'est pas exprimé, ce qui rend le complexe DCC incompetent et n'influence pas l'expression génique de l'unique chromosome X. Le mécanisme par lequel l'activation n'influence que de moitié l'expression des gènes est encore non connu.



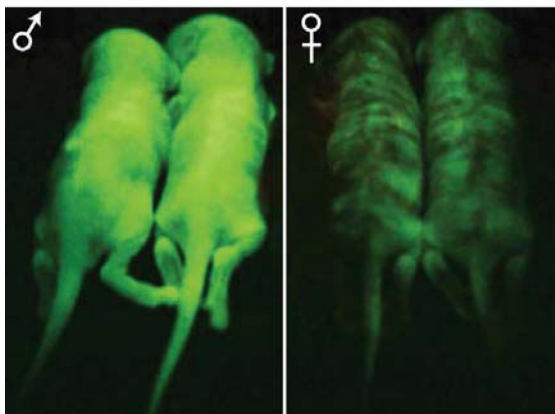


d'entrée vers les promoteurs actifs portant une marque reconnue par le DCC et permet son recrutement. Cette propagation semble être couplée à la transcription. Ensuite, le facteur DPY-30 participe probablement à la méthylation des histones H3K4, modifications caractéristiques des gènes liés au chromosome X chez le nématode. D'après la revue Ercan et Lieb, 2009

## B. Inactivation du chromosome X chez les mammifères

### 1. Principes généraux

C'est en 1961 que les bases théoriques de l'inactivation du chromosome X ont été posées par M Lyon (Lyon, 1961). Chez la souris, d'après des analyses phénotypiques de l'effet de mutations dans des gènes liés au chromosome X [*tabby* (Falconer, 1953) et *mottled* (Fraser *et al.*, 1953)], et qui régissent la couleur du pelage, M Lyon a observé que chez les femelles hétérozygotes, les mutations conduisent à la formation d'un pelage bigarré (*mottled* en anglais) consistant en la juxtaposition de zones de pelage de phénotype sauvage et de zones de phénotype mutant, tandis que chez les mâles porteurs de l'une ou l'autre de ces mutations, le pelage est toujours uniforme. Le même phénomène a été à nouveau observé par M Lyon pour un troisième mutant de gène lié au chromosome X (Lyon, 1961). En parallèle, l'apparition d'un mosaïcisme a été observé pour des translocations impliquant le chromosome X (Russell et Bangham, 1961 ; Russell Bangham, 1959). La même année des souris femelles aneuploïdes (XO) viables et fertiles ont été mises en évidence, ce qui a révélé d'une part que la détermination du sexe mâle dépend de la présence du chromosome Y (et non pas du nombre de chromosomes X comme chez la drosophile), d'autre part qu'un chromosome X unique suffit pour que le développement soit normal (Welshons et Russell, 1959). Pour illustrer ces observations phénotypiques, la **Figure 13** présente des souris transgéniques portant sur un chromosome X le gène codant la protéine fluorescente EGFP (Hadjantonakis *et al.*, 2001 ; pour revue, Hadjantonakis *et al.*, 2003).

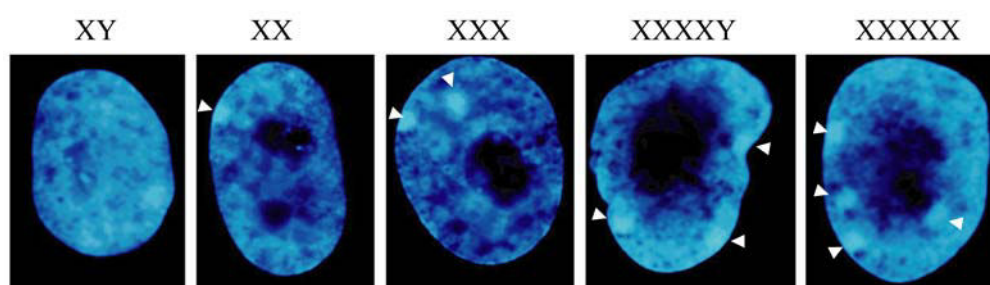


**Figure 13 : Illustration phénotypique de l'inactivation du chromosome X**

Les souris mâles ont une couleur de peau fluorescente, forte et unie qui témoigne de l'expression du gène EGFP de l'unique chromosome X présent dans chaque cellule. Par contre, les femelles arborent une fluorescence en mosaïque. Ce phénotype illustre l'inactivation aléatoire d'un des deux chromosomes X.

D'après Hadjantonakis *et al.*, 2001

M Lyon a alors proposé 1) l'existence d'une inactivation d'un des deux chromosomes X chez les femelles à un stade précoce du développement embryonnaire ; 2) un choix aléatoire du chromosome X maternel ( $X_m$ ) ou paternel ( $X_p$ ) inactivé ; 3) une stabilité de l'inactivation établie lors des divisions somatiques (Lyon, 1961 ; Lyon, 1962). Ce modèle rendait compte de l'effet clonal observé pour les mutations ou translocations liées au chromosome X. Il faisait également le lien avec des observations cytogénétiques antérieures. Ainsi, Barr avait décrit en 1949 la présence d'une structure condensée, associée à la membrane nucléaire, structure qui est révélée par la coloration de Nissl, dans des noyaux interphasiques de motoneurons de chattes (Barr et Bertram, 1949). Cette « chromatine sexuelle », dès lors appelée corpuscule de Barr, fut rapidement utilisée comme un moyen simple de détermination du sexe sur des préparations histologiques, avant d'être considéré comme un état hétérochromatique de l'un des deux chromosomes X (Ohno *et al.*, 1959 ; Ohno et Hauschka, 1960). Elle peut être observée dès le stade blastocyste tardif dans de nombreuses espèces, comme le chat, le singe et l'homme (Austin et Amoroso, 1957 ; Park, 1957), ce qui est en accord avec le modèle proposé par M Lyon. Plus tard, la caractérisation de nombreuses marques épigénétiques associées au domaine formé par le chromosome X inactif a permis de mieux définir ce corpuscule et permis de définir l'étape de l'embryogenèse où l'inactivation est mise en place. Le modèle proposé par M Lyon fut ensuite rapidement affiné pour tenir compte de nouvelles observations. Il fut en effet montré que des individus multisomiques pour le chromosome X, mais chez qui le nombre d'autosomes reste inchangé (AA:XXY; AA:XXXX; AA:XXXXY) (**Fig. 14**), n'ont qu'un seul X actif, les autres étant condensés (Giannelli, 1963 ; Grumbach *et al.*, 1963).



#### Figure 14 : Morphologie du chromosome X inactif

Plusieurs lignées de fibroblastes primaires ont été marquées au DAPI (46 XY, 46 XX, 47 XXX, 49 XXXXY et 49 XXXXX) et ont été analysées par microscopie électronique. Les images des noyaux analysés sont montrées. Les chromosomes inactifs apparaissent sous une forme dense et compacte au voisinage de la membrane nucléaire et sont indiqués par des flèches blanches. Le nombre de chromosomes X inactivés répond à la loi "n-1", dans laquelle n est égal au nombre de chromosomes X présents dans la cellule.

D'après Chadwick et Willard, 2001



Ainsi, un seul chromosome X doit être maintenu actif pour un lot d'autosomes, ce qui introduisait une notion de comptage (Lyon, 1962 ; Ohno *et al.*, 1964). L'inactivation du chromosome X, telle que nous venons de la décrire, s'applique essentiellement aux euthériens. En effet, chez les marsupiaux, l'inactivation est partielle, variable selon les tissus, et ne concerne que le chromosome X transmis par le père (Cooper *et al.*, 1993 ; Wakefield *et al.*, 1997). Les bases de l'inactivation étant ainsi posées, il restait à en comprendre les mécanismes moléculaires impliqués, aussi bien ceux permettant l'hétérochromatinisation et l'extinction transcriptionnelle d'un chromosome entier, que ceux contrôlant la mise en place de l'inactivation de manière opportune, au cours du développement embryonnaire. La caractérisation du centre d'inactivation du chromosome X, puis de l'ARN *Xist*, dans les années quatre-vingt dix, puis l'essor de l'épigénétique dans les années 2000, ont refaçonné la vision du phénomène d'inactivation sans pour autant la remettre en cause.

## **2. Inactivation du chromosome X au cours du développement**

### **2.1 Les étapes de l'embryogenèse chez la souris**

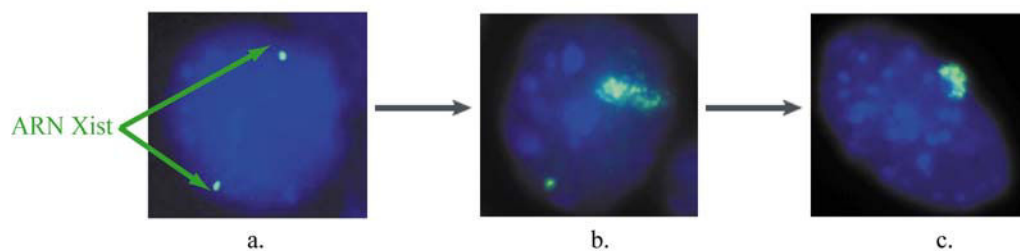
L'inactivation du chromosome X est un processus qui se déroule très tôt lors du développement embryonnaire. Nous rappellerons donc brièvement dans ce paragraphe les premières étapes du développement chez la souris.

Le développement préimplantatoire commence avec la fertilisation de l'ovocyte par le spermatozoïde. Rapidement après l'entrée du spermatozoïde, les pronuclei mâle et femelle migrent au centre de l'oocyte. Après 18 à 21 heures, les membranes pronucléaires se rompent et les deux pronucléi fusionnent pour former le zygote (œuf fécondé) réunissant le matériel génétique femelle et mâle. Par division cellulaire, deux blastomères apparaissent rapidement après la fécondation au premier jour du développement, puis successivement 4, 8 et 16 blastomères. Ce stade est appelé morula (l'œuf ressemble à ce stade à une petite mûre). Toutes ces divisions provoquent une compaction des cellules, décrite pour la première fois par W. H. Lewis et E. S. Wright en 1935. Les premiers événements de différenciation cellulaire induisent une nouvelle répartition des cellules dans la morula, stade nommé blastocyste. Les cellules périphériques subissent ensuite un phénomène de polarisation et se répartissent en une couche nommée trophoblaste primitif qui entoure toute la surface de l'œuf fécondé. Les cellules plus internes et initialement non polarisées se regroupent pour constituer la masse de l'embryoblaste, ou masse cellulaire interne (Inner Cell Mass, ICM). Ainsi, à 3,5 jours après fécondation (3,5E), le blastocyste est formé et contient 2 populations cellulaires distinctes : le trophectoderme (TE qui constituera les tissus extra-embryonnaires) et les cellules de la masse interne (ICM qui constitueront les tissus embryonnaires) (pour revue, Downs, 2001). Nous allons voir que lors de ces étapes de développement se succèdent des phases d'inactivation et de réactivation du chromosome X.

## 2.2 L'ARN Xist

La propriété majeure du chromosome X inactif (Xi) est le fait qu'il est recouvert par un ARN non-codant. Nous verrons d'ailleurs que celui-ci est l'élément moteur de l'initiation de l'inactivation. Cet ARN est appelé Xist (*X inactive specific transcript*), il est codé par le gène du même nom situé sur chaque chromosome X et ne s'exprime qu'à partir du chromosome X inactif (Xi) (Brockdorff *et al.*, 1992).

Son profil d'expression particulier a suggéré rapidement une implication potentielle dans l'inactivation du chromosome X (Penny *et al.*, 1996). Par exemple, les cellules ES de souris femelles indifférenciées (modèle cellulaire qui permet de récapituler les étapes de l'inactivation du chromosome X) présentent deux chromosomes X actifs. Ils expriment l'ARN Xist à un très faible niveau, qui est visualisé par l'approche RNA-FISH sous la forme d'un signal ponctuel (**Fig. 15**). Lors de la différenciation, un des deux chromosomes X est inactivé de manière aléatoire chez les femelles, mais pas chez les mâles qui gardent leur unique chromosome X actif (Rastan et Robertson, 1985). Plusieurs laboratoires ont montré que la différenciation des cellules ES femelles est accompagnée d'un fort renforcement de l'expression d'un des deux allèles de Xist, ce qui permet à cet ARN de recouvrir le chromosome X à partir duquel il est transcrit (**Fig. 15**). C'est sur le chromosome recouvert par l'ARN Xist que l'inactivation génique se met en place (Xi). Sur l'autre chromosome (Xa), ainsi que sur l'unique chromosome X des mâles, l'expression de Xist est complètement éteinte dans des stades plus tardifs de la différenciation.



**Figure 15 : Evolution de la quantité et de la localisation de l'ARN Xist au cours de la différenciation des cellules ES**

L'ARN Xist a été détecté par FISH grâce à l'emploi d'une sonde spécifique marquée par un fluorochrome (green dUTP).

a. Les cellules ES indifférenciées présentent des signaux qui mettent en évidence l'expression de l'ARN Xist sur les deux chromosomes X.

b. Les cellules ES en début de différenciation présentent une accumulation de Xist sur un des deux chromosomes X. Le chromosome X non inactivé continue à exprimer Xist sous sa forme "instable".

c. Dans les cellules ES différenciées, l'ARN Xist habille le chromosome inactif et l'expression du gène Xist est réprimée sur le chromosome actif.

D'après la revue Avner et Heard, 2001

### 2.3 Dynamique d'expression du gène *Xist* : inactivation soumise à empreinte parentale et inactivation aléatoire

Nous venons de voir que l'expression de l'ARN *Xist* est un élément important dans la mise en place de l'inactivation. Dans ce paragraphe, nous présenterons les régulations de l'expression de l'ARN *Xist* et nous décrirons leur caractère dynamique régissant plusieurs vagues d'inactivation du chromosome X au cours du développement embryonnaire.

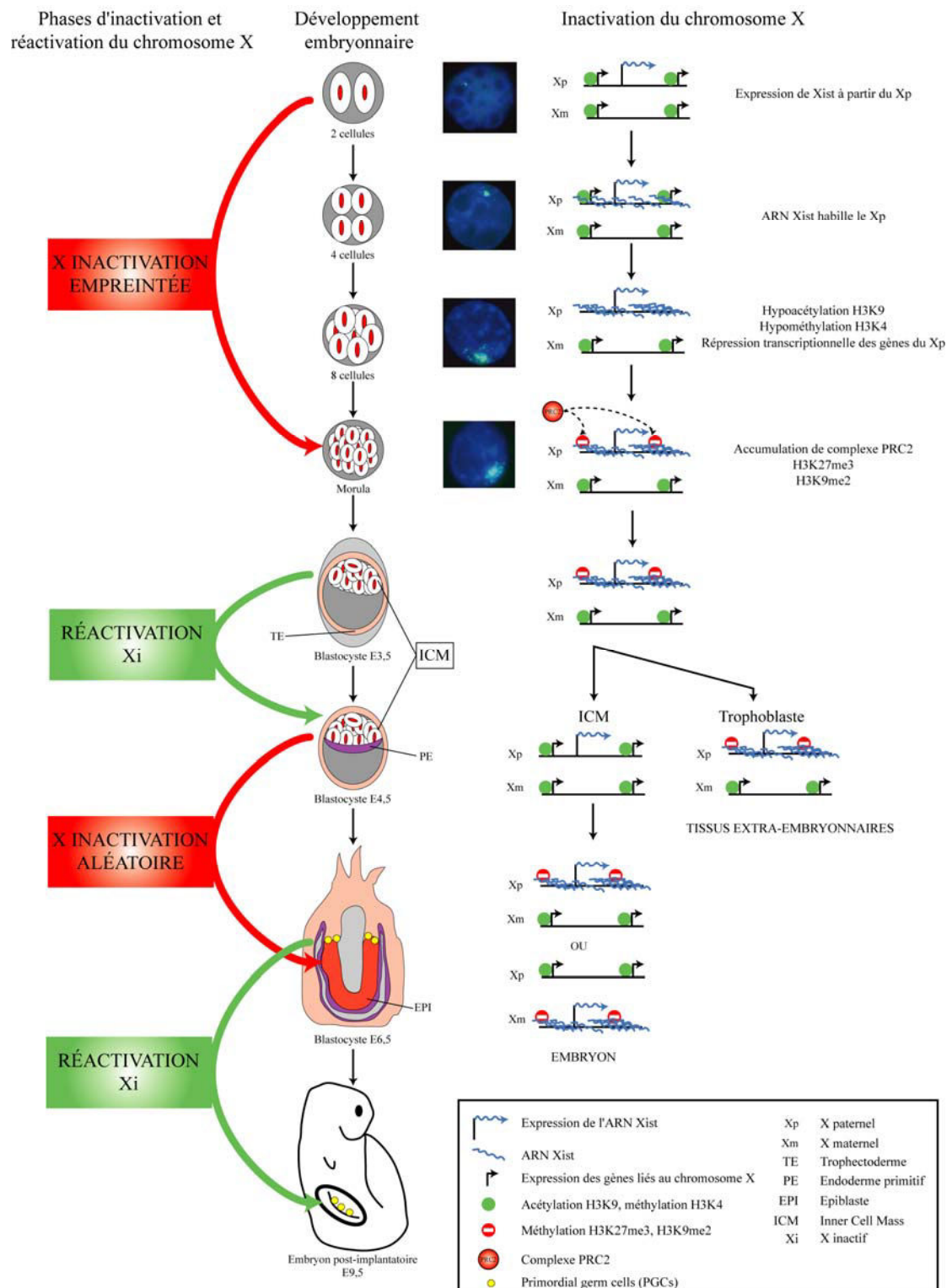
En effet, chez les organismes mammifères avec ou sans placenta, il est admis que deux types d'inactivation du chromosome X peuvent avoir lieu. L'inactivation soumise à l'empreinte parentale qui tend toujours à éteindre le chromosome X d'origine paternelle et ceci dans les étapes les plus précoces du développement. Cette inactivation perdure dans les tissus extra-embryonnaires tout au moins chez la souris et la vache (Dindot *et al.*, 2004 ; Xue *et al.*, 2002 ; Cooper *et al.*, 1993 ; Sharman, 1971 ; Takagi et Sasaki, 1975 ; West *et al.*, 1977). Par contre, plus tardivement au cours du développement, un autre processus (consécutif à la réactivation du chromosome paternel éteint précédemment) établit une inactivation aléatoire d'un des deux chromosomes X, ceci indépendamment de l'origine sexuelle du chromosome X inactivé. Ces événements tardifs ont lieu dans les cellules pluripotentes de la masse interne de cellules. Cette inactivation est stable, car elle est transmissible aux cellules filles et l'identité du chromosome inactif est figée.

Cette succession d'extinctions et d'activations transcriptionnelles des chromosomes X au cours du développement embryonnaire s'accompagne d'une expression différentielle de l'ARN *Xist* qui joue un rôle clé aussi bien dans l'inactivation du chromosome X soumise à l'empreinte parentale, que dans l'inactivation aléatoire (Huynh et Lee, 2003 ; Mak *et al.*, 2004; Okamoto *et al.*, 2004).

Dès le stade 2-cellules, l'ARN *Xist* est détecté par RNA-FISH sous la forme d'une accumulation de transcrits naissants. A ce stade, l'ARN *Xist* ne recouvre pas encore le chromosome X paternel (Okamoto *et al.*, 2004 ; Patrat *et al.*, 2009) (Xp par la suite ; Xm pour celui d'origine maternel). L'analyse de l'expression de 18 gènes liés au chromosome X à ce stade, met en évidence un activité transcriptionnelle identique sur les 2 chromosomes X, ce qui suggère que le Xp est actif avant la mise en place de l'inactivation soumise à l'empreinte parentale (Patrat *et al.*, 2009) (**Fig. 16**). Dès le stade 4-cellules, l'inactivation affecte systématiquement le Xp. L'analyse de l'expression du gène *Xist* montre qu'il est exprimé de manière monoallélique, uniquement à partir du Xp (Kay *et al.*, 1994 ; Matsui *et al.*, 2001 ; Nesterova *et al.*, 2001a ; Okamoto *et al.*, 2000 ; Zuccotti *et al.*, 2002), tandis que l'expression de l'allèle maternel du gène *Xist* est réprimée jusqu'au stade morula inclus (Nesterova *et al.*, 2001). L'expression de l'ARN *Xist* est accompagnée de la mise en place de marques épigénétiques qui concourent à la répression génique. Elles seront abordées en détails par la suite.

Alors que dans les tissus extra-embryonnaires le Xp reste inactif (trophectoderme par exemple), il est réactivé entre 3,5E et 4,5E dans les cellules de l'embryon (Huynh et Lee, 2003 ; Mak *et al.*, 2004 ; Okamoto *et al.*, 2004 ; Okamoto *et al.*, 2005). La perte de l'accumulation de l'ARN Xist est alors observée. Consécutivement à cette perte, il est admis que les deux chromosomes X deviennent vierges de toutes traces épigénétiques ce qui rend possible le passage de l'inactivation soumise à empreinte à un mode aléatoire d'inactivation (Navarro *et al.*, 2005).

Une nouvelle vague d'inactivation s'opère ensuite au niveau des cellules de la masse interne de cellule (ICM) au stade blastocyste en ciblant aléatoirement le Xp ou le Xm par une sur-expression mono-allélique du gène *Xist*. Cette inactivation appelée inactivation aléatoire se met en place à un stade compris entre 5,5E et 6,75E (Sugimoto *et al.*, 2000 ; Hadjantonakis *et al.*, 2001). L'état inactif du chromosome X est ensuite maintenu de façon clonale tout au long de l'embryogenèse pour n'être réversé que dans la lignée germinale primordiale femelle (PGC). En effet, durant la maturation des PGC, la majeure partie de la chromatine est reprogrammée à 13,5E pour devenir à nouveau pluripotente (pour revue, Senner et Brockdorff, 2009 ; Hajkova *et al.*, 2008, Johnston, 1981 ; Monk et McLaren, 1981 ; Tam *et al.*, 1994 ; Nesterova *et al.*, 2002, de Napoles *et al.*, 2007 ; Sugimoto et Abe, 2007) (**Fig. 16**).



**Figure 16 : Dynamique de l'expression du gène *Xist* au cours du développement**

La première vague d'inactivation est celle dite d'inactivation soumise à empreinte parentale. Au stade 2-4 cellules, l'expression de *Xist* à partir du Xp provoque systématiquement son inactivation en mettant en place les marques de l'hétérochromatine. Alors que l'inactivation du Xp est maintenue dans les tissus extra-embryonnaires (TE et PE), dans les cellules de la masse cellulaire interne (ICM) de l'embryon, le Xp se réactive. Ces cellules présentent alors deux chromosomes X actifs. A E4,5, l'inactivation aléatoire survient au moment où les premiers événements de différenciation de l'épiblaste ont lieu. Alors qu'elle est stable dans toutes les autres cellules de l'individu, les chromosomes X se réactivent dans les lignées germinales.

D'après les revues, Reik et Lewis, 2005 ; Heard, 2005 ; Okamoto *et al.*, 2005 ; Senner et Brockdorff, 2009 ; Navarro et Avner, 2009

## 2.4 Les systèmes d'étude de l'inactivation du chromosome X

Comme l'inactivation du chromosome X soumise à empreinte parentale se déroule très tôt au cours du développement embryonnaire, son étude sur l'embryon est difficile. Nous axerons l'exposé sur l'inactivation aléatoire qui peut être étudié en utilisant des modèles plus facile à mettre en place. Nous présenterons d'abord ces différents systèmes.

### 2.4.1 Cellules somatiques (humaines, murines et hybrides)

Plusieurs caractéristiques de l'X inactivé ont d'abord été identifiées dans des cellules somatiques, du fait de la facilité à étudier ces cellules. De plus, les cellules somatiques hybrides (humaine/murine) ont permis d'analyser séparément les chromosomes Xa et Xi et d'en définir les caractéristiques. Ce système a permis aussi d'identifier les gènes qui échappent à l'inactivation (Clemson *et al.*, 1998). Néanmoins, bien que les cellules somatiques aient beaucoup d'avantages, elles ne permettent pas de récapituler les étapes de l'inactivation et d'avoir une vision dynamique du processus. Par ailleurs, comme l'inactivation du chromosome X est un phénomène développemental, l'utilisation d'embryons et de cellules embryonnaires a beaucoup contribué à la compréhension de ce processus.

### 2.4.2 Embryon et cellules ES

Des embryons créés par substitution de leur noyau par le noyau de cellules somatiques présentent une inactivation préférentielle du Xi provenant de la cellule donneuse dans les tissus extra-embryonnaires, ce qui suggère la persistance d'une mémoire épigénétique du Xi après transfert du noyau (Eggan *et al.*, 2000 ; Nolen *et al.*, 2005). L'utilisation d'embryons et des expériences de transfert de noyaux ont permis de mieux définir la nature de l'empreinte parentale sur l'X.

Les cellules ES ont été un élément moteur dans l'analyse du processus d'inactivation de l'X du fait de la facilité relative à les cultiver et surtout du fait qu'elles récapitulent les différentes étapes de la mise en place de l'inactivation. Elles sont isolées à partir de la masse interne du blastocyste et peuvent être mises en culture. Elles sont pluripotentes et immortelles, car lorsqu'elles sont placées dans des conditions appropriées de culture, elles s'auto-renouvellent de façon illimitée. Le facteur clé de cet auto-renouvellement est le facteur LIF (Leukemia Inhibitory Factor), qui active une voie de signalisation inhibant la différenciation (pour revue, Burdon *et al.*, 2002). L'élimination du facteur LIF du milieu de culture déclenche la différenciation des cellules ES et permet l'initiation de l'inactivation du chromosome X (pour revue, Fluckiger, 2003).

Les succès obtenus avec les cellules ES murines ont conduit à entreprendre des études similaires sur les cellules embryonnaires humaines (hES). Toutefois, la mise en place de l'inactivation du chromosome X dans les cellules humaines ne résulte pas comme pour les

cellules ES de souris d'un processus unique bien défini. Le Xi peut être présent avant la différenciation et dans certaines lignées, l'expression de Xist est abolie, ce qui provoque une réactivation du Xi (Hoffman *et al.*, 2005 ; Hall *et al.*, 2008 ; Shen *et al.*, 2008 ; Silva *et al.*, 2008). Dans l'état actuel des connaissances, les raisons pour lesquelles certaines cellules ES humaines présentent deux chromosomes X actifs sont méconnues. On ne sait pas si ces cellules représentent un état épigénétique naïf dans la population cellulaire ou si la réactivation du Xi est due à des conditions de cultures sous-optimales.

### 3. Les acteurs protéiques de l'inactivation aléatoire

Consécutivement à l'expression de l'ARN Xist, des protéines sont recrutées sur le chromosome X inactif. Nous allons consacrer ce paragraphe à la présentation des acteurs protéiques impliqués ou supposés être impliqués dans la mise en place de l'inactivation : protéines PcG, macroH2A, SAF-A, ATRX, BRCA1, PARP-1, smcHD1 et CTCF.

#### 3.1 Les protéines du groupe polycomb : PcG

Les protéines du groupe polycomb PcG ont été initialement découvertes chez la drosophile par Struhl en 1981. Elles jouent un rôle important dans les mécanismes épigénétiques et particulièrement dans l'inactivation du chromosome X (pour revue, Heard, 2005). Les PcG sont des répresseurs de la transcription. Elles ont de nombreux homologues chez l'homme, la souris, le poulet ou le xénope, mais aussi chez le nématode. Les protéines PcG sont classées en 2 catégories dénommées PRC1 et PRC2 (Polycomb Repressive Complexes) (pour revue, Levine *et al.*, 2004). Le complexe PRC2 est impliqué dans la triméthylation des histones H3 en position 27. PRC1 reconnaît ces modifications et permet l'ubiquitinylation de l'histone H2A en position 119 (H2AK119ub). Ces deux complexes concourent à la répression génique (Cao *et al.*, 2002 ; Czermin *et al.*, 2002 ; Muller *et al.*, 2002 ; Wang *et al.*, 2004 ; Cao *et al.*, 2005).

##### 3.1.1 PRC2

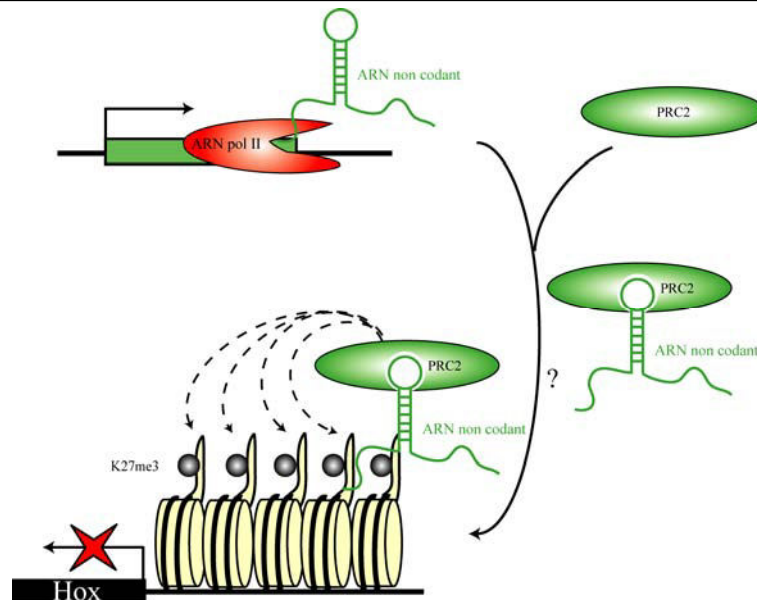
Le complexe PRC2 d'une masse moléculaire d'environ 600 kDa, a été initialement identifié chez la drosophile et appelé ESC/E(Z) (Ng *et al.*, 2000 ; Jones *et al.*, 1998). Il porte une activité histone méthyl-transférase. Chez les mammifères, le complexe correspondant appelé Eed/E(z)h2 est impliqué dans la mise en place des méthylation H3K27me3. Il permet de former des diméthylations H3K9 ceci en moindre proportion (Cao *et al.*, 2002 ; Czermin *et al.*, 2002 ; Muller *et al.*, 2002). Ce complexe est constitué des protéines Eed (Embryonic ectoderm development), Ezh2 (Enhancer of zeste 2), Suz12 (Suppressor of zeste 12), ainsi que des protéines RbAp46 et 48 (Cao *et al.*, 2002). Ezh2 porte l'activité histone méthyltransférase (Plath *et al.*, 2003 ; Silva *et al.*, 2003). Les protéines Eed et Suz12 sont nécessaires à la

stabilité du complexe et à l'activité HMTase (pour revue, Cao et Zhang 2004), RbAp46 et 48 permettraient la fixation du complexe aux protéines histones (Cao *et al.*, 2002).

Le complexe PRC2 fut le premier à être identifié comme essentiel à la maintenance de l'état inactif du Xi dans le trophectoderme de souris (Wang *et al.*, 2001 ; Kalantry *et al.*, 2006). De plus, des expériences d'immunofluorescence ont révélé que la protéine Eed est localisée sur le Xi dans les cellules ES en voie de différenciation, cette localisation étant concomitante avec la mise en place des H3K27me3 (Silva *et al.*, 2003 ; Plath *et al.*, 2004). Les 2 autres composants du complexe PRC2, Suz12 et Ezh2 s'associent à Eed sur le chromosome X inactif dans les embryons murins précoces (Okamoto *et al.*, 2004 ; Plath *et al.*, 2003) et au cours de la différenciation des cellules ES (Plath *et al.*, 2003 ; Silva *et al.*, 2003). Néanmoins, l'activité du complexe PRC2 n'est pas indispensable à l'inactivation du chromosome X dans les cellules embryonnaires (Kalantry et Magnuson, 2006 ; Schoeftner *et al.*, 2006). L'existence de facteurs encore inconnus portant des fonctions redondantes à celles du complexe PRC2 n'est donc pas exclue.

Il a également été démontré que le recrutement de PRC2 est dépendant de l'ARN Xist (Plath *et al.*, 2003 ; Kohlmaier *et al.*, 2004 ; Plath *et al.*, 2004). Plusieurs études ont suggéré que ce complexe est capable de s'associer à des molécules d'ARN. Ainsi, lors d'expériences d'immunoprécipitation à partir de fibroblastes humains, suivies par des analyses par RT-PCR, une association de Suz12 avec l'ARN non-codant HOTAIR a été détectée. Il s'agit d'un ARN non-codant de 2,2 kb exprimé à partir du *locus* HoxC. Cet ARN réprime *in trans* le *locus* HoxD, qui lui joue un rôle majeur dans la régulation des processus développementaux (Rinn *et al.*, 2007). Le fait que l'ARN U1, utilisé comme contrôle, n'interagisse pas avec Suz12 suggère fortement que l'interaction avec l'ARN HOTAIR est spécifique. Les auteurs proposent un modèle dans lequel HOTAIR permettrait le recrutement du complexe PRC2 et provoquerait le ciblage et la mise en place des méthylations H3K27me3 au *locus* HoxD (**Fig. 17**). Selon ce modèle, l'ARN permet le ciblage des gènes à réprimer. Comme nous l'avons vu précédemment, l'ARN Kcnq1ot1 est aussi associé aux protéines Ezh2 et Suz12 dans le placenta murin et ceci au niveau des gènes dont il régule l'expression (Pandey *et al.*, 2008). En ce qui concerne l'inactivation du chromosome X, des expériences d'immunoprécipitation de l'ARN (RIP) ont permis de montrer que dans les cellules ES non différenciées de souris, Suz12 et Ezh2 sont recrutées par la région des A-repeats, cette région de l'ARN Xist est cruciale pour l'initiation de l'inactivation (Zhao *et al.*, 2008). Comme nous l'avons déjà mentionné, le complexe PRC2 interagit avec les méthylations H3K27me3 qu'il catalyse et joue un rôle dans la propagation et l'amplification de ces marques épigénétiques (Hansen *et al.*, 2008).





**Figure 17 : Modèle de régulation de l'expression des gènes *Hox* par les ARN non-codants (HOTAIR) via les enzymes de modifications des histones**

Le recrutement du complexe PRC2 est programmé par la production d'un ARN non-codant en *trans* (HOTAIR), qui permet le ciblage de l'activité du complexe PRC2 au locus d'intérêt. Le recrutement de ce complexe conduit à la méthylation des histones H3 en position 27 et à la répression transcriptionnelle des gènes *Hox* au voisinage.

D'après Rinn *et al.*, 2007

### 3.1.2 PRC1

Comme nous l'avons déjà mentionné, le complexe PRC1 est impliqué dans le remodelage de la chromatine conduisant à l'extinction de la transcription de la chromatine (Francis *et al.*, 2001 ; Francis *et al.*, 2004 ; King *et al.*, 2002 ; Mohd-Sarip *et al.*, 2006 ; Shao *et al.*, 1999). Le complexe PRC1 porte une activité ubiquitine ligase E3 dirigée contre l'histone H2A (Cao *et al.*, 2005 ; de Napoles *et al.*, 2004 ; Elderkin *et al.*, 2007 ; Wang *et al.*, 2004). Chez les mammifères, le corps du complexe comporte les protéines Bmi-1 (B cell-specific Mo-MLV integration site 1), certaines protéines de la famille Pc (Polycomb appelé aussi Cbx), de la famille Ph (Polyhomeotic), Ring1a et 1b (Ring-finger protein 1 et 2 connus aussi sous le nom de Ring1 et Rnf2, respectivement). Ring1a/b porte l'activité ligase E3 dirigée contre H2AK119 (Cao *et al.*, 2005 ; de Napoles *et al.*, 2004 ; Wang *et al.*, 2004).

Chez les mammifères, le complexe PRC1 est retrouvé associé au Xi. Il inclut les protéines Ring1a et Ring1b, qui sont essentielles à la mise en place des H2AK119ub1 sur le Xi dans des lignées de fibroblastes murins (de Napoles *et al.*, 2004 ; Fang *et al.*, 2004). La composition précise du complexe PRC1 varie au cours du processus d'inactivation et en fonction du type cellulaire (Plath *et al.*, 2004 ; pour revue, Heard, 2005). Ainsi, la protéine Phc1 pourrait jouer un rôle lors de la phase d'initiation de l'inactivation du chromosome X,

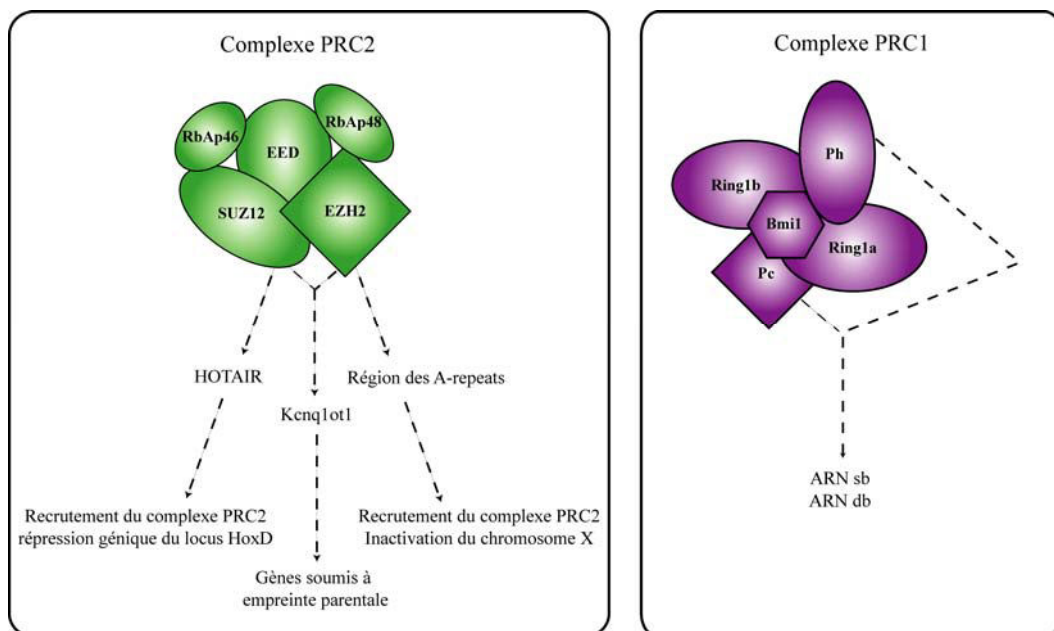
puis elle serait progressivement remplacée par Phc2, en association notamment avec Bmi-1 et Cbx2 (Plath *et al.*, 2004).

Nous avons vu que les méthylations H3K27me3 augmentent l'affinité de la chromatine pour les protéines polycomb à chromodomains, telle que Pc (Fischle *et al.*, 2003 ; Bernstein *et al.*, 2006). Cela se traduit par le fait que la perte des H3K27me3 dans les cellules ES de souris dans lesquelles le gène *Eed* a été délété compromet la fixation de Ring1b (Boyer *et al.*, 2006). Il faut noter que PRC1 peut être retrouvé dans des régions dépourvues de H3K27me3 (Plath *et al.*, 2004) et donc pourrait être recruté par un autre mécanisme faisant appel à des facteurs protéiques comme la protéine PHO, une protéine fixant aux PRE (Polycomb Response Elements) (Brown *et al.*, 1998 ; Klymenko *et al.*, 2006 ; Mihaly *et al.*, 1998). PHO peut interagir directement avec le complexe PRC1 *in vitro* (Mohd-Sarid *et al.*, 2005). De plus, la protéine Ring1b et la modification H2AK119ub1 peuvent être détectées dans des cellules délétées pour le gène *Eed* (Schoeftner *et al.*, 2006). Un recrutement de Ring1b indépendant de PRC2 est donc possible lors de l'inactivation du chromosome X. Ce recrutement peut passer par la formation d'autres complexes incluant Ring1b comprenant l'histone déméthylase Fbx110/Jhdm1 récemment identifiée (Sanchez *et al.*, 2007).

Le mécanisme de recrutement du complexe PRC1 sur le chromosome X inactif reste encore mal connu, il pourrait mettre en œuvre l'interaction de PRC1 à des molécules d'ARN. L'interaction des protéines Ph1 des expériences de retard sur gel (ou Rae28) (Zhang *et al.*, 2004) et Cbx7 du complexe PRC1 (Bernstein *et al.*, 2006) avec des ARN simples ou doubles brins a été mis en évidence par des expériences de gel retard. De plus, l'interaction de Cbx7 avec le Xi dépend de l'association de l'ARN Xist avec ce chromosome (Bernstein *et al.*, 2006).

Enfin, il a été démontré récemment que le complexe PRC1 s'associe à la chromatine et y reste associé lors d'expériences de réplication *in vitro* ce qui suggère un rôle possible de ce complexe dans les mécanismes d'héritages épigénétiques lors des divisions cellulaires. Ceci ouvre des pistes d'études en vue de comprendre les processus permettant le transfert de caractères géniques à la descendance cellulaire (Francis *et al.*, 2009).

La **Figure 18** présente un état des lieux des interactions des protéines des complexes PRC1 et PRC2 avec les molécules d'ARN.



**Figure 18 : Interactions des protéines des complexes PRC1 et PRC2 avec les ARN**

Les protéines des complexes PRC1 et PRC2 sont représentées et les identités des ARN cibles ainsi que leurs rôles sont indiqués.

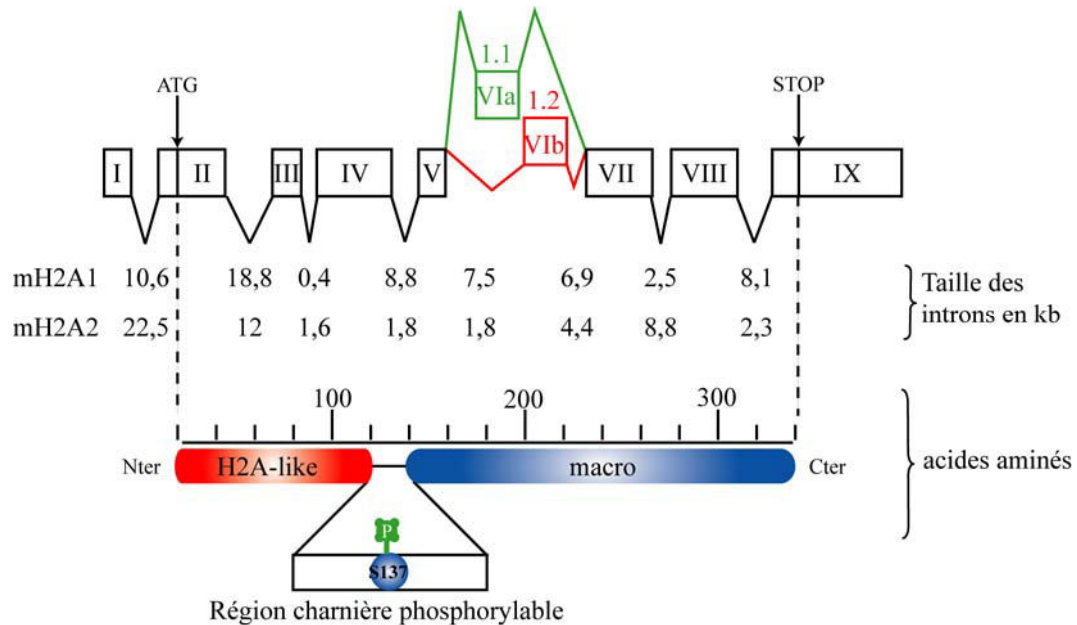
Les protéines Ezh2 et Suz12 du complexe PRC2 sont recrutées par les ARN non-codants HOTAIR, Kcnq1ot1 et la région des A-repeats de l'ARN Xist. Ce recrutement permet la mise en place des marques H3K27me3.

Les protéines Pc et Ph du complexe PRC1 interagissent avec des molécules d'ARN simples ou doubles brins et permettent la mise en place des H2AK119ub1.

D'après Hekimoglu et Ringrose, 2009

### 3.2 MacroH2A

MacroH2A est un variant d'histone enrichi sur le Xi (Costanzi et Pehrson, 1998 ; Mermoud *et al.*, 1999; Rasmussen *et al.*, 2000 ; Nusinow *et al.*, 2007). Elle a la particularité d'être de grande taille (42 kDa, soit trois fois celle de l'histone canonique H2A) (Pehrson et Fried, 1992). La partie N-terminale de la protéine portant l'*histone fold* comporte un domaine C-terminal, appelé "domaine macro" (*macrodomain*) ou NHR pour *Non Histone Region* (Pehrson et Fried, 1992 ; Allen *et al.*, 2003; Chakravarthy *et al.*, 2005). Il existe deux isoformes non-alléliques, macroH2A1 et macroH2A2 (Chadwick et Willard, 2001; Costanzi et Pehrson, 2001). De plus, l'ARN transcrit à partir du gène *mH2A1* subit un épissage alternatif conduisant aux formes mH2A1.1 et mH2A1.2, différant par seulement 30 acides aminés (Chadwick et Willard, 2001b) (**Fig. 19**). Enfin, macroH2A et ses variants peuvent être modifiés (Hernandez-Munoz *et al.*, 2005 ; Chu *et al.*, 2006 ; Ogawa *et al.*, 2005). En particulier, macroH2A peut présenter une phosphorylation au niveau de sa sérine 137 (mH2AS137ph) (Bernstein *et al.*, 2008). Dans la suite du texte, le terme « macroH2A » englobera les deux isoformes « macroH2A1 » et « macroH2A2 », les deux variants d'épissage « macroH2A1.1 » et « macroH2A1.2 » étant aussi inclus.



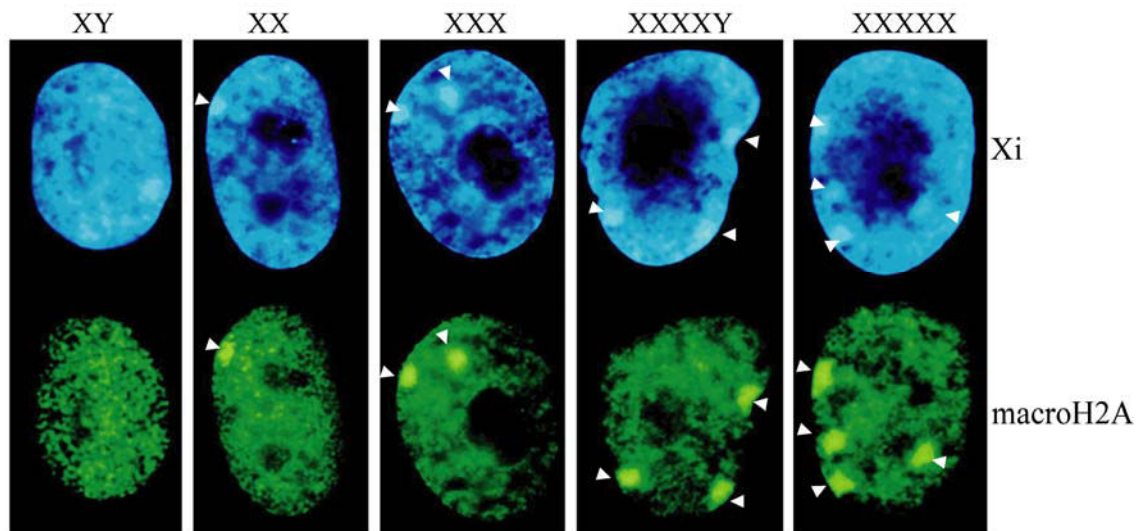
### Figure 19 : Composition en exons des ARNm des variants de macroH2A et domaines protéiques correspondants

L'épissage alternatif à l'origine des ARNm de macroH2A1.1 et macroH2A1.2 est indiqué en vert et en rouge. L'exon VIa est spécifique de la forme macroH2A1.2, alors que l'exon VIb est spécifique de la forme macroH2A1.1. Les tailles des introns des ARN pré-messagers de macroH2A1 et macroH2A2 sont indiquées sous la représentation schématique du gène.

Les domaines H2A-like et macro de la protéine macroH2A sont respectivement indiqués en rouge et en bleu. Entre ces 2 domaines, la région charnière comprend la sérine 137 qui est phosphorylable, mH2A1 non phosphorylée colocalise avec le Xi de façon Xist-dépendante, alors que mH2AS137ph en est exclue.

D'après Kustatscher *et al.*, 2005 ; Chadwick et Willard, 2003 ; Bernstein *et al.*, 2008

L'utilisation de cellules portant "n" chromosomes X permet l'observation de "n-1" foyers de macroH2A (Chadwick et Willard, 2001a). Ces expériences ont permis de suggérer une localisation sur l'X dépendante du nombre de Xi (**Fig. 20**). MacroH2A1 est en effet enrichie sur le Xi (Costanzi et Pehrson, 1998 ; Mermoud *et al.*, 1999 ; Rasmussen *et al.*, 2000 ; Nusinow *et al.*, 2007 ; Chadwick et Willard, 2002 ; Rasmussen *et al.*, 2000 ; Mermoud *et al.*, 2001). Cet enrichissement aurait pu refléter une densité globale en nucléosomes sur ce chromosome (Perche *et al.*, 2000), néanmoins, des expériences combinant la technique CHIP à des puces à ADN (*ChIp on Chip*) confirment un enrichissement relatif de macroH2A1 sur le Xi 1,5 fois supérieur à celui des autosomes (Mietton *et al.*, 2009).



**Figure 20 : Foyers de concentration de la protéine macroH2A et co-localisation avec le chromosome X inactif**

La protéine macroH2A est détectée par immunofluorescence dans plusieurs lignées de fibroblastes primaires marquées au DAPI (46 XY, 46 XX, 47 XXX, 49 XXXXY et 49 XXXXX). Les cellules sont analysées par microscopie et les images des noyaux analysés sont présentées. La ligne supérieure correspond à la localisation du Xi. Indiqué par des flèches blanches, il apparaît sous une forme dense et compacte au voisinage de la membrane nucléaire. La localisation de macroH2A est illustrée dans la série d'images au bas de la figure (flèches blanches), elle se concentre en foyers qui co-localisent avec le Xi.

D'après Chadwick et Willard, 2001

La localisation de macroH2A1 sur le Xi est dépendante de *Xist* (Beletskii *et al.*, 2001; Csankovszki *et al.*, 1999). Dans plusieurs lignées cellulaires murines, l'immunoprécipitation de la chromatine à l'aide d'un anticorps dirigé contre macroH2A1 démontre que *Xist* et macroH2A1 appartiennent aux mêmes complexes ribonucléoprotéiques (Gilbert *et al.*, 2000). L'expression ectopique d'un transgène *Xist* sur un autosome est suffisante pour promouvoir une accumulation de macroH2A sur ce chromosome (Rasmussen *et al.*, 2001). Ces observations montrent que *Xist* dirige le positionnement de macroH2A1 sur le futur Xi, ce qui a pu être vérifié par la délétion conditionnelle de *Xist*. Celle-ci entraîne une délocalisation de macroH2A1 (malgré un maintien de la répression). L'utilisation d'un transgène inducible intégré sur le chromosome X, codant pour différents mutants de *Xist*, a permis de montrer que la partie 3' de l'ARN *Xist* est nécessaire au recrutement de macroH2A1 (Wutz *et al.*, 2002 ; Beletskii *et al.*, 2001 ; Csankovszki *et al.*, 1999). L'ensemble de ces données démontre que la présence de l'ARN *Xist* est nécessaire et suffisante pour promouvoir l'incorporation de macroH2A1 sur le chromosome X inactif.

Comme nous l'avons déjà mentionné, la protéine macroH2A est modifiée post-traductionnellement. La forme phosphorylée de macroH2A (mH2AS137ph) dans la lignée cellulaire HEK293 est exclue du Xi et enrichie durant la mitose (Bernstein *et al.*, 2008). Ainsi,

cette sous-population mH2AS137ph ne semble pas jouer de rôle dans le processus d'inactivation du chromosome X. De plus, il a été mis en évidence que les protéines CULLIN3 et SPOP, qui portent des activités ubiquitines ligases, sont responsables de l'ubiquitinylation de macroH2A. En l'absence de l'une ou l'autre de ces enzymes, macroH2A n'est plus localisée sur le Xi, alors que la localisation de l'ARN Xist n'est pas modifiée (Hernandez-Munoz *et al.*, 2005 ; Takahashi *et al.*, 2002 ; Hernandez-Munoz *et al.*, 2005). Les protéines CULLIN3 et SPOP sont donc impliquées dans la localisation de macroH2A1 sur le Xi.

Toutefois, l'implication forte de mH2A dans l'inactivation du chromosome X est peu probable car les souris knock-out pour mH2A1 sont viables, fertiles et n'ont pas de défaut d'inactivation (Changolkar *et al.*, 2007). Ceci est d'autant plus vraisemblable que mH2A est exprimée en quantité équivalente chez la femelle et chez le mâle (Costanzi et Pehrson, 1998 ; Mermoud *et al.*, 1999 ; Rasmussen *et al.*, 2000). L'ARN Xist est nécessaire à la mise en place de l'inactivation mais en absence de Xist, le chromosome X déjà inactif (dépourvu en macroH2A) ne se réactive pas (Beletskii *et al.*, 2001; Csankovszki *et al.*, 1999). Il en est de même pour l'absence de macroH2A1. Comme Xist, macroH2A ne semble donc pas nécessaire au maintien de la répression de la transcription du Xi (Hernandez-Munoz *et al.*, 2005).

### 3.3 Les autres facteurs protéiques associés au Xi

#### 3.3.1 SAF-A

Chez l'homme, une autre protéine a été mise en évidence au niveau du Xi. La protéine SAF-A (Scaffold Attachment Factor A), qui est essentielle au développement embryonnaire (Roshon et Ruley, 2005). Elle est impliquée dans certaines régulations transcriptionnelles (Eggert *et al.*, 2001 ; Kim et Nikodem, 1999), dans l'acétylation des histones par son interaction avec la protéine p300/CBP (Martens *et al.*, 2002), ainsi que dans la réparation de l'ADN (Berglund et Clarke, 2009). Elle a été décrite il y a longtemps comme un des composants de la matrice nucléaire (Fackelmayer *et al.*, 1994). Dans les lignées HEK293, la protéine SAF-A est significativement enrichie sur le Xi et cette localisation dépend de son site de fixation aux ARN, le domaine RGG (Helbig et Fackelmayer, 2003). Elle peut être extraite du Xi par un traitement à la RNase (Fackelmayer, 2005). Il est donc envisageable que SAF-A soit liée à ce chromosome par l'intermédiaire de l'ARN Xist. SAF-A pourrait contribuer à l'immobilisation de l'ARN Xist sur le Xi et la coopération de ces 2 partenaires pourrait permettre la formation d'une structure stable dans le territoire du Xi. Cette structure servirait de plate-forme pour la maintenance de l'état silencieux du Xi (Fackelmayer, 2005).

### 3.3.2 ATRX

Un nouveau partenaire protéique du Xi a été identifié récemment. Il s'agit de la protéine ATRX (Alpha Thalassemia/mental Retardation X-linked) impliquée dans le remodelage de la chromatine et appartenant à la famille des hélicases SWI/SNF2 ATP-dépendantes. Son association à l'hétérochromatine constitutive avait été mise en évidence et on savait qu'elle jouait un rôle important dans le contrôle de la méthylation de l'ADN, en particulier, celui des régions répétées du génome humain (Gibbons *et al.*, 2000 ; McDowell *et al.*, 1999). Des études récentes suggèrent également son implication dans la formation d'hétérochromatine, dans l'alignement des chromosomes lors de la méiose, dans la cohésion des chromosomes dans les cellules somatiques et dans la régulation transcriptionnelle (Baumann *et al.*, 2008 ; De la Fuente *et al.*, 2004 ; Gibbons *et al.*, 2000 ; McDowell *et al.*, 1999). Sa délétion provoque une désorganisation des tissus extra-embryonnaires associée à un défaut d'inactivation du chromosome Xp (soumis à empreinte parentale) et conduit à la mort de l'embryon (Garrick *et al.*, 2006). Récemment, une étude a mis en évidence l'accumulation d'ATRX sur le Xi dans les cellules somatiques, trophoblastiques mais aussi dans les cellules ES en voie de différenciation. Elle co-localise avec mH2A1.2 et H2AK119ub qui sont comme nous l'avons vu précédemment, deux marqueurs établis du chromosome X inactif. Etant donnée son association tardive avec le Xi dans les cellules ES en voie de différenciation (8 jours après différenciation), il a été suggéré qu'ATRX intervienne non pas dans la phase d'initiation de l'inactivation mais dans la phase tardive consistant à maintenir l'état inactif du Xi (Baumann *et al.*, 2008).

### 3.3.3 BRCA1

La participation de la protéine suppresseur de tumeur BRCA1 à la localisation correcte de l'ARN Xist sur le Xi est encore beaucoup débattue mais elle est importante à mentionner. Cette protéine appartient à la famille des ubiquitines E3 ligases à domaine RING finger. L'importance de ce domaine et de l'activité E3 ligase putative de BRCA1 se traduit par la fréquence de mutations du RING finger dans les cancers héréditaires liés à une déficience en protéine BRCA1 (Brzovic *et al.*, 2003 ; Brzovic *et al.*, 2001 ; Hashizume *et al.*, 2001 ; Ruffner *et al.*, 2001). Cette protéine est impliquée dans les points de contrôle du cycle cellulaire, dans la réparation de l'ADN (pour revue, Deng *et al.*, 2006 ; pour revue, Venkitaraman, 2002) et dans l'inactivation méiotique (Turner *et al.*, 2004). Sa déplétion entraîne un phénotype de létalité embryonnaire chez la souris et une instabilité chromosomique (McCarthy *et al.*, 2003). Il a été démontré que la protéine BRCA1 co-localise avec le Xi en phase S tardive (Diaz-Perez *et al.*, 2005 ; Ganesan *et al.*, 2004 ; Pageau *et al.*, 2007 ; Silver *et al.*, 2007). De plus, la localisation de l'ARN Xist sur le Xi n'est plus observée dans plusieurs lignées ou cultures primaires de cellules tumorales provenant de cancers du

sein ou des ovaires et dans lesquelles BRCA1 est mutée (Ganesan *et al.*, 2002 ; Ganesan *et al.*, 2004 ; Silver *et al.*, 2007 ; Vincent-Salomon *et al.*, 2007 ; Xiao *et al.*, 2007). Il a également été observé que l'expression ectopique de BRCA1 a pour effet d'augmenter la quantité de Xist (Pageau *et al.*, 2007). Toutefois, aucun rôle de cette régulation de l'expression n'a encore été montrée dans l'inactivation du chromosome X (Pageau *et al.*, 2007). Autre argument en faveur d'un lien possible entre BRCA1 et la fixation de Xist sur le Xi, l'ARN Xist n'est pas détecté sur le Xi dans des cellules dont le taux de BRCA1 a été diminué par ARN interférence. Ces cellules ont tendance à réactiver un gène rapporteur lié au Xi (Ganesan *et al.*, 2002 ; Silver *et al.*, 2007). Enfin, lorsque la protéine BRCA1 est exprimée de manière ectopique dans une lignée de cellules cancéreuses dans laquelle BRCA1 est mutée, la localisation de Xist sur le Xi est restaurée (Ganesan *et al.*, 2002 ; Ganesan *et al.*, 2004 ; Silver *et al.*, 2007).

En dépit de cet ensemble de résultats qui semble convaincant, Xiao *et al.*, (2007) ont décrit des résultats complètement opposés, ce qui soulève la question d'une réelle implication de BRCA1 dans l'inactivation du chromosome X. Il faut noter que l'absence de concentration de Xist sur le Xi dans des tumeurs peut être liée à d'autres paramètres tels que la duplication du Xa qui est fréquente du fait des nombreux réarrangements génomiques dans les cellules cancéreuses (Sirchia *et al.*, 2005 ; Vincent-Salomon *et al.*, 2007, pour revue, Gudmundsdottir et Ashworth, 2006). La présence d'une forme mutée de BRCA1 ne conduit pas toujours à l'absence de concentration de Xist sur le Xi. Par ailleurs, l'absence de localisation de Xist est fréquemment observée dans des cellules tumorales provenant de cancers du sein qui ne sont pas causés par la mutation de BRCA1 (Ganesan *et al.*, 2005 ; Richardson *et al.*, 2006 ; Silver *et al.*, 2007; Sirchia *et al.*, 2005 ; Vincent-Salomon *et al.*, 2007 ; Xiao *et al.*, 2007). Toutes ces observations suggèrent que la co-localisation entre Xist et BRCA1 n'est peut-être pas spécifique du Xi, mais correspondrait plutôt à une association générale de BRCA1 à l'hétérochromatine constitutive au niveau des centromères (Pageau *et al.*, 2007 ; Pageau et Lawrence, 2006 ; Xiao *et al.*, 2007). Par ailleurs, comme BRCA1 intervient dans un grand nombre de processus, une déficience en BRCA1 peut affecter une voie de régulation essentielle à l'habillage du Xi par Xist sans pour autant que BRCA1 ait une action directe dans l'habillage.

#### 3.3.4 PARP-1

La protéine PARP-1 se lie aux nucléosomes et affecte la structure de la chromatine (Kim *et al.*, 2004 ; pour revue, Kim *et al.*, 2005). Elle fait partie de la famille des protéines PARP qui catalysent l'addition d'ADP-ribose sur des protéines cibles (pour revue, Ame *et al.*, 2004). Elle intervient dans le maintien et l'adaptation de la structure de la chromatine mais aussi dans la détection des dommages de l'ADN (Kim *et al.*, 2004 ; Rouleau *et al.*, 2007). Les



cellules HEK 293 dans lesquelles le gène *PARP-1* est éteint présentent une réactivation des gènes du Xi, ce qui suggère son implication dans la maintenance de l'hétérochromatine du Xi. Toutefois, les analyses ont été réalisées dans des cellules somatiques et il serait important pour conclure qu'elles soient répétées sur un modèle d'étude permettant de récapituler les phases de l'inactivation du chromosome X, tel que les cellules ES de souris ou l'embryon murin à proprement dit.

### 3.3.5 smcHD1

Lors d'un crible visant à identifier les modificateurs de variabilité concernant la méthylation des îlots CpG sur le Xi, le locus *Momme D1* a été identifié. La délétion de ce locus grand de 2,4 Mb, localisé sur le chromosome 17 de souris, provoque la mort des embryons femelles à mi-gestation, ainsi qu'une diminution importante de la méthylation des îlots CpG du Xi (Blewitt *et al.*, 2006). La protéine smcHD1 codée par ce locus et jusque là non caractérisée semble être responsable du phénotype (Blewitt *et al.*, 2008). La protéine smcHD1 co-localise avec le Xi. Elle n'est pas essentielle à l'expression de *Xist* mais est requise pour l'extinction transcriptionnelle et la méthylation des gènes du Xi (Blewitt *et al.*, 2008). SmcHD1 contient un domaine N-terminal d'hydrolyse de l'ATP et un domaine C-terminal « SMC hinge », trouvé dans les protéines SMC impliquées dans la condensation de la chromatine, la cohésion et la ségrégation des chromosomes. Chez les mammifères, les protéines du complexe SMC interagissent avec la protéine Snf2h, une enzyme de remodelage de la chromatine (Hakimi *et al.*, 2002) et avec la Dnmt3b (Geiman *et al.*, 2004).

### 3.3.6 CTCF

Des expériences d'immunofluorescence ont révélé une concentration élevée de la protéine CTCF sur le Xi, ce qui a suggéré son implication dans l'inactivation du chromosome X (Chadwick et Willard, 2003 ; Filippova *et al.*, 2005). C'est une protéine très conservée ayant de multiples fonctions dans la régulation de la transcription. Elle joue aussi un rôle important dans l'architecture nucléaire (Parelho *et al.*, 2008). CTCF a été retrouvée dans des domaines chromosomiques particuliers et permettrait de séparer les gènes inactivés des gènes échappant à l'inactivation (Filippova *et al.*, 2005 ; Barski *et al.*, 2007). De plus, CTCF est associée au promoteur du gène *Xist* (Navarro *et al.*, 2006, Kim *et al.*, 2006) mais aussi à celui de *Tsix* en association avec Yy1. Elle permet d'activer l'expression de *Tsix* sur le Xa (Chao *et al.*, 2002 ; Donohoe *et al.*, 2007). Yy1 est un facteur de transcription ubiquitaire et les mutants homozygotes *Yy1*<sup>-/-</sup> meurent précocément au cours du développement (pour revue, Gordon *et al.*, 2006). Plus récemment, il a été démontré que CTCF intervient dans l'appariement des 2 chromosomes X lors de l'initiation de l'inactivation étant donné que son extinction bloque cet

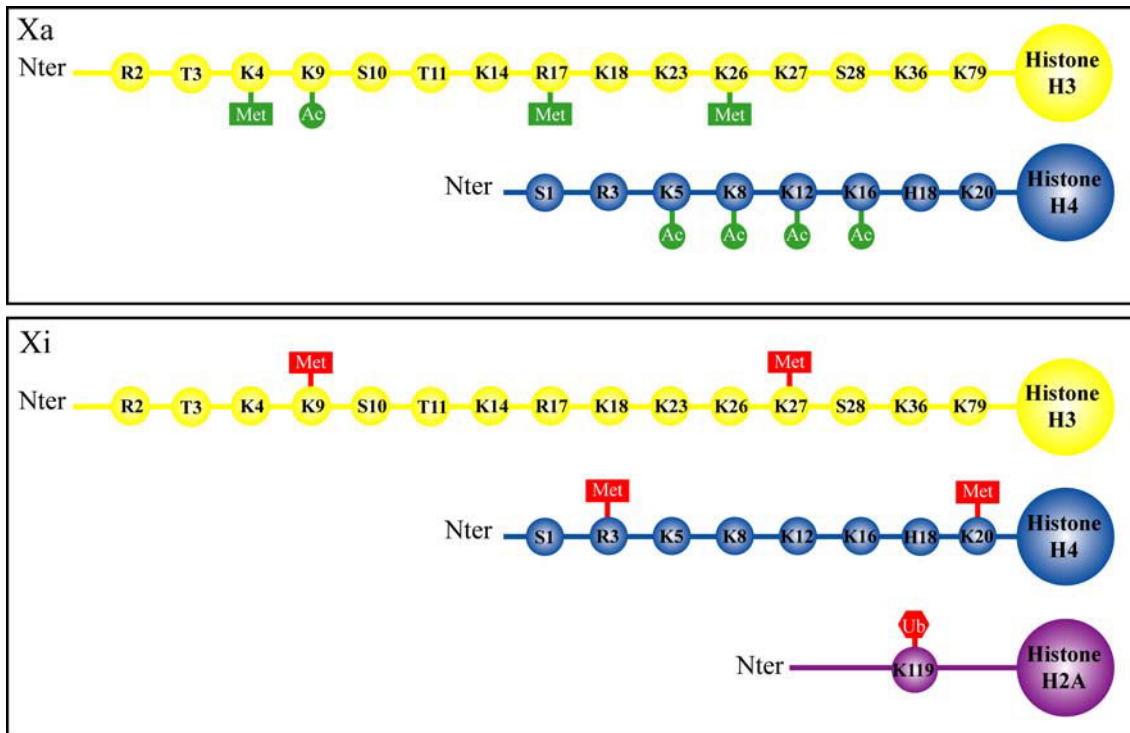
appariement (Xu *et al.*, 2007). Cette protéine multifonctionnelle joue donc vraisemblablement un rôle dans l'inactivation du chromosome X.

#### 4. Les caractéristiques du Xi

Nous venons d'aborder les différents acteurs protéiques impliqués ou susceptibles d'être impliqués dans le processus d'inactivation du chromosome X. L'expression de l'ARN Xist induit une cascade d'événements qui mène au recrutement direct ou indirect de ces facteurs protéiques. Certaines d'entre-eux comme les protéines du complexe PRC2 et PRC1 mettent en place les marques épigénétiques du Xi, que nous présenterons dans ce paragraphe. Nous aborderons également les autres caractéristiques du Xi telles que la méthylation de l'ADN, la réplication asynchrone et nous finirons en exposant les particularités des gènes échappant à la répression transcriptionnelle.

##### 4.1 Les marques épigénétiques du Xi

Comme nous l'avons déjà mentionné, l'expression de l'ARN Xist est le premier événement de l'initiation de l'inactivation qu'elle soit aléatoire ou empreintée. Sa sur-expression précède l'extinction transcriptionnelle des gènes le long du chromosome X (Kay *et al.*, 1993 ; Panning *et al.*, 1997 ; Sheardown *et al.*, 1997b ; Wutz et Jeanisch, 2000 ; Okamoto *et al.*, 2004). Cette capacité de l'ARN Xist à promouvoir l'extinction transcriptionnelle est encore à ce jour mal comprise, toutefois plusieurs études ont permis d'apporter de nombreuses informations. L'accumulation de l'ARN Xist est l'élément déclencheur d'une cascade d'événements aboutissant à la mise en place de la répression transcriptionnelle stable du chromosome X. Parmi ces événements, les méthylations des histones sont cruciales ce qui confère à l'X un profil de méthylation caractéristique des chromosomes actifs ou inactifs. En effet, le chromosome Xi présente des H3K27m3 (Rougeulle *et al.*, 2004). D'autres modifications telles que les H3K9me2 (Okamoto *et al.*, 2004), H3K9me3, H4K20m1 (Kolhmaier *et al.*, 2004), H2AK119ub1 apparaissent aussi sur le Xi (Smith *et al.*, 2004 ; Hernandez-Munoz *et al.*, 2005). Une diminution de H3K9ac, H3K4m2, H3K4m3, et la déacétylation globale de H4 est également observée (Keohane *et al.*, 1996 ; Heard *et al.*, 2001 ; Chaumeil *et al.*, 2002). Par contre, le Xa présente un enrichissement en H3K4me, H3K9ac, H3R17me, H3K26me, ainsi qu'une acétylation globale de l'histone H4 (**Fig. 21**).



**Figure 21 : Les différents types de modifications présentes sur les chromosomes X**

L'identité et la position des résidus modifiés sont indiquées par le code à 1 lettre des acides aminés suivi du chiffre de référence à partir de l'extrémité N-terminale.

Les modifications sont indiquées de la façon suivante : Ac pour acétylation, Met pour méthylation et Ub pour ubiquitinylation.

Les résidus lysine et arginine peuvent être méthylés, les lysines peuvent être acétylées et ubiquitinylées et les sérines et acides glutamiques peuvent être phosphorylés. Les modifications associées à la répression génique sont indiquées en rouge et celles associées à la chromatine active sont en vert.

Xa : X actif, Xi : X inactif.

D'après la revue Heard et Disteche, 2006

#### 4.2 Réplication tardive

Dans les cellules somatiques de mammifères, tandis que le Xa est répliqué en début de phase S, le Xi l'est à la fin. Cette propriété est très bien conservée chez les mammifères (Evans *et al.*, 1965 ; Priest *et al.*, 1967 ; Sharman, 1971 ; Taylor et Miner, 1968). La réplication tardive suit de peu l'apparition de la plupart des modifications d'histones (Chaumeil *et al.*, 2002 ; Okamoto *et al.*, 2005), et pourrait donc être liée à l'effet de ces modifications de la chromatine. Il a été montré qu'une fois l'inactivation mise en place, la délétion conditionnelle de *Xist* sur le Xi n'a pas d'effet sur la réplication tardive de ce dernier (Csankovszki *et al.*, 1999). Cette observation illustre bien le fait qu'une fois passée l'initiation de l'inactivation, la réplication asynchrone est maintenue au même titre que les autres caractéristiques du Xi et ceci indépendamment du gène *Xist*. Notons néanmoins que la réplication du Xi se fait plus tôt que celle du Xa dans les tissus extra-embryonnaires (Sugawara *et al.*, 1983), ce qui va à l'encontre de ce qui est observé dans les tissus embryonnaires.

### 4.3 Méthylation de l'ADN

La méthylation du Xi intervient tardivement et semble être importante pour le maintien de l'état inactif. Chez la souris, la majorité des promoteurs du Xi sont hyperméthylés en comparaison de ceux du Xa (Wolf *et al.*, 1984). Ceux des gènes échappant à l'inactivation ne sont pas méthylés (Goodfellow *et al.*, 1988). Chez l'homme, la relation entre méthylation de l'ADN et inactivation des gènes est plus compliquée. Par exemple les *loci* DXZ4 et DXS255 sont méthylés sur le Xa et non méthylés sur le Xi (Boyd et Fraser, 1990 ; Giacalone *et al.*, 1992). De plus, des expériences de restriction sensibles aux méthylations suggèrent que le Xi humain est hypo-méthylé par rapport au Xa, ce qui pose clairement un problème dans la conception actuelle des mécanismes impliqués. Cela suggère que la méthylation de l'ADN n'aurait pas le même impact sur la régulation de l'expression génique chez la souris et chez l'homme. Cette hypothèse n'est pas à exclure étant donné que les analyses comparatives entre homme et souris tendent à montrer que les processus régissant les mêmes fonctions ne sont pas équivalents.

### 4.4 Gènes échappant à l'inactivation

Tous les gènes du Xi ne sont pas inactivés. En effet, chez la souris, au moins 7 d'entre eux échappent à l'inactivation. Chez l'homme, il avait été proposé que 15 % des gènes restent actifs (pour revue, Brown et Grealley, 2003 ; Carrel *et al.*, 1999) mais une étude récente met en évidence un pourcentage de gène actif de seulement 5% (Johnston *et al.*, 2008). Cette « non-inactivation » conservée chez les eutheriens (Yen *et al.*, 2007) conduit à une expression bi-allélique des gènes et témoigne d'une résistance à la propagation de l'inactivation au niveau de certains *loci* (Lingenfelder *et al.*, 1998). A noter le fait que le niveau d'expression des gènes liés au Xi est rarement le même que celui des gènes correspondant au Xa. La sur-expression chez les femelles par rapport aux mâles de quelques gènes liés aux chromosomes X a récemment été mise en évidence (Johnston *et al.*, 2008), ce qui confirme que certains gènes échappent effectivement totalement à l'inactivation. Chez l'homme, les gènes non soumis à l'inactivation sont regroupés en régions échappant à l'hétérochromatinisation (Carrel et Willard, 2005 ; Jegalian et Page, 1998 ; Lahn et Page, 1997). Dans ces régions, la densité de régions répétées type LTR est plus faible (Tsuchiya *et al.*, 2004). Comme nous le verrons plus tard, les séquences répétées LINE-1 sont supposées être à l'origine de la propagation de l'inactivation (pour revue, Lyon, 1998).

Une approche bioinformatique a permis de montrer que les motifs (GATA)<sub>n</sub> sont enrichis de manière spécifique au niveau des gènes échappant à l'inactivation (McNeil *et al.*, 2006). Ainsi, l'appauvrissement en régions LTR et l'enrichissement en motifs (GATA)<sub>n</sub> pourraient conférer au matériel génétique une résistance localisée à la propagation de l'extinction transcriptionnelle.

Il est aussi intéressant de signaler que pour certains gènes échappant à l'inactivation, un facteur supplémentaire pourrait intervenir. En effet, l'extrémité 5' des gènes *Jarid1c*, *eif2s3x* et *EIF2S3X* qui ont cette propriété, recrute la protéine CTCF, connue pour avoir une fonction insulatrice (Filippova *et al.*, 2005). Néanmoins, il faut souligner que la présence de sites de fixation de CTCF autour d'un transgène inséré sur le chromosome X n'est pas suffisante pour lui permettre d'échapper à l'inactivation (Ciavatta *et al.*, 2006). Le travail de Chaumeil *et al.* (2006) met en évidence que le gène *Jarid1c* est situé en bordure du domaine Xist lors de la mise en place de l'inactivation. On peut imaginer qu'une boucle de chromatine se forme et qu'elle permet d'exclure les gènes échappant à l'inactivation du domaine Xist. Toutefois, des études sur ce même gène *Jarid1c* mettent en évidence des variations de son taux d'expression au cours du développement (Lingenfelter *et al.*, 1998), ce qui n'est pas le cas lors de la différenciation de cellules ES (Chaumeil *et al.*, 2006). Ceci pourrait refléter des limitations intrinsèques au modèle ES. Par ailleurs, il est à noter que le caractère actif des gènes échappant à l'inactivation de l'X présente une certaine variabilité inter-individuelle (Sudbrack *et al.*, 2001).

## C. Mécanistique de l'inactivation aléatoire : caractérisation du centre d'inactivation du chromosome X : le XIC

### 1. Le centre de l'inactivation XIC : un locus complexe

Pour déclencher l'inactivation du chromosome X, la cellule doit tout d'abord évaluer le nombre de chromosomes et sélectionner un chromosome à inactiver. Puis, l'initiation de l'inactivation doit avoir lieu sur le chromosome X sélectionné, la répression transcriptionnelle se mettant ensuite en place. Toutes ces étapes dépendent d'un locus appelé XIC (X inactivation center). L'insertion du XIC sur un autosome est suffisante pour promouvoir l'inactivation de ce chromosome entier (Lee et Jaenisch, 1997 ; Lee *et al.*, 1996). Le XIC porte donc à lui seul les éléments nécessaires et suffisants au bon déroulement de l'inactivation du chromosome X.

La région XIC est longue d'environ 1 Mb. Elle est localisée à proximité de la bande D de chaque chromosome X (pour revue, Clerc et Avner, 2003). Plusieurs gènes ont été identifiés dans cette région : les gènes *Xist*, *Tsix*, *Xpct* (Debrand *et al.*, 1998), *Tsx* (Simmler *et al.*, 1996), *Chic1* (Simmler *et al.*, 1997), *Cdx4* (Gamer et Wright, 1993) et *Nap1l2* (Rogner *et al.*, 2000; Rougeulle et Avner, 1996), et les gènes *Cnbp2*, *Ftx*, *Jpx* et *Ppnx* (Chureau *et al.*, 2002).

Le XIC murin contient au total 11 gènes : *Xpct*, *Cnbp2*, *Ftx*, *Jpx*, *Xist*, *Tsx*, *Tsix*, *Chic1* (anciennement *Brx*), *Cdx4*, *Ppnx* et *Nap1l2* (anciennement *Bpx*). Sept d'entre eux codent des protéines (*Xpct*, *Ftx*, *Tsx*, *Chic1*, *Cdx4*, *Ppnx* et *Nap1l2*) mais ne semblent pas être impliqués

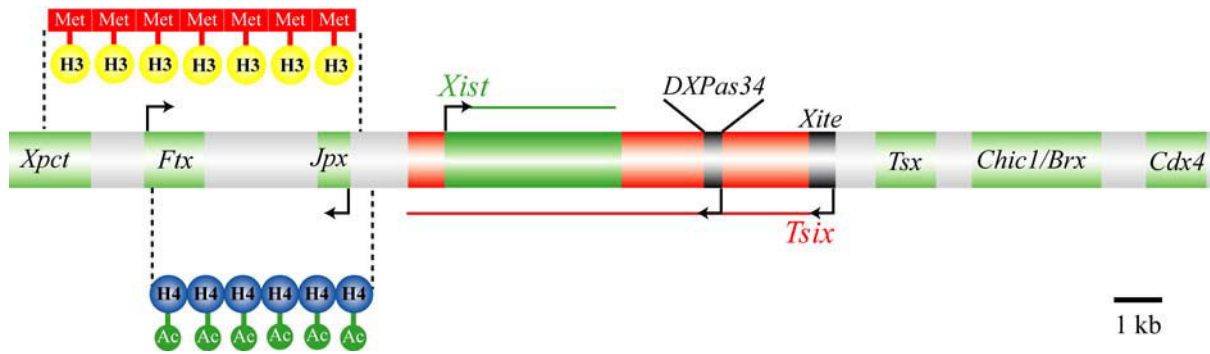
dans le processus d'inactivation. Quatre gènes (*Ftx*, *Jpx*, *Xist* et *Tsix*) produisent des ARN non-codants de longue taille. A ce jour aucune fonction n'a été associée à l'ARN transcrit à partir du gène *Ftx*. Par contre, dernièrement, il a été montré que le gène *Jpx* (sous sa forme chromatinienne) pourrait en interagissant avec le gène *Xist*, participer à la régulation de l'expression de ce dernier (Tsai *et al.*, 2008). Comme nous l'avons vu précédemment, *Xist* code un long ARN de même nom, crucial pour l'inactivation étant donné que son expression est le premier événement détectable lors de l'initiation de l'inactivation. L'ARN *Tsix*, synthétisé à partir du brin opposé à celui du gène *Xist*, régule l'expression de ce dernier. En effet, la diminution de l'expression de *Tsix* induit la sur-expression de *Xist* (Lee *et al.*, 1999).

Un autre locus important est la région *DXPas34*. Cette région minisatellite, longue d'environ 1,2 kb, est formée par un grand nombre de répétitions en tandem d'un motif de 34 pb. Elle est associée à un îlot CpG contenant le promoteur majoritaire de *Tsix* (Shibata et Lee, 2003 ; Vigneau *et al.*, 2006 ; Stavropoulos *et al.*, 2005 ; Cohen *et al.*, 2007).

Un élément supplémentaire appelé *Xite* a aussi été indentifié dans le locus *XIC*. Il correspond à un promoteur transcriptionnel d'où est initiée une forme longue de *Tsix* (Ogawa et Lee, 2003) (**Fig. 22**).

La caractérisation du locus *XIC* a permis de mettre en évidence qu'il est globalement déméthylé au niveau des CpG dans les cellules ES non différenciées et dans les embryons jusqu'au stade blastocyste (Avner *et al.*, 1998 ; Courtier *et al.*, 1995 ; Prissette *et al.*, 2001 ; Simmler *et al.*, 1993). Ensuite, les CpG sont méthylés, ceci aussi bien dans les tissus embryonnaires que dans ceux extra-embryonnaires (Courtier *et al.*, 1995 ; Prissette *et al.*, 2001).

Par des expériences d'immunofluorescence et d'immunoprécipitation de la chromatine du locus *XIC*, une accumulation de marques épigénétiques particulières a été détectée dans la région 5' du gène *Xist*. En effet, une région d'environ 340 kb qui s'étend de l'extrémité 3' du gène *Xpct* à celle du gène *Jpx* présente une concentration importante en H3K9me2, en H3K27me3, ceci quelque soit le sexe et le type cellulaire analysé (Heard *et al.*, 2001 ; Rougeulle *et al.*, 2004). De plus, une partie de cette région s'étendant du gène *Ftx* au gène *Jpx* est enrichie en H4 acétylée dans les cellules ES indifférenciées de femelles (O'Neill *et al.*, 1999). Il a été proposé que ces enrichissements représentent un centre de nucléation des marques épigénétiques sur le Xi, centre à partir duquel ces marques épigénétiques se propageraient tout au long du Xi au cours de la mise en place de l'inactivation du chromosome X (Heard *et al.*, 2001 ; Rougeulle *et al.*, 2004 ; O'Neill *et al.*, 1999 ; pour revue, Heard, 2004).



**Figure 22 : Centre d'inactivation du chromosome X chez la souris**

Seul le sens de transcription des gènes non-codants est indiqué. Le gène *Xist* est représenté en vert foncé et le gène *Tsix* en rouge. Les gènes *Ppnx*, *Cnbp2* et *Nap112* n'ont pas été représentés pour simplifier la figure. Ils sont localisés plus de 200 kb en amont de *Chic1/Brx1*. Les loci *Xite* et *DXPas34* sont indiqués en noir. La distribution des marques épigénétiques de la chromatine est représentée en rouge et jaune pour les méthylations de H3 et en vert et bleu pour les acétylations de H4.

D'après la revue Heard, 2005

## 2. Les étapes de l'inactivation aléatoire

La mise en place de l'inactivation aléatoire peut être divisée en cinq étapes : comptage, sélection d'un chromosome X, initiation, établissement et maintenance définitive de l'état silencieux. Il est admis que chez les mammifères, le comptage repose sur un mécanisme qui recense le ratio entre autosomes et chromosomes X et fournit une information permettant de mettre la mise en place d'une inactivation appropriée. En effet, en présence de plus de deux chromosomes X, un seul sera maintenu sous forme active. Les étapes d'initiation, d'établissement et de maintenance de l'état inactif qui suivent font intervenir les modifications de la chromatine, qui lorsque l'inactivation est établie se pérennisent de générations cellulaires en générations cellulaires. Comme nous l'avons décrit précédemment, le locus XIC est complexe et ses multiples fonctions régissent ces différentes étapes. Elles sont assurées par plusieurs éléments génétiques distincts. Nous décrirons les différentes étapes de l'inactivation aléatoire et leurs liens avec ces différents éléments du XIC.

### 2.1 Etapes de comptage et de sélection du chromosome X

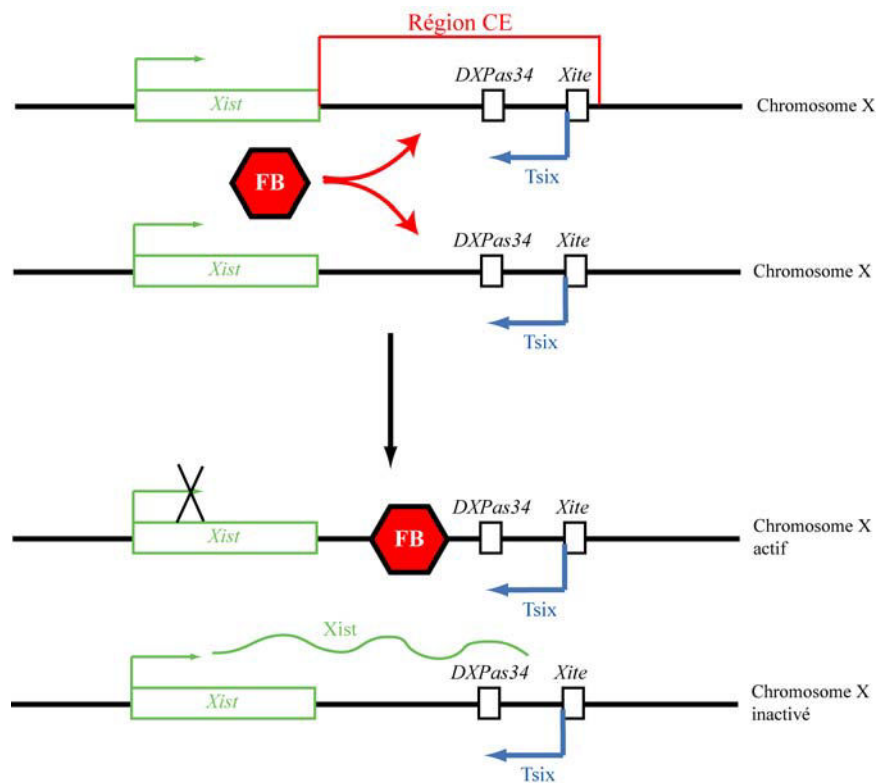
L'existence d'un mécanisme de recensement du nombre de chromosomes X est mis en évidence par le fait que les individus possédant un complément diploïde d'autosomes et un nombre variable de chromosomes X (i.e. AA:XXY ; AA:XXXX ; AA:XXXXY) ont toujours, et ceci indépendamment de leur sexe, un seul chromosome X actif. L'étape de comptage n'est donc pas un processus dépendant du sexe (Barr, 1963; Giannelli, 1963; Grumbach *et al.*, 1963). De plus, pour un génotype AA:XXXX, trois chromosomes X sont inactifs dans chaque cellule, alors que pour AAAA:XXXX, seulement 2 chromosomes X sont inactifs. Le

comptage des chromosomes consiste donc en la détermination 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).

### 2.1.1 Modèle mono-factoriel proposé pour les étapes de choix et comptage du chromosome X à inactiver

Un ensemble de données a servi de base pour essayer de construire un modèle de mécanisme de recensement du ratio chromosomes X/autosomes. Il a été observé que des translocations de fragments d'autosome sur l'X (X:autosome) conduisent à un blocage de l'inactivation (Rastan *et al.*, 1983), probablement parce que la symétrie existant entre les chromosomes X est rompue. Il y a alors inhibition de la sur-expression de *Xist* sur le chromosome X remanié qui n'est donc pas inactivé (Nicomedi *et al.*, 2007). Par ailleurs, une délétion de 65 kb en aval du gène *Xist* provoque la mise en place d'une inactivation inappropriée dans les cellules ne comportant qu'un seul chromosome X (XO, dérivés par sélection à partir d'une lignée femelle XX) (Clerc et Avner, 1998). Des réinsertions partielles du segment de 65 kb ont montré qu'un segment de 20 kb composé d'une partie du gène *Tsix* et des régions *Xite* et *DXPas34* est suffisant pour restaurer le phénomène de comptage et de choix. Ce segment joue donc un rôle majeur dans ce processus (Morey *et al.*, 2004). Il comporte les éléments *Xite* et *DXPas34* et une partie du gène *Tsix* et est appelé CE (pour Counting Element). Il pourrait fixer un facteur bloquant l'inactivation (BF). Comme ce facteur bloquant serait présent en quantité limitante, il ne pourrait se fixer que sur un seul chromosome X par cellule diploïde et protégerait ce chromosome X de l'inactivation. Le second chromosome X n'ayant pas fixé le facteur bloquant subirait alors l'inactivation. Dans ce cas, l'inactivation aléatoire serait due à l'équiprobabilité qu'aurait le facteur bloquant d'interagir avec l'un ou l'autre des chromosomes X et les processus de comptage et de choix seraient confondus (pour revue, Avner et Heard, 2001) (**Fig. 23**). Comme nous l'avons mentionné ci-dessus, le gène *Tsix* régule négativement l'expression du gène *Xist*. Or, la région CE contient les promoteurs de *Tsix*. Il est donc envisageable que le facteur bloquant régule positivement l'expression de *Tsix* et bloque l'inactivation du chromosome X à partir duquel il est exprimé. En accord avec cette hypothèse, le chromosome X portant une mutation abolissant la transcription de *Tsix* est toujours préférentiellement choisi pour être inactivé (Clerc et Avner, 1998 ; Lee et Lu, 1999 ; Luikenhuis *et al.*, 2001).





**Figure 23 : Modèle mono-factoriel proposé pour les étapes de comptage et de sélection du chromosome X à inactiver**

Le facteur bloquant (BF) en quantité limitante ne se fixerait que sur un seul chromosome X par cellule diploïde au niveau d'une région appelée CE (pour Counting Element). Cette fixation le protégerait de l'inactivation. Le second chromosome X n'ayant pas fixé le facteur bloquant subirait l'inactivation. L'inactivation aléatoire serait due à l'équiprobabilité qu'a le facteur bloquant d'interagir avec l'un ou l'autre des chromosomes X.

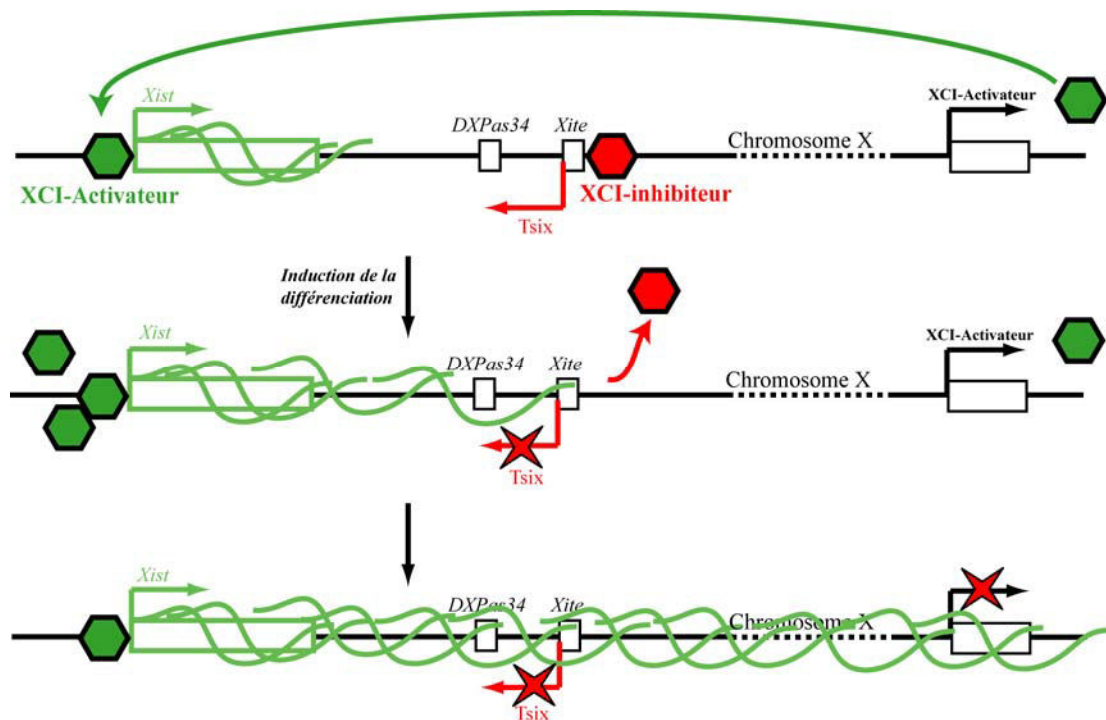
La région CE est localisée dans la région de 20 kb comprenant *Tsix*, *DXPas34* et *Xite*. Elle est indiquée par un crochet rouge sur le premier chromosome X schématisé.

D'après Avner et Heard, 2001 ; Morey *et al.*, 2004.

### 2.1.2 Modèle à deux facteurs proposé pour les étapes de comptage et choix du chromosome X à inactiver

A l'opposé des résultats obtenus dans les cellules (XO) (Clerc et Avner, 1998) que nous venons de décrire, la délétion du promoteur *Tsix* dans les cellules mâles (XY) n'entraîne pas l'inactivation du seul chromosome X présent (Lee and Lu, 1999). Or le modèle mono-factoriel prédit l'inactivation du chromosome X unique dans les cellules mâles. Aussi, pour expliquer ces observations, un facteur supplémentaire appelé facteur de compétence, exprimé uniquement dans les cellules femelles, a été introduit (Lee et Lu, 1999). Ce modèle à deux facteurs implique que l'inactivation n'est plus initiée par défaut : le processus de comptage engendre l'inactivation quand plus d'un X est présent par lot diploïde d'autosomes. Dans ce modèle, la quantité de facteur de compétence, produite par un seul chromosome X, serait titrée par la quantité de facteur bloquant produite par un complément diploïde d'autosomes ce qui empêcherait l'inactivation de se produire sur le chromosome X (**Fig. 24**). Des arguments en faveur de ce modèle ont été apportés très récemment par l'équipe de Gribnau. Selon cette

équipe, l'activateur du processus d'inactivation du chromosome X (XCI-activateur) stimulerait l'expression de *Xist* et son expression serait éteinte lors de l'inactivation, alors que le facteur bloquant ou inhibiteur de l'inactivation (XCI-inhibiteur) stimulerait l'expression de *Tsix* (Monkhorst *et al.*, 2008 ; Monkhorst *et al.*, 2009) (**Fig. 24**). Dans les cellules ES de génotype tétraploïde XXXX, l'initiation de l'inactivation du chromosome X est beaucoup plus importante que dans les cellules XXXY ce qui est en faveur de l'existence d'un facteur XCI-activateur codé par le chromosome X (Monkhorst *et al.*, 2008).



**Figure 24 : Modèle à deux facteurs proposé pour les étapes de comptage et de sélection du chromosome X à inactiver**

Avant différenciation, la transcription de *Tsix* réprime l'expression de *Xist* et la concentration du XCI-activateur est insuffisante pour permettre la répression de *Tsix*. Après différenciation, la concentration du XCI-activateur augmente pour atteindre un niveau requis pour initier l'inactivation. L'ARN *Xist* se propage et provoque l'extinction transcriptionnelle de *Tsix*, puis du gène codant le XCI-activateur. Cette dernière répression éviterait l'inactivation du second chromosome X.

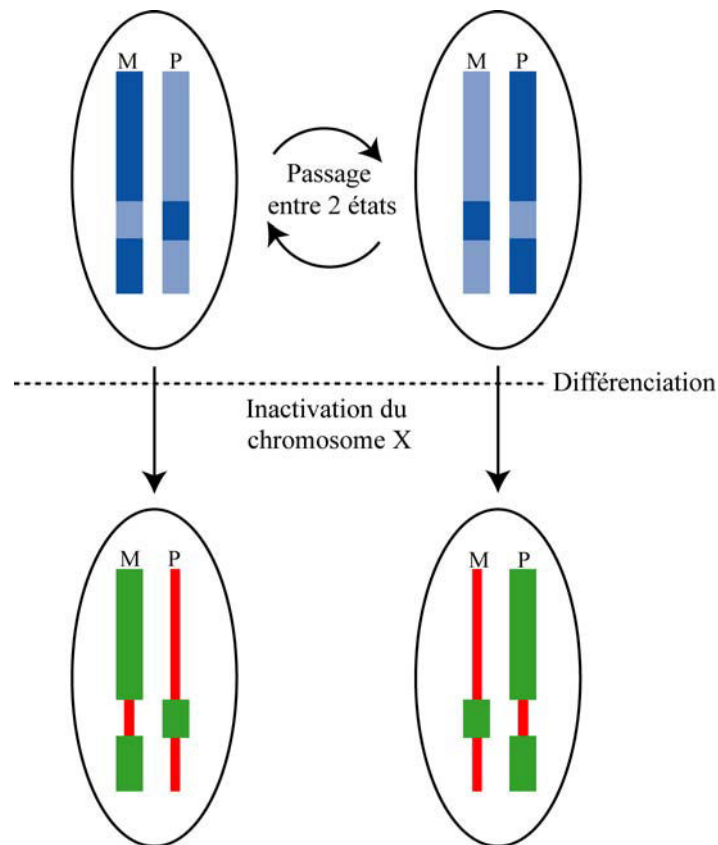
D'après Monkhorst *et al.*, 2008 ; Monkhorst *et al.*, 2009

Comme le chromosome exprimant *Tsix* deviendra le futur Xa et celui exprimant *Xist* sera le Xi, les régulateurs de l'expression de ces gènes doivent jouer un rôle fondamental dans l'inactivation aléatoire du chromosome X. Comme nous l'avons présenté ci-dessus, la transcription de *Tsix* est régulée par les loci *DXPas34* et *Xite* (Cohen *et al.*, 2007 ; Debrand *et al.*, 1999 ; Ogawa et Lee, 2003 ; Stavropoulos *et al.*, 2005 ; Vigneau *et al.*, 2006). La région *DXPas34* fixe la protéine insulatrice de la chromatine appelée CTCF, citée ci-dessus, et son partenaire, Yy1, qui est un activateur transcriptionnel (Chao *et al.*, 2002 ; Donohoe *et al.*,

2007). Ce complexe formé induit la transcription de *Tsix* et semble également bloquer l'accès aux activateurs de *Xist*, donc pourrait contribuer au choix du Xa. Ces résultats confirment le modèle de comptage et de choix du chromosome X à inactiver présenté ci-dessus.

### 2.1.3 Modèle de choix aléatoire basé sur un passage continu entre 2 états des 2 chromosomes X

Le modèle à deux facteurs peut être considéré comme stochastique étant donné que tous les chromosomes ont une probabilité équivalente de s'inactiver et que c'est la fixation aléatoire des facteurs activateurs et inhibiteurs qui guiderait le choix et le comptage (Monkhorst *et al.*, 2008 ; Monkhorst *et al.*, 2009). Certaines données vont néanmoins à l'encontre de ce modèle. Elles suggèrent que les chromosomes X alternent entre 2 états avant l'inactivation du Xi et seraient donc susceptibles de communiquer entre eux (Mlynarczyk-Evans *et al.*, 2006). Il a été démontré que les 2 chromosomes X ont des états différents avant l'inactivation. Des expériences de DNA-FISH réalisées sur différents *loci* et gènes du chromosome X, dont le XIC, dans des lignées sauvages de cellules ES, ont permis de mettre en évidence l'existence d'un signal ponctuel simple sur un allèle, et d'un signal dédoublé sur le second. Cet état est caractéristique des gènes liés aux chromosomes X et cette configuration « simplet-doublet » (SD) particulière a été nommée SIAR pour *SD "signals that are Independent of Asynchronous DNA Replication"* (Mlynarczyk-Evans *et al.*, 2006). De façon intéressante, l'analyse des SIAR par DNA-FISH dans les lignées mutées pour *Xist* et *Tsix* reflète leurs implications dans le contrôle de l'inactivation et suggère l'existence d'une relation entre l'apparence des chromosomes X observée par FISH avant l'inactivation et leurs devenir en tant que Xa ou Xi. Il a aussi été observé qu'en début d'inactivation dans les cellules ES femelles, les gènes d'un chromosome X sont successivement détectés sous la forme d'un signal unique puis d'un doublet. L'ensemble de ces observations a conduit à la proposition d'un modèle selon lequel les mécanismes de comptage et de choix sont en permanence actifs dans les cellules ES femelles indifférenciées et ceci avec un profil SIAR opposé du XIC et du reste des gènes du même chromosome X. Le devenir de chaque chromosome X dépend de son état lorsqu'un signal développemental induit l'initiation de l'inactivation (Mlynarczyk-Evans *et al.*, 2006) (**Fig. 25**).



**Figure 25 : Modèle de choix aléatoire basé sur un passage continu entre 2 états des 2 chromosomes X**

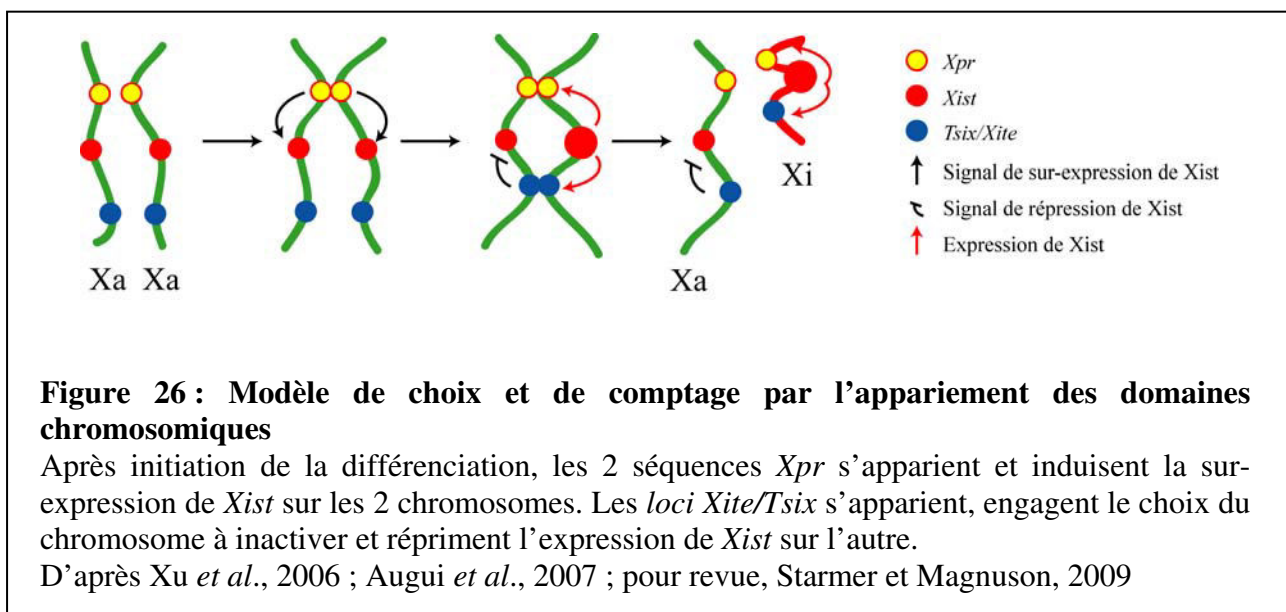
Sur la base de l'alternance des états SIAR, Mlynarczyk-Evans *et al.*, suggèrent un modèle selon lequel, dans les cellules indifférenciées, les chromosomes X alternent de manière coordonnée entre un état de Xa présomptif (chromosome bleu clair) et un état de Xi présomptif (chromosome bleu foncé). Le devenir des chromosomes dépend de leur état au moment où un signal développemental initie l'inactivation. Le centre d'inactivation du chromosome X présente également une alternance de configuration SIAR, mais elle est opposée à celle du reste du chromosome X, d'où sa représentation par une bande de couleur différente de celle du chromosome qui le porte.

D'après Mlynarczyk-Evans *et al.*, 2006

#### 2.1.4 Rôle potentiel de l'appariement des 2 chromosomes X dans le processus d'inactivation

Certaines données suggèrent que ces étapes de comptage et de choix ou que les étapes ultérieures pourraient impliquer une communication passagère entre les 2 chromosomes X (pour revue, Marahrens *et al.*, 1999). En effet, un rapprochement des chromosomes X est observé au cours de la différenciation de cellules ES femelles, dans la période d'initiation du recouvrement du chromosome X par Xist (Bacher *et al.*, 2006 ; Xu *et al.*, 2006). Très précisément, vers 1,5 à 4 jours de différenciation est détecté un rapprochement transitoire des régions XIC (Bacher *et al.*, 2006 ; Xu *et al.*, 2006). Il a aussi été observé que le XIC est inséré en multicopie dans un chromosome artificiel (YAC) de 460 kb, le YAC se rapproche transitoirement du XIC endogène au cours de la différenciation des cellules ES mâles *in vitro* (Bacher *et al.*, 2006). L'idée que l'appariement entre les 2 chromosomes X ait un lien fonctionnel avec le processus d'inactivation est renforcée par le fait que dans des cellules ES

femelles dont un des deux X porte une délétion de 65 kb en aval du gène *Xist*, ce qui abolit l'inactivation aléatoire (Clerc et Avner, 1998), il n'y a plus d'appariement entre les deux XIC. La réinsertion dans ces lignées mutantes de l'unité promotrice de *Tsix*, incluse dans les 65 kb délétés, permet de rétablir ce processus, suggérant un rôle fondamental de cette séquence promotrice et/ou de la transcription de *Tsix* dans la colocalisation des deux XIC. La réinsertion de cette séquence ne suffit pas néanmoins à recouvrir un choix aléatoire de l'X à inactiver (Xu *et al.*, 2006). Des analyses plus fines des éléments impliqués dans l'appariement entre les 2 régions XIC ont permis de démontrer que les loci *Tsix/Xite* (Xu *et al.*, 2006) et 1,6 kb de la région *DXPas34* sont capables de s'apparier avec leurs régions homologues. Cet appariement semble dépendre de la fixation de la protéine CTCF, car son extinction abolit l'appariement (Xu *et al.*, 2007). Il est donc proposé qu'à la fois CTCF et les éléments *Tsix* et *Xite* (et probablement d'autres facteurs) sont nécessaires au mécanisme d'appariement des chromosomes X (pour revue, Payer *et al.*, 2008). Une région de 200 kb, localisée en amont du gène *Xist* (« X-pairing region », Xpr), pourrait participer à l'appariement des 2 chromosomes X dans les cellules non différenciées juste avant ou pendant l'initiation de l'inactivation (Augui *et al.*, 2007). Il a été proposé que l'appariement de cette région intervienne avant celui des loci *Tsix/Xite*, toutefois son rôle dans le processus d'inactivation du chromosome X reste à définir (Fig. 26).



Il est encore difficile de savoir si la régulation de l'expression du gène *Xist* intervient dans les étapes de choix et de comptage. En effet, les délétions de *Xist* tendent à provoquer l'inactivation préférentielle du chromosome X sauvage (Penny *et al.*, 1996 ; Gribnau *et al.*, 2005). Toutefois, ces délétions peuvent soit altérer les processus de comptage et de choix ou entraîner une incapacité du chromosome muté à engager l'inactivation *in cis*.

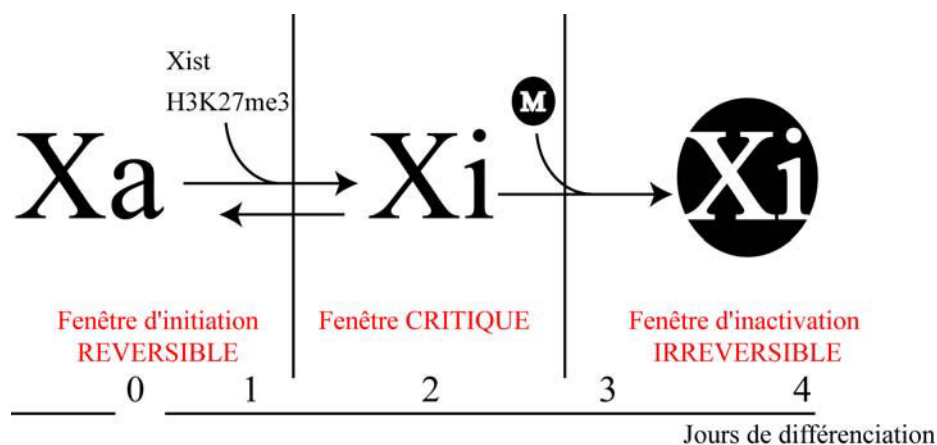
## 2.2 L'initiation

Ainsi comme nous l'avons déjà mentionné, après sa sur-expression, l'ARN *Xist* se propage sur le futur Xi. Cette sur-expression est le point de départ d'une cascade d'événements permettant la mise en place de marques épigénétiques essentielles à la stabilité de l'extinction transcriptionnelle.

### 2.2.1 Fenêtre d'opportunité de l'inactivation aléatoire

Plusieurs études ont permis de montrer que l'extinction transcriptionnelle médiée par l'ARN *Xist* ne peut se dérouler que dans une fenêtre de temps très stricte faisant suite à la différenciation cellulaire. Elle est restreinte aux cellules de l'embryon précoce (Savarese *et al.*, 2006). Dans beaucoup de cellules différenciées, dans lesquelles l'état silencieux du Xi est stable, l'expression de *Xist* n'a aucun effet (Sado *et al.*, 2004 ; Savarese *et al.*, 2006 ; Wutz et Jaenisch, 2000). L'étude des cellules ES en voie de différenciation a apporté de ce fait beaucoup d'informations sur l'initiation de l'inactivation. C'est en introduisant des transgènes d'ADNc *Xist* sous contrôle d'un promoteur inductible dans des cellules ES mâles qu'il a pu être montré que la sur-expression de *Xist* induit l'inactivation du chromosome porteur du transgène dans les cellules ES indifférenciées. Après différenciation, cette sur-expression ne provoque aucun phénotype particulier (Wutz et Jaenisch, 2000). De plus, l'induction de l'expression du transgène *Xist* à différents stades de différenciation a permis de définir avec précision la fenêtre d'opportunité au cours de laquelle *Xist* est capable d'induire l'inactivation du chromosome X. Il s'agit d'une période allant de l'état non différencié, jusqu'aux 24 premières heures après initiation de la différenciation. La répression transcriptionnelle initiale est réversible et dépend de l'expression de *Xist*. En revanche, l'état inactif du chromosome X devient irréversible après 48 h de différenciation et ceci même après arrêt de l'expression de *Xist* (Wutz *et al.*, 2002). Il reste donc une fenêtre dite critique de 24 h pendant laquelle la cellule a la capacité de consolider ou non l'inactivation. Comment la répression génique est maintenue durant cette période et comment l'état silencieux devient irréversible ? Afin de répondre à ces questions, des analyses fines de la mise en place des marques épigénétiques telles que les H3K27me3, en présence ou en absence de l'ARN *Xist* et à différents temps après l'initiation de la différenciation ont été réalisées (Kohlmaier *et al.*, 2004). La mise en place des H3K27me3 est rapidement observée après induction de l'expression de *Xist* et disparaissent 48 h après l'extinction de l'expression de *Xist*. Et de façon intéressante, quand *Xist* est ré-exprimé tardivement, la méthylation H3K27me3 est efficacement restaurée uniquement si *Xist* a été préalablement exprimé au moins durant les premières 48 à 72 h de différenciation, c'est-à-dire durant la fenêtre d'initiation et la fenêtre critique (Kohlmaier *et al.*, 2004). Pour expliquer ces observations, les auteurs ont proposé la notion de mémoire chromosomique qui est établie lors de l'expression de *Xist* dans un stade développemental

bien particulier correspondant aux fenêtres d'initiation et critique (**Fig. 27**) (Kohlmaier *et al.*, 2004). L'identification de facteurs protéiques exprimés durant cette fenêtre de temps pouvait apporter des informations. Dans cette optique, Agrelo *et al.* (2009) ont proposé que les protéines SATB1 et SATB2 puissent être impliquées. En effet, ces protéines sont exprimées durant cette période, elles co-localisent avec l'ARN *Xist* dans des lymphocytes et leur expression induit la répression transcriptionnelle médiée par *Xist* dans des fibroblastes embryonnaires qui ne sont normalement pas le siège de l'initiation de l'inactivation du chromosome X (Agrelo *et al.*, 2009). De plus, la répression des protéines SATB1 et SATB2 par ARN interférence dans les cellules ES diminue l'efficacité d'extinction transcriptionnelle de respectivement 20 et 26%. Néanmoins, la disruption du gène SATB1 chez la souris reste compatible avec un développement embryonnaire normal (Alvarez *et al.*, 2000). Il a été suggéré que ces protéines SATB participeraient à l'initiation de l'inactivation du chromosome X dans les cellules embryonnaires et ceci en association avec d'autres partenaires encore inconnus.



### Figure 27 : Fenêtre d'opportunité de l'initiation de l'inactivation du chromosome X

Les jours de différenciation sont indiqués sur la flèche de temps. Avant initiation de la différenciation, les chromosomes X sont actifs (Xa). L'inactivation du chromosome X dans ces cellules consécutives à la sur-expression de l'ARN *Xist* est réversible. Après initiation de la différenciation des cellules ES, l'expression de *Xist* induit une apparition des modifications H3K27me3 qui induit l'extinction génique (Xi). Durant cette phase dite de fenêtre critique, la cellule a la capacité de consolider ou non l'inactivation. Ensuite, après 48 heures de différenciation, une mémoire chromosomique est mise en place (lettre M dans l'ovale noir) grâce aux marques épigénétiques H3K27me3 et à *Xist*. Cette mémoire est complètement mise en place sur le Xi et l'inactivation devient irréversible et indépendante de *Xist*.

D'après Kohlmaier *et al.*, 2004

#### 2.2.2 Propagation de l'ARN *Xist*

Une autre question non résolue du processus d'inactivation du chromosome X est de savoir comment l'ARN *Xist* est fixé sur le chromosome X et comment il se propage le long de la chromatine et ceci spécifiquement sur le chromosome X à partir duquel il est transcrit.



Aucun modèle clair n'a été établi mais plusieurs observations ont permis de formuler des hypothèses.

L'expression ectopique d'une portion du chromosome X contenant *Xist* sur un autosome conduit au recouvrement de celui-ci par *Xist* et à l'extinction des gènes qui lui sont liés. Toutefois, cette extinction est fortement atténuée comparée à celle observée sur le chromosome X inactif (Duthie *et al.*, 1999 ; Rastan, 1983 ; Shao et Takagi, 1991). Ceci suggère l'existence d'une certaine spécificité pour le chromosome X soumis à l'inactivation médiée par *Xist*. Néanmoins, lorsque des transgènes du gène *Xist* sont insérés en multicopie dans un autosome, l'inactivation de ce chromosome semble être complète. Donc, l'expression de *Xist* à des niveaux supérieurs au niveau physiologique permettrait un recouvrement efficace de l'autosome par *Xist* et donc son inactivation (Lee *et al.*, 1996 ; Wutz et Jaenisch, 2000). Ces données sont en accord avec le fait 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 ce qui n'entraîne qu'une faible extinction transcriptionnelle (Popova *et al.*, 2006 ; Giorda *et al.*, 2008). L'absence de propagation de *Xist* à longue distance pourrait s'expliquer par le fait qu'il existerait sur le chromosome X des centres « relais » ou « tremplins » (*way stations* et *booster elements*, dans les termes des auteurs) favorisant la propagation du chromosome X (pour revue, Gartler et Riggs, 1983). En l'absence de ces centres relais, la sur-expression de *Xist* à un niveau élevé non physiologique est nécessaire pour obtenir une forte inactivation. M Lyon a proposé que les éléments répétés LINE-1 (L1) soient des cibles de haute affinité pour la fixation de l'ARN *Xist* (pour revue, Lyon, 1998). En effet, dans les cellules humaines et murines, la quantité des éléments LINE-1 sur le chromosome X est deux fois supérieure à celle retrouvée sur les autosomes. Elle est tout particulièrement élevée autour du gène *XIST* et, à l'inverse, conforme à la moyenne du génome au niveau des gènes échappant à l'inactivation (Bailey *et al.*, 2000 ; Ross *et al.*, 2005 ; Allen *et al.*, 2003 ; Chureau *et al.*, 2002; Nesterova *et al.*, 2001). Comme mentionné ci-dessus, dans les cellules ES portant la translocation du XIC sur le chromosome 4, la propagation de l'ARN *Xist* s'arrête au point de cassure. Or la région de 20 Mb bordant ce point de cassure est exceptionnellement pauvre en éléments LINE-1 (Popova *et al.*, 2006). Cet appauvrissement pourrait donc être impliqué dans ce défaut de propagation.

Récemment, l'utilisation de lignées cellulaires diploïdes et tétraploïdes exprimant l'ARN *Xist* fusionné à 16 motifs MS2, a permis l'analyse dynamique de la localisation intranucléaire de l'ARN *Xist*, ceci avec distinction de l'origine allélique de cet ARN. Cette analyse a apporté une preuve supplémentaire que l'ARN *Xist* reste sur le Xi en *cis* et qu'aucun échange entre chromosomes X n'est possible. Ce maintien de *Xist* sur le chromosome à partir duquel il est exprimé pourrait s'expliquer par la forte concentration en éléments LINE-1 autour du gène *Xist* (Jonkers *et al.*, 2008). Cela n'exclut néanmoins pas que

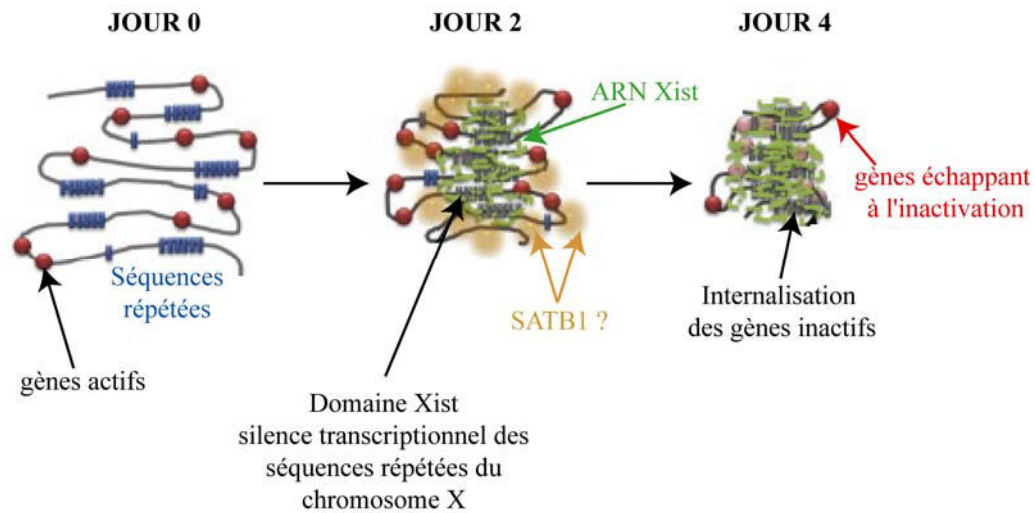


d'autres régions d'ADN ou des marques épigénétiques spécifiques du chromosome X soient impliquées dans la propagation de l'ARN Xist. Toutefois, toutes ces données renforcent l'idée d'un rôle important des éléments LINE-1 dans son ancrage sur le chromosome X.

### 2.2.3 Le domaine Xist

Lorsque l'ARN Xist recouvre le futur Xi, ce dernier adopte une conformation condensée extrêmement stable. Il a été montré que l'interaction chromatine/ARN permet la formation d'un compartiment nucléaire en général répressif pour la transcription. Il a été proposé que les éléments de la matrice nucléaire interviennent dans la formation de ce compartiment. Notons néanmoins que l'existence d'une matrice nucléaire est controversée par certains auteurs. L'implication de composants de la matrice de l'ADN repose sur le fait qu'après hydrolyse par des DNases et extraction saline, l'ARN XIST reste en place dans le noyau, alors que lorsque la matrice nucléaire est extraite, l'ARN XIST y est retrouvé (Clemson *et al.*, 1996). L'idée d'un rôle de la matrice nucléaire dans la localisation de l'ARN XIST est confortée par le fait que la protéine SAF-A est enrichie sur le Xi et que cet enrichissement dépend de son domaine d'interaction avec les ARN, ce qui suggère une interaction fonctionnelle possible avec XIST (Fackelmayer, 2005 ; Helbig et Fackelmayer, 2003).

Au cours de son inactivation, l'organisation du futur Xi évolue. En effet, le Xi se structure avec formation d'un corps interne de séquences répétées, entouré par une bordure de gènes, le tout recouvert par l'ARN Xist. Ce compartiment nucléaire répressif enrichi en H3K27me3 est appelé domaine Xist (Chaumeil *et al.*, 2006 ; Clemson *et al.*, 2006). Les gènes échappant à l'inactivation sont situés à l'extérieur de ce domaine. Le modèle suivant a été proposé pour la mise en place de ce domaine : l'événement le plus précoce après l'habillage du chromosome X par l'ARN Xist est l'exclusion de l'ARN polymérase II et l'extinction transcriptionnelle de la fraction riche en séquences répétées du chromosome X (Chaumeil *et al.*, 2006). La formation de ce compartiment transcriptionnellement silencieux a lieu avant même l'extinction des gènes. La relocalisation des gènes dans le domaine Xist permet ensuite l'extinction transcriptionnelle efficace (**Fig. 28**). Nous verrons par là suite, que cette relocalisation est dépendante de la région des A-repeats, région de l'ARN Xist cruciale pour l'inactivation. Les protéines SATB1 et SATB2 que nous avons mentionnées dans la partie précédente pourraient avoir un rôle dans ce processus. Il a en effet été montré dans un autre système, les thymocytes, qu'elles sont impliquées dans la régulation de la stucturation de la chromatine et la régulation des gènes par réorganisation nucléaire de séquences chromosomiques (Alvarez *et al.*, 2000 ; Cai *et al.*, 2003). SATB1 et SATB2 pourraient donc intervenir de manière similaire lors de la relocalisation des gènes inactifs du Xi à l'intérieur du domaine Xist (Agrelo *et al.*, 2009 ; pour revue, Chow et Heard, 2009) (**Fig. 28**).



**Figure 28 : Modèle proposé pour la formation du domaine Xist**

Dans les cellules femelles non différenciées, les chromosomes X sont toujours actifs. Après différenciation, l'ARN Xist recouvre le Xi et se structure selon un corps interne de séquences répétées transcriptionnellement inactives, qui est entouré par un ensemble de gènes. Les gènes destinés à être inactifs sont ensuite relocalisés dans le domaine Xist et l'extinction transcriptionnelle devient alors efficace. Les protéines SATB1 et SATB2 pourraient être impliquées dans cette relocalisation.

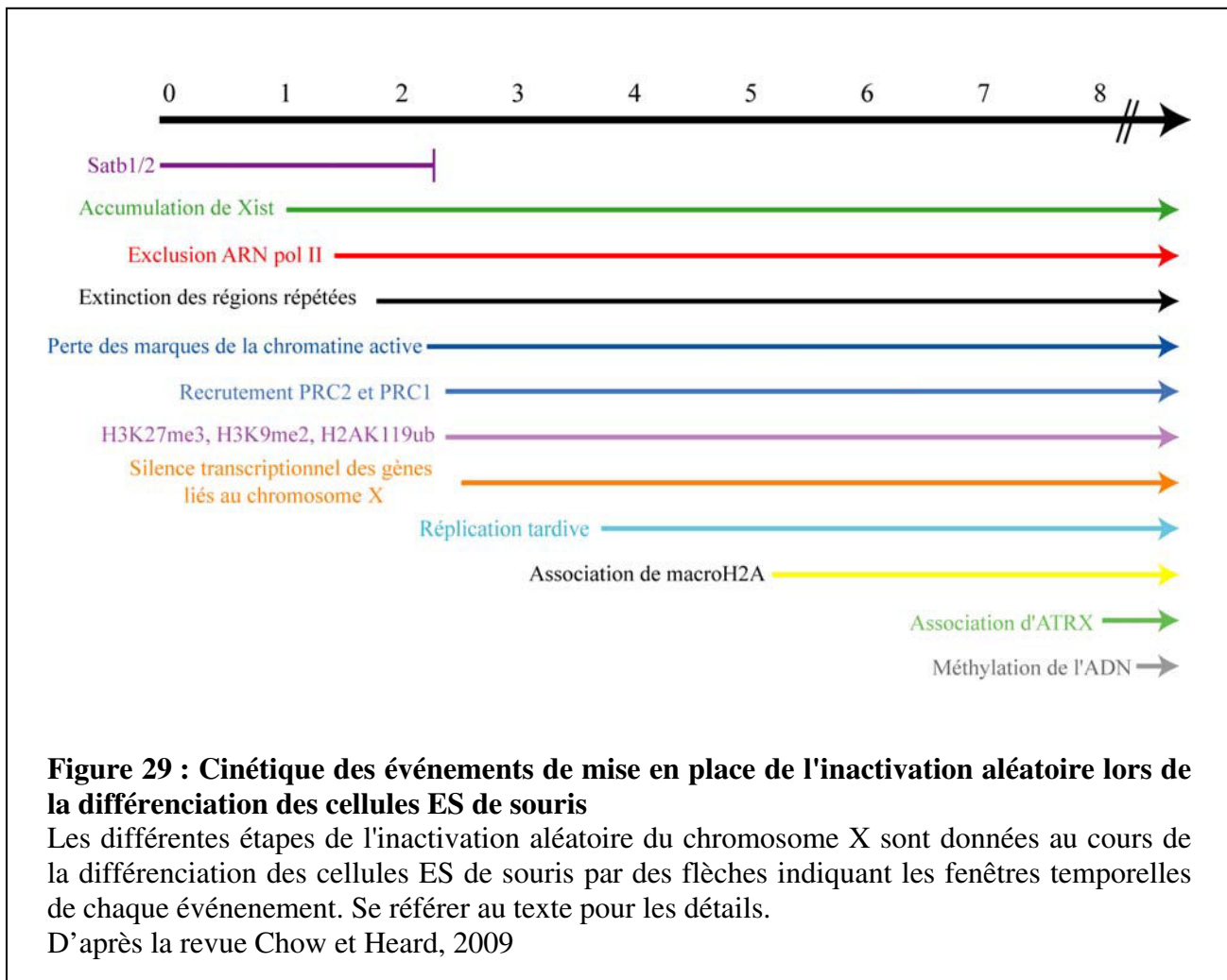
D'après la revue Chow et Heard, 2009

### 2.3 La mise en place des marques épigénétiques et maintenance de l'état réprimé

Nous avons décrit précédemment les étapes précoces de l'inactivation du chromosome X. Afin de comprendre comment sont mises en place les marques épigénétiques et comment est maintenu l'état réprimé, nous allons présenter dans ce paragraphe la cinétique des événements de l'inactivation aléatoire lors de la différenciation des cellules ES.

Les données récentes suggèrent que la protéine SATB1 est exprimée durant la fenêtre d'opportunité de l'inactivation, c'est-à-dire de l'état indifférencié jusqu'à 48 à 72 h après l'initiation de la différenciation. Cette protéine est donc susceptible d'être impliquée dans les étapes précoces d'inactivation (Kohlmaier *et al.*, 2004 ; Alegro *et al.*, 2009). Dès un jour après stimulation de la différenciation, l'ARN Xist est sur-exprimé et son accumulation sur le futur Xi permet la formation du domaine Xist et la relocalisation des gènes à inactiver à l'intérieur de ce domaine. Cet événement est suivi rapidement par une cascade d'événements, le premier étant l'exclusion de l'ARN polymérase II (Plath *et al.*, 2003 ; de la Cruz *et al.*, 2005b). Les séquences répétées du Xi sont réprimées et les marques de l'euchromatine, telles que les H3K9ac, H3K4m2, H3K4m3, et l'acétylation globale de H4 sont perdues (Keohane *et al.*, 1996 ; Heard *et al.*, 2001 ; Chaumeil *et al.*, 2002). Dès 48 h après le début de la différenciation, les complexes PRC2 et PRC1 sont recrutés et apposent les H3K27m3 (Rougeulle *et al.*, 2004) et aussi les H3K9me2 (Heard *et al.*, 2001 ; Okamoto *et al.*, 2004), H3K9me3, H4K20m1 (Kohlmaier *et al.*, 2004), et les H2AK119ub1 (De Napolés *et al.*, 2004 ; Hernandez-Munoz *et al.*, 2005). Puis le futur Xi s'associe aux variants d'histone

macroH2A, et à la protéine ATRX et finalement, les îlots CpG sont méthylés (pour revue, Chow et Heard, 2009) (**Fig. 29**).



Lorsque l'état inactif est établi, il faut que celui-ci soit maintenu tout au long des divisions cellulaires. Certaines des modifications épigénétiques détectées au cours de l'inactivation du chromosome X sont d'ailleurs aussi détectées sur le Xi dans des cellules métaphasiques (Boggs *et al.*, 2002 ; Chadwick et Willard, 2004 ; Chaumeil *et al.*, 2002 ; Mak *et al.*, 2002 ; Mermoud *et al.*, 2002 ; Peters *et al.*, 2002). L'état d'inactivation du Xi est très stable, car même le traitement de cellules somatiques femelles par des combinaisons de drogues, comme la TSA qui inhibe l'acétylation des histones, et la 5-azacytidine qui inhibe la méthylation des CpG, associées à la délétion conditionnelle de *Xist* n'entraîne pas la perte de macroH2A et la réactivation du gène *Hprt* lié au chromosome X est détectée seulement dans un faible nombre de cellules (de l'ordre de 5%) (Csankovszki *et al.*, 2001).

Étant donné que des cinétiques différentielles d'apparition pour les différentes marques épigénétiques de l'X sont observées au cours de l'inactivation, il est admis qu'elles jouent des rôles distincts dans ce processus. Ainsi, les marques précoces pourraient participer à la mise en place de l'état inactif, alors que l'incorporation de macroH2A et la méthylation

des îlots CpG permettraient le verrouillage de la chromatine dans un état inactif (Kohlmaier *et al.*, 2004).

De plus, comme déjà mentionné, la 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), ne résultent pas en une réactivation globale du Xi, ceci bien que la localisation de macroH2A (Csankovszki *et al.*, 1999) et l'enrichissement en H3K27me3 (Zhang *et al.*, 2007) disparaissent du Xi et que certains gènes soient réactivés (Csankovszki *et al.*, 2001 ; Zhang *et al.*, 2007).

Par contre, il est clair que la méthylation de l'ADN est un paramètre clé pour le maintien de l'état répressif, car la délétion du gène *Dnmt1* codant l'enzyme responsable de cette méthylation provoque la réactivation du chromosome X dans l'embryon. Il faut noter que cette délétion n'a par contre aucun effet sur l'inactivation soumise à l'empreinte parentale. Ceci suggère que l'étape de maintenance diffère selon le type d'inactivation (Sado *et al.*, 2000).

Des données récentes suggèrent un rôle de la protéine smcHD1 dans l'inactivation aléatoire du chromosome X car dans des embryons mutés dans le gène *smcHD1*, la méthylation des îlots CpG des gènes liés à l'X et la répression des gènes sont perturbées. De plus, la maintenance des deux types d'inactivations (soumise à empreinte parentale et aléatoire) est perturbée, ce qui montre son importance (Blewitt *et al.*, 2006 ; Blewitt *et al.*, 2008).

### 3. Régulation par le gène *Xist*

#### 3.1 Structure du gène *Xist*

Découvert en 1991 chez l'homme (*XIST*), puis chez la souris (*Xist*), le gène *Xist* (*X inactive specific transcript*) est le premier gène à avoir été caractérisé comme n'étant exprimé qu'à partir du chromosome X inactif (Xi) dans les cellules somatiques (Borsani *et al.*, 1991 ; Brown *et al.*, 1991), ce qui a rapidement suggéré son implication dans l'inactivation du chromosome X (Penny *et al.*, 1996).

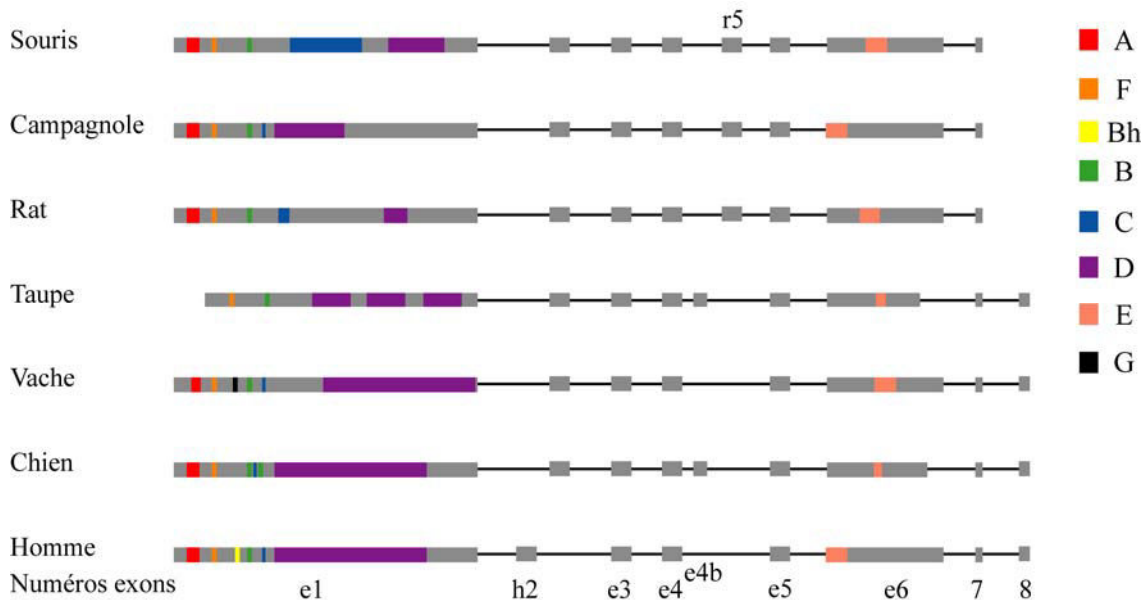
Les séquençages des gènes *Xist* de différents mammifères ont permis d'en établir l'architecture. Le gène *Xist* est long de 35 kb et huit exons ont été caractérisés chez l'homme (Brockdorff *et al.*, 1992 ; Brown *et al.*, 1992 ; Chureau *et al.*, 2002), la vache (Chureau *et al.*, 2002), la souris (Chureau *et al.*, 2002 ; Sheardown *et al.*, 1997b ; Simmler *et al.*, 1996), le campagnol (Nesterova *et al.*, 2001). Deux de ces exons (1 et 7) sont très longs (l'exon 1 d'environ 9 kb et exon 7 de 6 à 7 kb). Les séquences récemment établies pour la taupe, le rat et le chien montrent que la majorité des exons sont conservés entre les espèces, à l'exception de l'exon 2 chez l'homme et de l'exon 5 chez les rongeurs et révèlent l'existence d'un nouvel exon nommé "4 b" chez le chien et la taupe (Yen *et al.*, 2007) (**Fig. 30**). La conservation de

séquence des exons est importante : entre la souris (*Xist*) et l'homme (*XIST*) (66%), entre la souris et la vache (62%). Par contre, la séquence des introns est peu conservée (Chureau *et al.*, 2002).

Trois promoteurs contrôlent l'expression du gène *Xist*. Ils ont été désignés P0 (6 kb en amont des deux autres), P1 et P2. Dans les cellules somatiques différenciées, environ 75 % des transcrits sont issus du second promoteur (Johnston *et al.*, 1998). Nous y reviendrons par la suite. Récemment, une activité promotrice supplémentaire a été mise en évidence. En effet, 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). Nous reviendrons aussi sur ce point plus tard.

L'ARN *Xist* transcrit à partir des promoteurs P0, P1 ou P2 peut ne pas être épissé et comporte environ 35000 nt et 17000 nt après avoir été épissé. Il existe d'autres variants de *Xist* qui résultent soit d'une terminaison précoce de la transcription, soit de mécanismes d'épissage alternatif. Les variants de terminaison ne contiennent pas l'exon 8 chez la souris, et les exons 7 et 8 chez l'homme (Brockdorff *et al.*, 1992; Hong *et al.*, 1999 ; Hong *et al.*, 2000). Les variants d'épissage sont modifiés à l'extrémité 3' et sont raccourcis d'environ 5 kb par rapport aux aux ARN de 17000 nt (Chureau *et al.*, 2002).

L'ARN *Xist* est composé de plusieurs régions contenant des séquences répétées en tandem. Certaines d'entre elles sont extrêmement divergentes d'une espèce à l'autre alors que d'autres sont fortement conservées (Clemson *et al.*, 1996 ; Nesterova *et al.*, 2001) (**Fig. 30**). Chez la souris, 5 de ces régions contenant des éléments répétés sont localisées dans l'exon 1, la 6<sup>ième</sup> dans l'exon 7. Elles sont nommées de A à F (Brown *et al.*, 1992 ; Brockdorff *et al.*, 1992 ; Nesterova *et al.*, 2001 ; Brockdorff *et al.*, 2002). L'exon 1 chez la vache et l'homme renferme des répétitions supplémentaires (régions Bh et G) (Yen *et al.*, 2007). Nous verrons dans la partie suivante que certaines de ces régions sont importantes pour les différentes fonctions de l'ARN *Xist*. Par exemple, la région des A-repeats localisée à l'extrémité 5' de l'exon 1 est cruciale pour l'initiation de l'inactivation, sa délétion provoque un défaut dans la mise en place de ce processus (Wutz *et al.*, 2002 ; Hoki *et al.*, 2009).



**Figure 30 : Conservation de l'architecture et des séquences répétées du gène *Xist* chez les euthériens**

Le numéro de chaque exon est donné sous le rectangle gris lui correspondant. Les introns sont indiqués par des traits. L'exon 2 chez l'homme (h2), l'exon 5 chez les rongeurs (r5) et l'exon supplémentaire 4 (4b) chez le chien et la taupe ne sont pas conservés. Les répétitions sont indiquées par des rectangles de couleurs. Chez la taupe, la région des A-repeats n'est pas conservée et une région supplémentaire est observée chez la vache et l'homme (régions Bh et G).

D'après la revue Brockdorff, 2002 ; Yen *et al.*, 2007

### 3.2 Caractérisation fonctionnelle du gène *Xist*

Les délétions de la région promotrice et du premier exon du gène *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 provoquent un biais d'inactivation, à savoir que l'inactivation a toujours lieu sur le chromosome X sauvage (Marahrens *et al.*, 1997). Ces délétions altèrent donc le mécanisme d'inactivation, d'où l'idée que *Xist* est un facteur clé de ce processus. L'introduction d'un ADNc de *Xist* dans des autosomes sous le contrôle d'un promoteur inductible par la tétracycline (Wutz et Jaenisch, 2000) a permis de montrer que l'induction de l'expression de *Xist* est suffisante pour inactiver un chromosome contenant son gène, cela sans que la différenciation soit initiée et même dans un contexte mâle. Cette étude a donc apporté la preuve formelle que l'ARN *Xist* non-codant est primordial pour l'inactivation aléatoire. Le rôle de *Xist* dans l'inactivation soumise à empreinte parentale a été montré par le fait que des mutations dans *Xist* transmises par le père, provoquent la mort des embryons du fait d'une défaillance dans le développement des tissus extra-embryonnaires (Marahrens *et al.*, 1997 ; Hoki *et al.*, 2009).

Afin de déterminer les éléments fonctionnels de l'ARN Xist, Anton Wutz *et al.* (2002) ont généré de nombreux ADNc Xist porteurs de diverses délétions. Ces ADNc étaient placés sous le contrôle d'un promoteur inducible (promoteur répondant à la doxycycline) et ont introduit individuellement par recombinaison au locus Hprt (hypoxanthine phosphoribosyltransferase *locus*) du chromosome X dans des cellules ES mâles (Wutz *et al.*, 2002). Ils ont pu ainsi étudier, après induction de l'expression du transgène, la localisation de l'ARN Xist par RNA-FISH et l'efficacité d'inactivation par estimation de la mortalité cellulaire. Ces auteurs ont ainsi pu identifier à la fois les séquences de Xist impliquées dans la répression transcriptionnelle et celles impliquées dans sa localisation sur le Xi. Une très grande région correspondant environ aux deux tiers de l'ARN Xist et contenant les 6 régions conservées contenant des répétitions semble nécessaire pour la localisation optimale de l'ARN Xist sur le Xi (**Fig. 31**, mutant  $\Delta$ SPs). Ces éléments impliqués dans la localisation semblent avoir des fonctions redondantes (Wutz *et al.*, 2002), ce qui donnerait à l'ARN sa capacité à recouvrir le chromosome X même si des délétions apparaissent (mutant  $\Delta$ P). Une illustration de la redondance fonctionnelle des régions répétées est apportée par le fait que la délétion de l'exon 4, très conservé entre les espèces, ne provoque pas de délocalisation hors du Xi (Caparros *et al.*, 2002).

Les expériences réalisées par l'approche citées ci-dessus et la délétion de la région des A-repeats dans l'embryon ont permis de mettre en évidence l'importance de cette région dans l'extinction transcriptionnelle (**Fig. 31**, mutant  $\Delta$ SX). Toutefois, la région des A-repeats n'est pas impliquée dans la localisation de Xist (mutant  $\Delta$ SX), la formation du domaine Xist, et l'extinction transcriptionnelle de séquences répétées intergéniques et introniques du chromosome X (Wutz *et al.*, 2002 ; Chaumeil *et al.*, 2006 ; Clemson *et al.*, 2006 ; Hoki *et al.*, 2009). De plus, les transgènes de Xist portant la délétion de la région des A-repeats sont capables d'induire l'apposition des marques hétérochromatiques dépendantes de Xist (Wutz *et al.*, 2002). La région des A-repeats est donc nécessaire à la mise en place de l'inactivation mais intervient dans une étape bien définie comme nous le verrons plus loin.

Souris			Inactivation	Localisation	
	A	—————	Xist	+	+
	F	—————	$\Delta$ SX	-	+
	B	—————	$\Delta$ P	+	+
	C	—————	$\Delta$ SPs	-	-
	D	—————	$\Delta$ E	+/-	+/-
	E	—————			

**Figure 31 : Eléments fonctionnels impliqués dans la localisation et dans la répression transcriptionnelle du Xi médiée par l'ARN Xist**

L'ARN Xist est présenté avec ses répétitions indiquées par des rectangles de différentes couleurs. Au dessous sont représentés certains transgènes générés par Wutz *et al.*, pour caractériser les éléments fonctionnels de l'ARN Xist. Les phénotypes associés aux différents transgènes sont indiqués à droite.  
D'après Wutz *et al.*, 2002

#### 4. Régulation de l'expression du gène *Xist*

La régulation de l'expression du gène *Xist* dépend de plusieurs mécanismes : la régulation négative par son antisens *Tsix* et l'intervention de facteurs de transcription impliqués dans la pluripotente.

##### 4.1 Régulation de l'expression du gène *Xist* par *Tsix*

###### 4.1.1 Structure du gène *Tsix*

Le gène *Tsix*, antisens de *Xist* a été découvert en 1999 par emploi de diverses méthodes : de RNA-FISH (Lee *et al.*, 1999), Northern blot (Debrand *et al.*, 1999), RT-PCR (Debrand *et al.*, 1999 ; Lee *et al.*, 1999 ; Mise *et al.*, 1999) et clonage (Mise *et al.*, 1999). Il est long de 40 kb, chevauche complètement le gène *Xist* et la plupart des sites possibles de terminaison de *Tsix* sont localisés en amont de *Xist*, plus précisément à moins de 1,5 kb en amont du site d'initiation de la transcription du gène *Xist* (Debrand *et al.*, 1999 ; Lee *et al.*, 1999 ; Mise *et al.*, 1999 ; Ogawa et Lee, 2003 ; Sado *et al.*, 2001). Le gène *Tsix* produit un ARN *Tsix* qui est polyadénylé et épissé de manière alternative d'où l'existence de plusieurs types de variants (Debrand *et al.*, 1999 ; Lee *et al.*, 1999).

- Sites alternatifs de la transcription

La présence de certains variants est liée à l'existence de plusieurs sites d'initiation de la transcription. L'origine du transcrit majoritaire est située au sein d'un îlot CpG associé au minisatellite *DXPas34* (Avner *et al.*, 1998 ; Courtier *et al.*, 1995 ; Cohen *et al.*, 2007 ; Lee *et al.*, 1999). Un groupement de sites d'initiation de la transcription a également été caractérisé au locus *Xite* (Ogawa et Lee, 2003 ; Stavropoulos *et al.*, 2005) (Cf Fig. 22).

- Les variants d'épissage

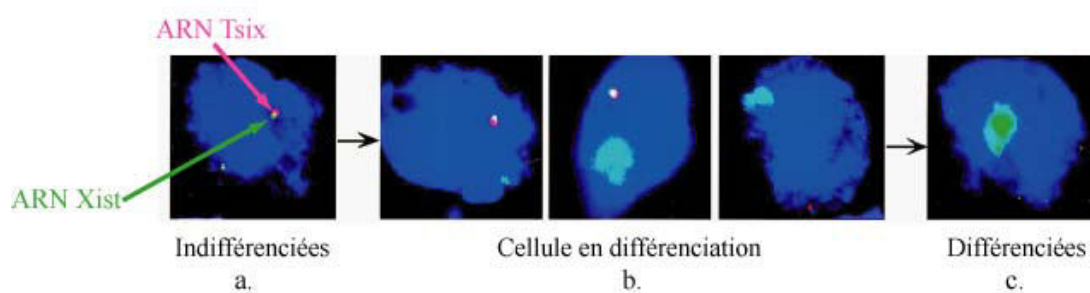
Un grand nombre des variants d'épissage de *Tsix* ont été détectés. Les formes épissées représentent environ 60 % des transcrits totaux (Shibata et Lee, 2003). Néanmoins, l'importance fonctionnelle des formes épissées de *Tsix* n'est pas claire, car la mutation de tous les sites d'épissage du gène *Tsix* n'altère par l'inactivation du chromosome X (Sado *et al.*, 2006).

###### 4.1.2 Caractérisation fonctionnelle de *Tsix*

Dans les cellules indifférenciées, les ARN *Tsix* et *Xist* sont coexprimés par les 2 chromosomes X dans les cellules ES femelles. Au début de la différenciation, l'expression de



l'ARN *Tsix* est maintenue à partir du chromosome qui restera actif et permet la répression de l'expression de *Xist*, de même que dans les cellules indifférenciées et chez les mâles. Par contre, au niveau du futur Xi, l'expression de l'ARN *Tsix* est inhibée, ce qui engendre la sur-expression de l'ARN *Xist* et la mise en place de l'inactivation (Lee et Lu, 1999 ; pour revue, Boumil et Lee, 2001) (**Fig. 32**). Comme nous l'avons déjà mentionné, des mutations entraînant l'extinction hétérozygote de l'expression de *Tsix* conduisent obligatoirement à la sur-expression de *Xist* sur le chromosome X muté et à sa répression (Lee et Lu, 1999 ; Luikenhuis *et al.*, 2001 ; Morey *et al.*, 2001 ; Sado *et al.*, 2001 ; Sun *et al.*, 2006 ; Vigneau *et al.*, 2006 ; Shibata et Lee, 2004). Par ailleurs, la sur-expression de *Tsix* placé sous contrôle d'un promoteur inductible dans un des deux chromosomes X dans les cellules ES en voie de différenciation abolie la possibilité du chromosome X sur-exprimant *Tsix*, d'exprimer *Xist* et donc d'être inactivé (Stavropoulos *et al.*, 2001). *Tsix* est donc considéré comme un élément régulant l'inactivation du chromosome X par antagonisme avec *Xist*.



**Figure 32 : Interaction dynamique entre *Xist* et *Tsix* durant la différenciation des cellules ES**

La détection des deux transcrits est réalisée par RNA-FISH. L'ARN *Xist* a été détecté par une sonde spécifique marquée à la fluoroscéine (signal vert). L'ARN *Tsix* a été détecté par une sonde spécifique marquée au "texas red" (signal rouge).

Dans les cellules ES non différenciées, les ARN *Tsix* et *Xist* sont coexprimés par les 2 chromosomes X à un taux basal (a.). Au début de la différenciation (b.), l'expression de l'ARN *Tsix* est maintenue sur le chromosome qui restera actif. *Tsix* inhibe l'accumulation de l'ARN *Xist*. Au niveau du chromosome destiné à être inactivé, l'expression de l'ARN *Tsix* est inhibée. Ceci conduit à une accumulation de l'ARN *Xist* permettant la mise en place de l'inactivation. Dans les cellules différenciées (c.), un large domaine correspondant au Xi est détectable par localisation de l'ARN *Xist*, en revanche, les expressions de *Xist* et *Tsix* sont réprimées sur la Xa (c.).

D'après la revue, Boumil et Lee, 2001

#### 4.1.3 Comment *Tsix* régule l'expression du gène *Xist*

Des études ont démontré que la régulation de *Xist* par *Tsix* pourrait impliquer les modifications de la chromatine. Dans les cellules ES femelles indifférenciées (qui expriment *Tsix* et *Tsix* qui régule négativement *Xist*), la région promotrice de *Xist* est enrichie en H3K4me2 et H3K9me3, mais appauvrie en H3K4me3, H3K9ac et H3K27me3 (Navarro *et*

*al.*, 2006). Lorsque *Tsix* est tronqué, une accumulation des marques de la chromatine active telles que l'augmentation des H3K4me2-3 et des H3K9ac mais aussi l'hypométhylation de l'ADN et une diminution des H3K9me3 sont observées. Ceci correspond à un basculement de la chromatine du promoteur de *Xist* d'un état inactif à un état actif (Cf Fig. 17) (Navarro *et al.*, 2005 ; Navarro *et al.*, 2006 ; Sado *et al.*, 2005 ; Ohhata *et al.*, 2008 ; Shibata et Yokota, 2008). *Tsix* est donc capable d'influencer l'apposition des marques épigénétiques au niveau du promoteur de *Xist* et il permettrait de moduler l'état hétérochromatinien du promoteur du gène *Xist* sur le Xa dans le but de bloquer son expression (McDonald *et al.*, 1998 ; Navarro *et al.*, 2005 ; Sado *et al.*, 2005 ; Sun *et al.*, 2006)..

Cependant, la délétion de *Tsix* ne provoque pas seulement une augmentation des marques euchromatiniennes au niveau du promoteur du gène *Xist*. Elle provoque aussi l'augmentation des H3K27me3 qui est une marque épigénétique caractéristique de la chromatine silencieuse (Navarro *et al.*, 2006 ; Sun *et al.*, 2006 ; Shibata et Yokota, 2008). Ces apparentes contradictions peuvent être expliquées par le caractère bivalent de certains gènes régulant le développement des cellules ES qui sont marqués à la fois par les H3K27me3 et par les H3K4me2-3 (Bernstein *et al.*, 2006).

Un autre mode possible de régulation de *Xist* par *Tsix* qui avait été envisagé reposent sur la déstabilisation des transcrits *Xist* par *Tsix* suivie d'une hydrolyse de l'ARN double brin (Panning *et al.*, 1997 ; Sheardown *et al.*, 1997a). Cependant, l'analyse de la stabilité de *Xist* dans les cellules n'exprimant plus *Tsix* n'est pas modifiée au cours de la différenciation de cellules ES (Sun *et al.*, 2006). De plus, dans des cellules ES déficientes pour Dicer, l'ARN *Xist* est capable d'habiller le chromosome X et de recuter les complexes PcG. De même la sur-expression de *Xist* dans les cellules ES mâles induit la répression des gènes liés aux chromosomes X (Nesterova *et al.*, 2008 ; Ogawa *et al.*, 2008 ; Kanellopoulou *et al.*, 2009). Donc, la dégradation des ARN *Xist* par un mécanisme de type RNAi dépendant de *Tsix* semble peu probable.

#### 4.1.4 Régulation de l'expression de *Tsix*

Le promoteur minimal du gène *Tsix* est constitutivement actif mais ne produit qu'à un taux basal de transcrit *Tsix*. D'où l'idée qu'il existe des séquences régulatrices (Stavropoulos *et al.*, 2005). Ces séquences régulatrices correspondent aux *loci Xite* et *DXPas34*, que nous avons déjà mentionné dans la présentation du locus XIC. Ces 2 *loci Xite* et *DXPas34* activent la transcription de *Tsix* (Cohen *et al.*, 2007 ; de Napoles *et al.*, 2007 ; Ogawa *et al.*, 2008 ; Stavropoulos *et al.*, 2005 ; Vigneau *et al.*, 2006).

Il a été montré par des expériences d'immunoprécipitation de la chromatine que le locus *DXPas34* fixe la protéine CTCF (*CCCTC-binding factor*) (Chao *et al.*, 2002 ; Donohoe *et al.*, 2007). La plupart des sites de fixation de CTCF dans *DXPas34* sont associés à

des sites de fixation du facteur Yy1 (Donohoe *et al.*, 2007 ; Kim *et al.*, 2006). Yy1 est un facteur de transcription à doigts de zinc de la famille Gli-Kruppel (Kim *et al.*, 2006). Il est l'orthologue de la protéine PHO, membre de la famille des protéines polycomb chez la drosophile (Atchison *et al.*, 2003 ; Wilkinson *et al.*, 2006). *In vivo*, l'inactivation de *CTCF* par RNAi est létale avant l'implantation. L'inactivation par *knock-out* de *Yy1* est également létale dans la période péri-implantatoire, si elle est réalisée sur les deux allèles (Donohoe *et al.*, 2007). En l'absence de Yy1 ou de CTCF dans des cellules ES mâles, l'expression de *Tsix* est fortement diminuée et celle de *Xist* augmentée (Donohoe *et al.*, 2007). Des expériences de co-immunoprécipitation et des tests d'activité luciférase réalisées avec des constructions rapporteurs suggèrent que ces deux protéines pourraient collaborer pour activer l'expression de *Tsix* (Donohoe *et al.*, 2007). Ces protéines semblent donc être importantes pour la régulation de *Tsix*. Toutefois leur mécanisme d'action est encore méconnu.

Il a été démontré que la région ADN *Xite* (Ogawa et Lee, 2003) interagit physiquement avec le gène *Tsix* (Xu *et al.*, 2006). Cette interaction est dépendante du stade de développement et du sexe. L'interaction entre *Xite* et *Tsix* est détectée dans les cellules ES femelles et mâles non-différenciées et disparaît progressivement après différenciation. Il a été proposé que cette interaction permettrait d'activer l'expression de *Tsix* et donc permettrait à un des chromosomes X de rester actif. Un tel système de régulation a déjà été décrit pour certains *loci* ayant un profil de régulation complexe. Par exemple, chez la souris et l'homme, les gènes de la  $\beta$ -globine interagissent avec des régions de « contrôle du locus » (LCR) et ces interactions régulent l'expression génique par la formation d'un « centre de chromatine active » (active chromatin hub) correspondant à une structure ordonnée hautement permissive à la transcription (Zhou *et al.*, 2006 ; pour revue, de Laat et Grosveld, 2003 ; Patrinos *et al.*, 2004 ; Tolhuis *et al.*, 2002).

#### 4.2 Répression de *Xist* et pluripotence

Nous avons vu que l'expression du gène *Xist* est un élément essentiel dans les différentes phases d'inactivation et d'activation transcriptionnelle successives des chromosomes X au cours du développement embryonnaire. C'est la raison pour laquelle il est primordial que l'expression de *Xist* soit hautement régulée. Jusque-là, il était considéré que l'expression de *Xist* est surtout régulée négativement par son antisens *Tsix*. Il a été montré récemment qu'il existe en plus un système indépendant de *Tsix*. L'ARN *Xist* étant exprimé qu'après la perte du caractère pluripotent des cellules, il a été proposé que la régulation transcriptionnelle de *Xist* soit dépendante de la perte de facteurs assurant la pluripotence des cellules. Des expériences réalisées sur les cellules de la masse interne de l'embryon (Mak *et al.*, 2004 ; Okamoto *et al.*, 2004), dans les PCG (Nesterova *et al.*, 2002) et sur des fusions de cellules somatiques (Takagi *et al.*, 1983) ont conduit à proposer un rôle possible des facteurs

transcriptionnels de pluripotence Nanog et Oct4. En effet, l'expression de Nanog a un profil opposé à celui de l'expression de *Xist* (Chambers *et al.*, 2003 ; Mitsui *et al.*, 2003 ; Yamaguchi *et al.*, 2005) et la réactivation du chromosome X dans les cellules de la masse interne n'est observée que dans les cellules dans lesquelles Nanog est exprimé (Mak *et al.*, 2004). Récemment des preuves formelles de l'implication de ces facteurs de pluripotences ont été apportées par des expériences d'immunoprécipitation de la chromatine. En effet, l'équipe de P Avner a démontré que les facteurs Nanog, Oct4 et Sox2 se fixent au niveau de la séquence ADN codant l'extrémité 3' de l'intron 1 du gène *Xist* dans les cellules non différenciées qu'elles soient mâles ou femelles (Navarro *et al.*, 2008). Cette interaction est perdue dans les cellules différenciées et ne dépend pas de la transcription de *Tsix* (Sun *et al.*, 2006 ; Navarro *et al.*, 2008 ; Navarro *et al.*, 2006). Dans les cellules mâles, l'extinction de l'expression de ces 3 facteurs provoque une perte totale de la pluripotence, un défaut d'inactivation et une forte augmentation de l'ARN *Xist* (Navarro *et al.*, 2008). Ces données sont en accord avec le fait que *Xist* est sur-exprimé dans des cellules dans lesquelles l'intron 1 du gène *Xist* a été remplacé par une cassette IRES-GFP, et ceci sans diminution notable de l'expression de *Tsix* (Sado *et al.*, 2006).

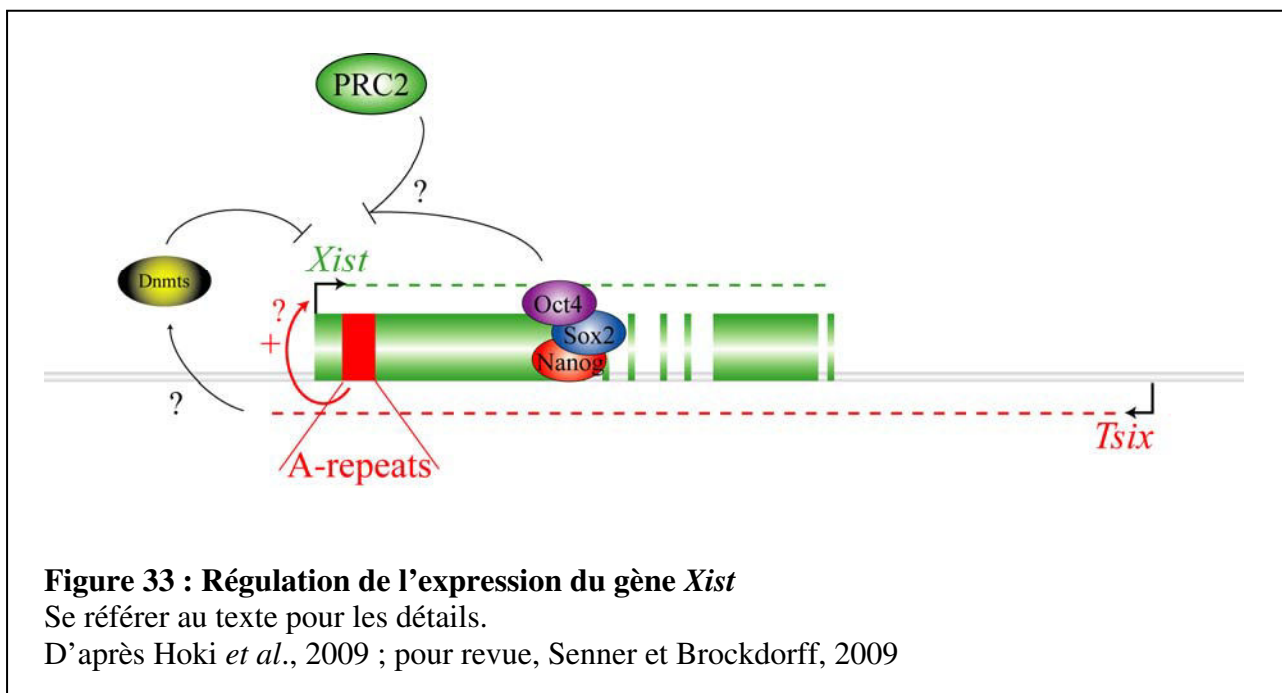
#### 4.3 Méthylation du promoteur du gène *Xist*

Il a été démontré dans les cellules ES indifférenciées que le fait de transcrire *Xist* influence la méthylation de son propre promoteur (rétrocontrôle négatif). En effet, la délétion d'une région de 9 kb ou l'insertion de site de terminaison de la transcription en amont du promoteur de *Xist* entraînent une augmentation de la transcription de *Xist* dans les cellules indifférenciées XY. Cette activation est liée à une hypométhylation du promoteur (Nesterova *et al.*, 2008). Il y a alors inactivation inappropriée de l'unique chromosome X chez la cellule mâle. La méthylation du promoteur *Xist* avant le début de l'inactivation semble dépendre des taux relatifs de transcription sens et antisens (Nesterova *et al.*, 2008).

#### 4.4 Vision globale actuelle du mécanisme de répression de l'expression du gène *Xist*

Un modèle global de régulation de l'expression de *Xist* peut être maintenant envisagé. Lorsque *Xist* est réprimé, son promoteur porte les marques de l'hétérochromatine (méthylation des îlots CpG par les Dnmt et méthylation des histones par le complexe PRC2) alors que lorsqu'il est actif, il porte les marques de l'euchromatine (hypométhylation de l'ADN et acétylation des histones). La balance entre la transcription sens et celle de l'antisens le long du promoteur de *Xist* influence le profil de modifications épigénétiques de ce promoteur. La mise en place des modifications de type hétérochromatiniennes dépend de l'expression de *Tsix* (Navarro *et al.*, 2006 ; McDonald *et al.*, 1998 ; Sado *et al.*, 2005 ; Sun *et al.*, 2006) et de la transcription sens (**Fig. 33**). Les méthylations des histones sont mises en place par le

complexe PRC2, en particulier, par la protéine Eed et ceci est renforcé par la transcription de *Tsix*, dans ce cas, le promoteur de *Xist* est inactivé (Shibata *et al.*, 2008). De plus, indépendamment à *Tsix*, les 3 facteurs de pluripotence Nanog, Oct4 et Sox2 se fixent sur la région intronique 1 de *Xist* et permettent sa répression. Une régulation post-transcriptionnelle a été récemment démontré car la délétion de la région des A-repeats provoque un blocage de l'inactivation du chromosome X lié à une diminution de la quantité de l'ARN *Xist* (Hoki *et al.*, 2009) qui pourrait elle, être due à une diminution de la  $\frac{1}{2}$  vie de l'ARN. Tous ces systèmes de régulation sont le reflet de l'extrême complexité de l'inactivation du chromosome X.



## 5. La région des A-repeats : l'ultime acteur

### 5.1 Caractérisation fonctionnelle de la région des A-repeats

Comme nous l'avons déjà mentionné, l'ARN *Xist* est composé de plusieurs régions comprenant des séquences répétées en tandem (Clemson *et al.*, 1996 ; Nesterova *et al.*, 2001) (Cf Fig. 30), nous allons voir que la région localisée à l'extrémité 5' de l'exon 1 de l'ARN *Xist*, nommée la région des A-repeats, est cruciale pour l'inactivation (Wutz *et al.*, 2002 ; Hoki *et al.*, 2009).

Chez l'embryon murin, la délétion mono-allélique paternelle de la région des A-repeats provoque un défaut d'inactivation dans les tissus extra-embryonnaires (inactivation du chromosome X liée à l'empreinte parentale) et met en évidence l'incapacité du chromosome X muté à être inactivé (Hoki *et al.*, 2009).

Lorsque l'ARN *Xist* est transcrit à partir d'un ADNc dans lequel la région des A-repeats est délétée, ceci sous contrôle d'un promoteur inductible, il peut s'accumuler sur le chromosome X dans les cellules ES mâles murines, mais est incapable d'initier l'inactivation

(Wutz *et al.*, 2002). Dans des cellules humaines HEK293, lorsque l'ADN correspondant à l'ARN Xist sans la région des A-repeats est intégré en simple copie, toujours sous le contrôle d'un promoteur inductible, on observe un blocage complet de la localisation de l'ARN Xist et une absence d'inactivation du chromosome X. Donc, il semble que dans les cellules humaines, cette région soit essentielle pour la localisation de Xist sur le Xi et la répression génique (Chow *et al.*, 2007). Toutes ces observations mettent en évidence que la région des A-repeats est très importante pour l'inactivation du chromosome X, que l'inactivation soit soumise à empreinte parentale (études dans l'embryon) ou aléatoire (étude dans les cellules ES).

## 5.2 Régulation de l'expression de la région 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 compose 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 - d'une part que l'utilisation différentielle des promoteurs est coordonnée avec l'avancement de la différenciation – 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 (Johnston *et al.*, 1998) puisque la région des A-repeats participe à l'initiation de l'inactivation. 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.

Comme nous l'avons déjà mentionné, 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 (Zhao *et al.*, 2008).

## 5.3 Connaissances actuelles sur le(s) rôle(s) de la région des A-repeats

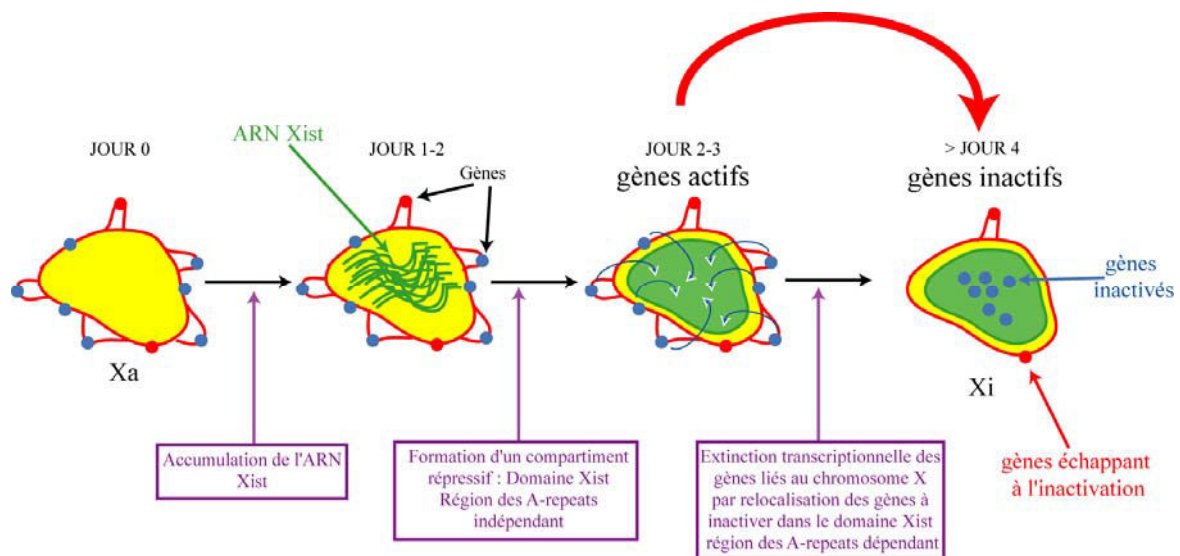
### 5.3.1 Importance de la région des A-repeats sous sa forme ADN

Au niveau génomique, la région des A-repeats est un élément nécessaire à la régulation de l'expression de l'ARN Xist (Hoki *et al.*, 2009) (Cf Fig. 33). En effet, la délétion

de la région des A-repeats provoque un blocage de l'inactivation du chromosome X lié à une diminution de la quantité de l'ARN Xist (Hoki *et al.*, 2009). Ces résultats sont en accord avec des données précédentes obtenues à partir d'un système rapporteur qui montraient l'existence d'une forte activité stimulatrice du promoteur Xist de la région des A-repeats (Allaman-Pillet *et al.*, 2000). La région ADN de la région des A-repeats est donc importante en tant qu'activateur transcriptionnel en plus de son importance au sein de l'ARN. Il est d'ailleurs important de noter que toute approche délétionnelle visant à caractériser les fonctions de cette région affecte à la fois les fonctions de l'ARN et de l'ADN, d'où l'importance de trouver des stratégies permettant de séparer la fonction de l'ADN de celle de la région correspondante de l'ARN.

### 5.3.2 Importance de la région des A-repeats sous sa forme ARN

D'autres études ont montré que la région des A-repeats induit la répression transcriptionnelle par la relocalisation des gènes à inactiver dans le domaine Xist. En effet, Chaumeil *et al.*, (2006) ont étudié la cinétique précise des premiers événements de la mise en place du domaine Xist, 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 Xist, la répression des gènes est liée à leur repositionnement de l'extérieur vers l'intérieur du domaine Xist. Au départ, le domaine d'exclusion de la polymérase II apparaît essentiellement formé de séquences répétées non géniques, dont l'expression est réprimée (Chaumeil *et al.*, 2006 ; Clemson *et al.*, 2006). La première phase de répression transcriptionnelle des séquences répétées, ainsi que la mise en place du domaine Xist ne sont pas affectées par l'absence de la région des A-repeats de l'ARN Xist. Par contre, cette région est requise à la fois pour l'inactivation des gènes du chromosome X (et non des séquences répétées) et leur relocalisation (Chaumeil *et al.*, 2006). Un modèle d'inactivation en 2 étapes a été proposé : la formation d'un compartiment répressif, puis le recrutement des gènes dans ce compartiment, la deuxième étape étant dépendante de la région des A-repeats (**Fig. 34**). Il a été récemment démontré que les protéines SATB1 et SATB2 pourraient avoir un rôle dans ce processus. Ces protéines étaient connues pour être impliquées dans la régulation génique dans les thymocytes (Dickinson *et al.*, 1992 ; Alvarez *et al.*, 2000 ; Cai *et al.*, 2003) et qu'elles intervenaient dans la réorganisation nucléaire de séquences chromosomiques lors du développement lymphocytaire (Cai *et al.*, 2003). Elles interviendraient de manière similaire dans la relocalisation des gènes inactifs du Xi à l'intérieur du domaine Xist (Agrelo *et al.*, 2009 ; pour revue, Chow et Heard, 2009). Il serait intéressant de tester par des expériences d'immunoprécipitation des ARN, si la région des A-repeats interagit avec les protéines SATB1/B2.



**Figure 34 : Intervention 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 bleus et rouges) des chromosomes X (en jaune) sont actifs. En début de différenciation, l'ARN Xist commence à recouvrir le chromosome X choisi pour être inactivé. Puis, la formation du domaine Xist est associée à l'exclusion de la machinerie de transcription et à la répression transcriptionnelle des séquences répétées, toutes deux indépendantes de la région des A-repeats. L'inactivation des gènes liés au chromosome X implique un processus dépendant de la région des A-repeats, qui est concomitant à la mise en place de marques de type hétérochromatique sur le Xi. Les gènes du chromosome X sont relocalisés dans le domaine d'ARN Xist. Les gènes échappant à l'inactivation sont maintenus en dehors de ce domaine.

D'après Chaumeil *et al.*, 2006

La découverte de partenaires protéiques de la région des A-repeats est une étape indispensable pour comprendre comment a lieu l'inactivation du chromosome X. Des expériences d'immuno-précipitation réalisées dans la perspective d'identifier ces protéines ont permis de montrer que les protéines du complexe PRC2 (Ezh2 et Suz12) se lient à l'ARN Xist (Zhao *et al.*, 2008). Dans les cellules ES non différenciées, le complexe PRC2 interagissait seulement avec la région des A-repeats. Après différenciation (et donc sur-expression de *Xist*), le complexe PRC2 interagissait en plus avec la région 3' de Xist. Trois et six jours après la différenciation, la fixation du complexe PRC2 à la région des A-repeats n'est pas retrouvée ce qui suggère que cette fixation intervient avant et au début de l'initiation de l'inactivation. De ce fait, il a été suggéré que le complexe PRC2 soit recruté par la région ARN des A-repeats et soit transféré sur la chromatine afin de catalyser la mise en place des H3K27me3 après différenciation.

La région des A-repeats n'est peut-être pas la seule à fixer le complexe PRC2 car des transgènes inductibles dans lesquels la région des A-repeats est délétée sont encore capables d'induire les méthylations H3K27me3 sur le chromosome X. Les différentes régions de l'ARN Xist pourraient donc avoir des fonctions redondantes (Kohlmaier *et al.*, 2004 ; Plath *et*

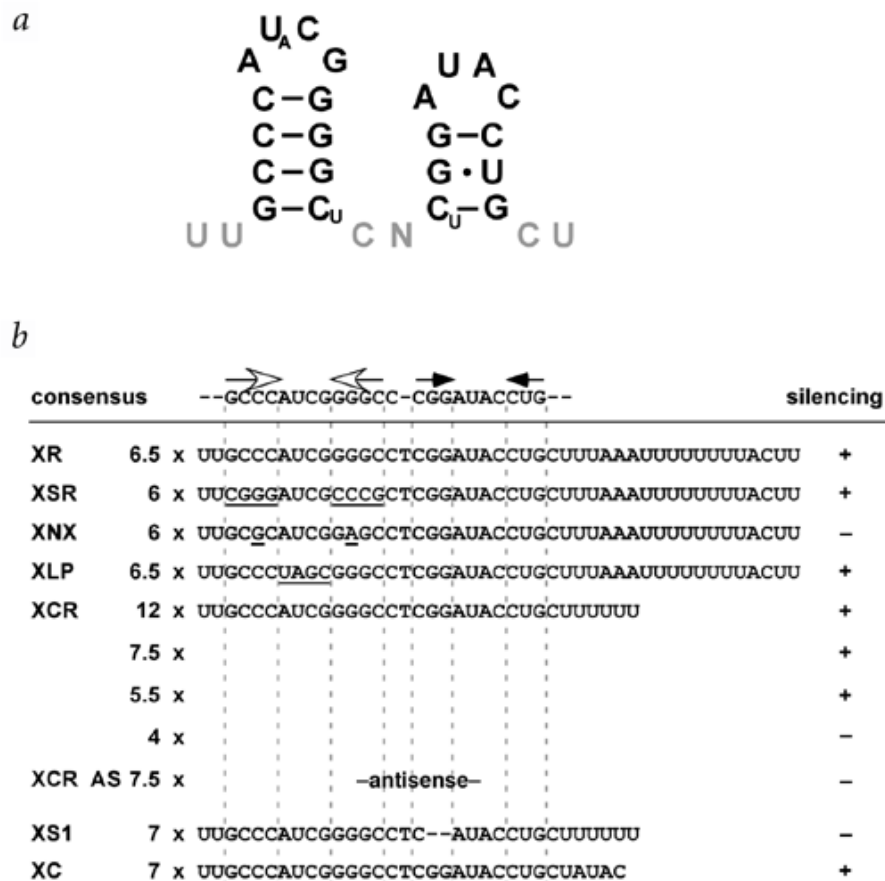


*al.*, 2003). Enfin, deux études montrent que le complexe PRC2 n'est pas indispensable à l'initiation de l'inactivation (Schoeftner *et al.*, 2006 ; Kalantry et Magnuson, 2006). Il y aurait donc à la fois redondance de certaines fonctions des régions de l'ARN Xist et des facteurs protéiques impliqués dans l'inactivation aléatoire. Néanmoins, la région des A-repeats est pour le moment le seul élément de l'ARN Xist ayant été montré avoir un caractère indispensable pour l'inactivation.

#### 5.4 Composition et structure secondaire de la région des A-repeats

La région des A-repeats murine renferme 419 nt chez la souris. Elle 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)** et une demi-répétition de ce motif (**GCCCAACGGGGC**). Ces répétitions sont séparées par des régions très riches en résidus U. Chez l'homme, 8 répétitions entières sont présentes et elles sont aussi séparées par des régions riches en résidus C et U. Chaque répétition comporte deux séquences complémentaires riches en G et C, GCCC et GGGC. Etant donné que la structure secondaire d'un ARN conditionne son action, la structure 2D potentielle de chaque répétition avait été prédite par voie informatique. Selon le modèle obtenu, les séquences GCCC et GGGC formeraient une petite structure en tige-boucle (Wutz *et al.*, 2002) et les triplets (C/U)GG et (C/U)U(G/A) de la partie 3' des répétitions formeraient une seconde structure tige-boucle (**Fig. 35a**). Néanmoins, la seconde structure tige-boucle n'est pas thermodynamiquement stable et la structure 2D de la région des A-repeats doit forcément impliquer les segments reliant les répétitions non pris en compte dans la prédiction.

En réinsérant un nombre variable de répétitions dans le gène *Xist*, il a été montré qu'il faut au minimum 5,5 répétitions du motif pour que l'inactivation soit efficace (**Fig. 35b**, mutant XR 5.5). La mutation des bases composant les boucles proposées par l'approche informatique n'influence pas l'inactivation (**Fig. 35b**, mutant XLP). En revanche, des mutations dans les éléments proposés former les hélices des structures tige-boucle abolissent l'inactivation (**Fig. 35b**, mutant XNX) (Wutz *et al.*, 2002). A ce stade, il n'est pas possible de savoir si c'est la séquence ou la structure de ces éléments qui est importante. Le fait que la génération de mutations compensatoires restaurant les structures tige-boucle semblait restaurer l'inactivation suggérait néanmoins une importance de la structure 2D quelle qu'elle soit (**Fig. 35b**, mutant XSR) (Wutz *et al.*, 2002). Il avait aussi été proposé par des expériences de retard sur gel que des mutants dans lesquels les structures tige-boucle potentielles étaient déstabilisées n'interagissent pas avec le complexe PRC2, alors que les ARN ayant la séquence sauvage sont capables d'interagir (Zhao *et al.*, 2008). Ce qui renforçait l'idée que la structure 2D pourrait être importante.



### Figure 35 : Analyse de la région des A-repeats

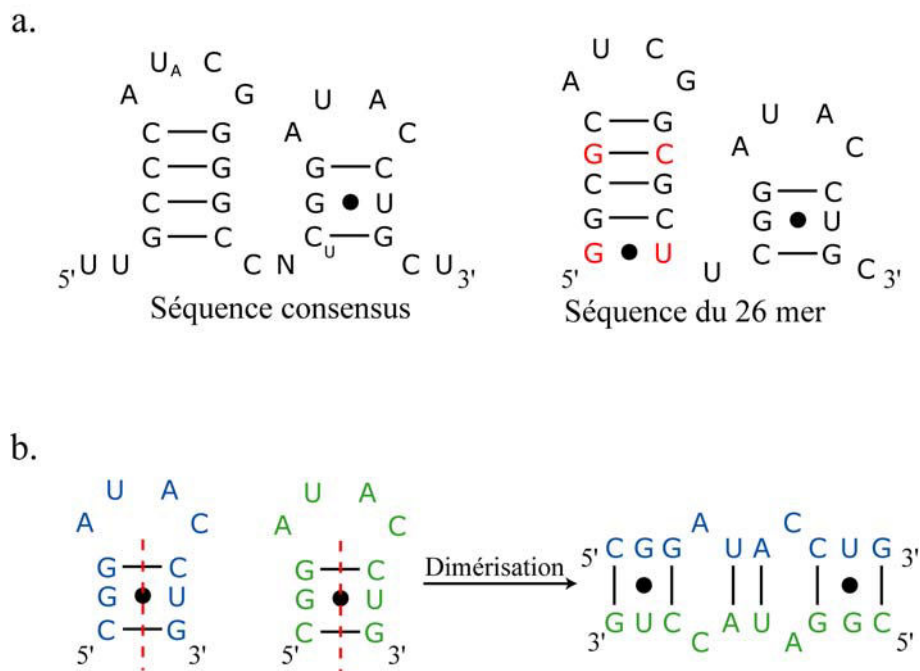
a. Prédiction de la structure secondaire d'une répétition de la région des A-repeats

b. Séquence des monomères de l'élément de base de la région des A-repeats type sauvage ou muté par Wutz et al. (2002). Le nombre des 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

Comme nous l'avons mentionné ci-dessus, la prédiction de la structure 2D réalisée n'était pas optimale. En plus, les prédictions informatiques de structure 2D ne sont pas très fiables en absence de données expérimentales. Il était donc important de réaliser l'analyse expérimentale de la structure 2D de la région des A-repeats. Récemment, il a été montré par RMN qu'un élément répété de la région des A-repeats (26 nt) pris individuellement 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 (Duszczuk *et al.*, 2008) (**Fig. 36**). En plus du fait que cette répétition soit étudiée hors de son contexte, il faut remarquer que la séquence utilisée pour cette analyse ne correspondait pas à la séquence consensus des répétitions. Elle comporte les séquences GGCGC et GCGCU au lieu des séquences GCCC et GGGC dont nous avons parlé précédemment (**Fig. 36**).



**Fig 36 : Modèle de dimérisation des éléments répétés de la région des A-repeats proposé par Duszczyk *et al.* (2008)**

a. Comparaison entre la séquence consensus des éléments répétés de la région des A-repeats et la séquence utilisée par Duszczyk *et al.*, 2008. Les différences sont indiquées en rouge.

b. Modèle de dimérisation de la partie 3' des éléments répétés de la région des A-repeats





## Objectifs

La complexité du processus d'inactivation du chromosome X décrite dans cette introduction souligne fortement les difficultés à surmonter pour atteindre une parfaite connaissance de ce phénomène à l'échelle moléculaire. Néanmoins, il est acquis que la région des A-repeats, segment de 500 nt localisé à l'extrémité 5' de l'ARN Xist joue un rôle central dans ce processus. Lorsque j'ai commencé à développer l'étude de l'ARN Xist au laboratoire avec Athanase Visvikis (2005), les mécanismes moléculaires sur lesquels reposait l'action de la région des A-repeats étaient inconnus. Étant donné que la structure 2D des ARN joue en général un rôle important dans leur mécanisme d'action, que le laboratoire est spécialisé dans l'étude expérimentale de la structure des ARN en solution et qu'aucune étude de la structure 2D de l'ARN Xist en particulier de la région des A-repeats n'avait été réalisée, mon premier objectif a été de réaliser la première analyse expérimentale de la région des A-repeats. Ceci a impliqué la mise en œuvre d'expériences de cartographie en solution, de FRET en collaboration avec M Blaud et d'études phylogénétiques.

Au cours de ces travaux structuraux, plusieurs publications ont apporté des informations suggérant que la région des A-repeats recrute des facteurs importants pour l'inactivation du chromosome X. Il était donc important de développer une approche permettant l'analyse la plus exhaustive possible des partenaires protéiques de la région des A-repeats. Dans cet objectif, j'ai employé une méthode de chromatographie d'affinité basée sur la fixation de la séquence MS2 en 3' de l'ARN étudié afin de la fixer sur billes d'amylose par une protéine de fusion MBP/MS2. Les partenaires fonctionnels de la région des A-repeats étant supposés être présents dans les noyaux de cellules ES de souris avant et après initiation de la différenciation, j'ai dû dans un premier temps apprendre à préparer des extraits nucléaires à partir de ces cellules dans l'équipe de P Avner en collaboration avec A Visvikis et P Clerc. J'ai ensuite dû définir les conditions de formation et de purification des complexes RNP ; les protéines des complexes purifiés ont été analysées par spectrométrie de masse en collaboration avec l'équipe d'A Van Dorsselaer. En parallèle, l'équipe de Lee a identifié par des expériences d'immuno-précipitation des ARN, les protéines Ezh2 et Suz12 comme interactants de la région des A-repeats (Zhao *et al.*, 2008). Ayant également identifié ces protéines et la plupart des protéines du complexe PRC2, nous nous sommes fixé comme objectif d'identifier les séquences et les éléments structuraux de la région des A-repeats requis pour le recrutement efficace de ce complexe. Un article sur ces travaux a été rédigé et soumis à PLoS Biology (Chapitre C).

En plus des protéines du complexe PRC2, l'analyse des protéines présentes dans les RNP formées par la région des A-repeats mettait en évidence un grand nombre d'autres

candidats partenaires. Afin d'obtenir des arguments en faveur d'une signification biologique de l'association observée entre certaines protéines et la région des A-repeats, j'ai entrepris une étude de leur localisation dans les cellules ES de souris indifférenciées et dans les cellules en voie de différenciation par immunofluorescence. Cette étude est présentée dans le chapitre D de la partie « Résultats et Discussion ».

Deux des protéines candidates, les protéines PTB et KSRP sont des protéines abondantes se liant directement à l'ARN en particulier sur des segments riches en pyrimidines. Nous avons émis l'hypothèse qu'en se fixant à la région des A-repeats, elles pouvaient stabiliser des structures tige-boucle et peut-être faciliter la fixation du complexe PCR2 à la région des A-repeats. Un autre objectif a donc consisté à vérifier les capacités d'interaction de ces protéines avec la région des A-repeats par des expériences *in vitro*. Nous avons également testé si la fixation de ces protéines sur la région des A-repeats pouvait stabiliser certaines des structures tige-boucle par dichroïsme circulaire en collaboration avec M Blaud. Cette partie de l'étude est présentée dans le chapitre D de la partie « Résultats et Discussion ».

En collaboration avec P Clerc de l'équipe de P Avner et en vue de réaliser une analyse fonctionnelle de la région des A-repeats par une approche de génétique, nous avons réalisé des constructions génétiques. Comme nous l'avons explicité dans l'introduction, à la fois la région ADN correspondant à la région des A-repeats et la région ARN elle-même sont importantes dans l'inactivation du chromosome X. Evidemment, toute approche délétionnelle au niveau de l'ADN, visant à caractériser les fonctions de la région des A-repeats affecte à la fois la fonction de l'ARN et celle de l'ADN. Pour différencier l'importance la région ARN par rapport à celle de l'ADN *in cellulo*, une possibilité était d'insérer de nouveaux sites d'épissage dans le transcrit Xist, afin d'éliminer la région des A-repeats de l'ARN au sein de l'ARN par épissage. L'idée était ensuite d'introduire un gène contenant ces sites d'épissage dans des cellules ES de souris par recombinaison homologue (pour revue, Bode *et al.*, 2002). Une telle lignée cellulaire pouvait également être utilisée pour produire des mutants de Xist dans lesquels la structure secondaire de la région des A-repeats telle que nous l'avons proposée est mutée. Ce travail a débuté en collaboration avec l'équipe de P Avner. Les stratégies utilisées pour mener à bien cette étude et l'état actuel des constructions sont présentés dans la chapitre E de la partie « Résultats et Discussion ».

Mon travail de thèse a donc porté à la fois sur l'analyse de la structure secondaire de la région des A-repeats, sur l'identification de ses partenaires protéiques, l'étude de leur interaction et enfin sur la génération d'outils pour une étude fonctionnelle.

## **MATÉRIEL ET MÉTHODES**



## II. Matériel et Méthodes

### A. Matériel

#### 1. Souches et plasmides utilisés

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

**La souche DH5 $\alpha$  d'*E. coli*** (*supE44*,  $\Delta$ *lacU169* ( $\Phi$ 80 *lacZ* $\Delta$ M15), *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*) (Hanahan, 1983) a été utilisée pour la production et la purification de plasmides.

**La souche BL21 codon plus (DE3) d'*E. coli*** ( $F^-$ , *ompT*, *hsdS<sub>B</sub>* ( $r_B^-$ ,  $m_B^-$ ), *dcm*, *gal*,  $\lambda$ (DE3), pLysS, Cm<sup>r</sup>) (Studier & Moffatt, 1986) a été utilisée pour la production de protéines recombinantes.

##### 1.2 Les cellules d'organismes eucaryotes

**Cellules HeLa** : N°ATCC® : CCL-2™ (Scherer et al, 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 reins.

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

##### 1.3 Les plasmides utilisés

###### 1.3.1 Plasmides recombinants utilisés pour la production *in vitro* d'ARN

Les fragments correspondant aux séquences des positions +1 à +1137 et +330 à +796 des gènes Xist murin et humain, respectivement, ont été amplifiés par PCR à partir d'ADN génomique avec des oligonucléotides spécifiques portant les sites de restriction *HindIII/XbaI*. Les fragments amplifiés sont clonés dans le vecteur pCR2.1. La validité des séquences clonées est vérifiée par séquençage. Puis, les sites de restrictions permettent de sous-cloner ces fragments dans le vecteur pUC18. L'oligonucléotide sens porte également le promoteur de l'ARN polymérase T7 ce qui permet de mettre sous contrôle de cette dernière la transcription des fragments clonés. Les vecteurs produits permettent la production des ARN-1137 et ARN-AH et sont nommés pARN-1137 et pARN-AH. Ces vecteurs ont également servi de matrice pour l'amplification par PCR de l'ensemble des fragments que nous avons cloné par la même stratégie dans le vecteur pUC18 afin de réaliser leur transcription au cours de l'analyse de la structure secondaire de la région des A-repeats. Les constructions pARN-AM, pARN-SLS1H

ont permis la synthèse des ARN-AM et ARN-SLS1H. Les bornes des produits d'amplification et les oligonucléotides qui ont servis à la production de toutes ces constructions sont présentés **dans la section C des résultats**.

Le vecteur pAdML3 fourni par le Dr Luhrmann contient le motif MS2. La restriction de ce plasmide par les sites *XbaI/EcoRI* a permis de l'isoler et de le sous-cloner dans le vecteur pUC18. Le vecteur ainsi produit est appelé pMS2. A partir des constructions pARN-1137 et pARN-AH, plusieurs fragments ont été amplifiés par PCR avec un couple d'oligonucléotides permettant d'ajouter les sites de restriction *HindIII/XbaI* à chaque extrémité des produits d'amplification ainsi que le promoteur de l'ARN polymérase T7 à l'extrémité 5'. Les fragments clonés dans le vecteur pMS2 permettent de produire des ARN fusionnés au motif MS2. Les vecteurs produits sont appelés pAH-MS2, pARN-1R-MS2, pARN-2RH-MS2, pARN-SLS1H-MS2, pARN-2RM-MS2, pARN-SLS1M-MS2, pARN-AM-MS2, pARN-5'M-MS2 et pARN-5'H-MS2. Les bornes des produits d'amplification et les oligonucléotides qui ont servis à la production de ces constructions sont présentés dans **le chapitre B des résultats**.

Pour l'étude de l'épissage *in vitro* de la région des A-repeats, la construction mutée a été produite à partir du vecteur pARN-1137 et clonée dans le vecteur pUC18 en utilisant les sites *HindIII/XbaI*. Cette construction est appelée pUC-Xistm1137. Ce dernier a été réduit dans le but de produire un ARN de plus petite taille compatible avec les expériences d'épissage *in vitro* (850 nts). Pour cela, une amplification par PCR avec le couple 3901/3897 a été réalisée et le produit d'amplification a été cloné dans le vecteur pUC18. Le vecteur généré est appelé pUC-Xistm. La stratégie complète de production de ce vecteur est détaillée **dans le chapitre E des résultats**. Leurs équivalents sauvages correspondent aux vecteurs pUC-Xistwt1137 et pUC-Xistwt. Les séquences des oligonucléotides utilisés pour réaliser ces constructions sont indiquées dans le tableau suivant :

Constructions	Oligonucléotide 1 5'-3' Oligonucléotide 2 5'-3'
pUC-Xistm1137 et pUC-Xistwt1137	3755 - AAGCTTAAATTTAATACGACTCACTATAGGGCTTGCTCCAGCCATGTT
	3794 - CATATGAAAAAAGAGGCATAGATATTCCAGGTAGTGC
	3795 - CATATGCTCAGGTGCTGCTGATCGTTTGGTGC
	3754 - TCTAGACGCGCAACACCGCACACTAA
pUC-Xistm et pUC-Xistwt	3901 - AAGCTTAAATTTAATACGACTCACTATAGGGTTCCCTACCACCAAGCCTT 3897 - TCTAGAGCGTAACTGGCTCGAGAATA

### 1.3.2 Les vecteurs d'expression eucaryote

Le plasmide pGL3 a été utilisé pour transférer les cellules eucaryotes. Ce plasmide contient le gène de la luciférase « fire-fly » sous contrôle du promoteur provenant du virus SV40. Ce plasmide a servi à réaliser les constructions permettant de tester l'épissage de la

région des A-repeats du gène *Xist* *in vivo*. Les différentes stratégies de production des vecteurs sont détaillées dans le **chapitre E** des résultats. Les fragments ont été amplifiés par des amorces permettant d'ajouter les sites de restriction *KpnI/HindIII* aux bornes des produits d'amplification. Leur clonage dans le vecteur pGL3 a donné naissance aux vecteurs nommés pXist, pXistwt, pXistm.

Les séquences des oligonucléotides utilisés pour réaliser ces constructions sont indiquées dans le tableau suivant :

Construction	Oligonucléotide 1 5'-3' Oligonucléotide 2 5'-3'	Utilisation
pUC-Xistm	3792 - TTTGATATTTGTAATAAGGTCAGTCAATTTTCATTTTAAAGG 3793 - TGAAAAATTGACTGACCTTATTACAAATATCAAAAAGTATGGAG	Insertion du site 5' d'épissage
	3895 - TATTTGTAATAAGGTAAGTCAATTTTCATTTTAAAGGAAT 3896 - AATGAAAAATTGACTTACCTTATTACAAATATCAAAAAGTA	
	3893 - TATGGTTTTGTGAGTTATTAACACCTGGAATATCTATGCCTCTTTTT 3894 - CATAGATATTCAGGTGTTAATAACTCACAAAACCATATTTCCATCCT	Insertion de la boîte de branchement
	3755 - AGCTTAAATTTAATACGACTCACTATAGGGCTTGCTCCAGCCATGTT 3794 - CATATGAAAAAAGAGGCATAGATATTCAGGTAGTGC	Insertion de la séquence polypyrimidine
	3795 - CATATGCTCAGGTGCTGCTGATCGTTGGTGC 3754 - TCTAGACGCGCAACACCGCACACTAA	Insertion du site 3' d'épissage

Constructions	Oligonucléotide 1 5'-3' Oligonucléotide 2 5'-3'	Utilisation
pXistm et pXistwt	4362 - GGTACCCGAGCGATACATGGTTTGT 4364 - AGATCTAAATGTTCCCCAAAGCTCC	Clonage du promoteur de <i>Xist</i>
	4363 - AGATCTCCCTACCACCAAGCCTT 4366 - AAGCTTCCGCGCAACACCGCACACT	
	4942 - CCCGGGCTCACCCAGAGACACAGACTC	Clonage de la séquence mutée ou sauvage en amont su promoteur de <i>Xist</i> dans le pGL3

### 1.3.3 Les vecteurs utilisés pour la production des protéines recombinantes

Le plasmide pET28 a été utilisé dans le but de produire les protéines PTB, KSRP-4KH et KSRP-34 et permet leur purification grâce à une étiquette « 6xHis » fusionnée à l'extrémité N-terminale (Novagen).

A partir d'un vecteur plasmidique fourni par la société RZPD (Hs. 172550), nous avons amplifié la région des positions +23 à +1710 codant la protéine PTBP1 avec des amorces (couple 4676/4677) permettant d'insérer des sites de restriction (*BamHI*) compatibles avec le sous-clonage dans le vecteur d'expression pET28b appelé alors pET28b-PTB. Nous avons cloné à partir du plasmide pGEX4T1-KSRP-4KH (fourni par F BAKLOUTI, CGMC, Lyon) le fragment codant les 4 domaines KH de la protéine KSRP humaine des positions +387 à +1703 dans le vecteur pET28b en utilisant les sites *EcoRI/XhoI* et généré le plasmide

pET28b-KSRP-4KH et généré ainsi le plasmide. Les séquences des oligonucléotides utilisés pour réaliser ces constructions sont indiquées dans le tableau suivant :

Constructions	Oligonucléotide 1 5'-3'
	Oligonucléotide 2 5'-3'
pET28b-PTB	4676 - GGATTCGACGGCATTGTCCCAGATAT 4677 - CTCGAGTCAGATGGTGGACTTGGAGAAGG

Le vecteur pLox-FRT-PTK a été produit après 2 étapes d'amplification par PCR avec les couples d'oligonucléotides 5607/5608 et 5609/5610 à partir du vecteur PTK fourni par P Clerc (Laboratoire de génétique murine, Institut Pasteur, Paris). Il nous a servi de vecteur de sélection positive/négative lors des expériences de recombinaison médiées par recombinaison. Les séquences des oligonucléotides utilisés pour réaliser ces constructions sont indiquées dans le tableau suivant :

Construction	Oligonucléotide 1 5'-3' Oligonucléotide 2 5'-3'	Utilisation
pLox-FRT-PTK	5607 – AAAAGGTACCAAAATTAATTAATAAAAAACAATTGATAACTTCGTATAATGTATGCT ATACGAAGTTATGAAGTTCCTATTCC	Insertion du site Lox en 5'
	5608 - TTTTCCACCGCGGTGGTTTTGGCGCGCCTTTTGTGCGACATAACTTCGTATAATGT ATGCTATACGAAGTTATGAAGTTCCTATAC	Insertion du site Lox en 3'
	5609 – CTATACGAAGTTATGAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAG GAACTTCCAGTCGGCCGTAAGCTCGAAT	Insertion du site FRT en 5'
	5610 - GCTATACGAAGTTATGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGAAT AGGAACTTCGGTCTTTCCGCCTCAGAAGCC	Insertion du site FRT en 3'

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

## 2. Milieux de culture, tampons et solution 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 de culture LB 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.

**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.

**Blotting buffer 10X** : Tris 500 mM ; glycine 145 mM.

**Solution de transfert** : Blotting buffer 1X ; éthanol 20% (v/v).

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

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

**Solution de bleu 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 désionisée.

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

**Tampon Protéinase K 1 X** : EDTA 25 mM pH 7,4 ; NaCl 0,3 M ; SDS 2% ; Tris-HCl 20 mM pH 7,5.

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

### 3. Anticorps

Tous les anticorps utilisés au cours de ce travail sont répertoriés dans le **Tableau I**. Les anticorps secondaires proviennent des sociétés Invitrogen et Abcam.

Dénomination de l'anticorps	Clonalité	Utilisation	Origine
anti-PTB (NA63)	monoclonal de souris	IF (150°), IP	Calbiochem
anti-PTB (ab5642)	polyclonal de mouton	Western-blot (3000°)	abcam
anti-KSRP (Ab5)	polyclonal de souris	Western-blot (1000°)	D. Black, University of California
anti-Nucléoline	polyclonal de lapin	Western-blot (8000°), IF (100°)	S. Muller, Université de Strasbourg
anti-SAF-A	monoclonal de souris	Western-blot (1000°)	Abcam
anti-Sin3a (sc-994)	polyclonal de lapin	Western-blot (500°), IF (100°)	Santa cruz
anti-Rod1 (sc-15251)	polyclonal de mouton	Western-blot (500°), IF (50°)	Santa cruz
anti-DEK (sc-30213)	polyclonal de lapin	Western-blot (500°), IF (50°)	Santa cruz
anti-TIAR (sc-1749)	polyclonal de mouton	Western-blot (50°), IF (50°)	Santa cruz
anti-lin28a	polyclonal de rat	Western-blot (1000°), IF (1000°)	A. Poleskaya, Université Paris-Sud
anti-FBP1 (sc-48821)	polyclonal de lapin	Western-blot (500°), IF (50°)	Santa cruz
anti-Ezh2 (sc-25383)	polyclonal de lapin	Western-blot (1000°), IP	Santa cruz
anti-Suz12 (ab12073)	polyclonal de lapin	IP	Abcam
anti-RbAp46 (ab3535)	polyclonal de lapin	Western-blot (500°), IP	Abcam
anti-RbAp48 (ab488)	monoclonal de souris	Western-blot (500°), IP	Abcam
anti-Eed (ab4469)	polyclonal de lapin	Western-blot (1000°), IP	Abcam
anti-Suz12 (3C1.2)	polyclonal de mouton	Western-blot (1000°)	Cell signaling
anti-Sam68 (sc-1238)	monoclonal de souris	Western-blot (15000°), IF (250°)	Santa cruz
anti-nucléophosmine (3541)	polyclonal de lapin	Western-blot (8000°), IF (200°)	Cell signaling

#### Tableau I : Anticorps utilisés au cours de ce travail

La dénomination, la clonalité, l'utilisation (IF pour immunofluorescence, IP pour immunoprécipitation) et l'origine des anticorps utilisés au cours de ce travail sont présentées dans ce tableau.

## B. Méthodes

### 1. Méthodes utilisées pour l'analyse de la structure secondaire d'un ARN

#### 1.1 Techniques de cartographie des ARN par utilisation de sondes enzymatiques et chimiques

##### 1.1.1 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 (MgCl <sub>2</sub> 30 mM ; Spermidine 10 mM ; NaCl 50 mM ; Tris-HCl 200 mM, pH 7,5)	6 µl
MgCl <sub>2</sub> (1 M)	0,6 µl
rNTP (25 mM UTP/ATP/GTP/CTP)	5 µl
RNasin (20 UE/µl)	1 µl
ARN polymérase T7	3 µl
H <sub>2</sub> O	qsp 30 µl

Le mélange réactionnel est incubé pendant 4h à 37°C. La matrice ADN est ensuite dégradée par ajout de 1 µl 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 de 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 12 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, repris dans de l'eau stérile puis quantifié.

##### 1.1.2 Analyse de la structure secondaire d'un ARN en solution par utilisation de sondes enzymatiques

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. Pour cette méthode, seuls les ARN non radiomarqués ont été utilisés. Le mélange réactionnel suivant est réalisé :

Transcrit non radiomarqué (1,2 pmol/µl)	1 µl
ARNt de levure (5 µg/µl)	1 µl
Tampon D	3,6 µl
H <sub>2</sub> O	qsp 13 µl
Incubation 10 min à 65°C pour la dénaturation de l'ARN	

MgCl<sub>2</sub> (62,5 mM) 1 µl

Incubation 10 min à température ambiante pour renaturer l'ARN

Les réactions ont été réalisées en présence de 3,125 mM de MgCl<sub>2</sub> comme indiqué ci-dessus, mais également en présence de 10 et 20 mM de MgCl<sub>2</sub>.

Le mélange est ensuite incubé avec l'enzyme (T1 : 0,02 ou 0,0375 U/µl ; T2 : 0,025 ou 0,0375 U/µl et V1 : 2.5x10<sup>-3</sup> ou 5x10<sup>-3</sup> U/µl) et le volume réactionnel est ajusté à 20 µl avec du TpD. 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 µg 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 µg) et de 0,3 M d'Acétate de sodium final. Le culot d'ARN est alors lavé dans 200 µl d'éthanol à 70%, séché, puis resuspendu dans 7µl d'eau. Les ARN modifiés sont stockés à -80°C. L'analyse des coupures enzymatiques se fait par extension d'amorce avec un oligonucléotide radiomarqué sur 1 µl d'ARN clivés, ce qui permet une révélation par fluorographie après électrophorèse sur gel dénaturant.

### 1.1.3 Analyse de la structure secondaire d'un ARN en solution par utilisation de sondes chimiques

Nous avons utilisé les sondes chimiques DMS (diméthylsulfate) et le CMCT (le N-Cyclohexyl-N'-[2-(N-Méthylmorpholino)-éthyl] Carbodiimid-4-Toluonosulfonate). Le DMS modifie spécifiquement les résidus A en position N1A, les résidus C en positions N3C et les résidus G en position N7G. Le CMCT modifie les positions N3U et à un plus faible degré N1-G (**Fig. 37**). Pour l'utilisation de ces sondes chimiques, le mélange réactionnel suivant est réalisé :

Transcrit non radiomarqué (1,2 pmol/µl)	1 µl
ARNt de levure (5 µg/µl)	1 µl
Tampon D	3,6 µl
H <sub>2</sub> O	qsp 13 µl
Incubation 10 min à 65°C pour la dénaturation de l'ARN	
MgCl <sub>2</sub> (62,5 mM)	1 µl

Incubation 10 min à température ambiante

Le mélange est ensuite incubé avec l'agent chimique (DMS : 1 µl d'une solution DMS/EtOH diluée au 1/4 ou au 1/8 (V/V) ; CMCT : 4 ou 5 µl d'une solution à 180 mg/ml) et le volume réactionnel est ajusté à 20 µl avec du tampon D. Les modifications par le DMS sont arrêtées après 10 min d'incubation à 30°C par ajout du tampon Stop DMS (AcNa 0,15 M ; β-mercaptoéthanol 0,1 M ; EDTA 0,01 mM ; Tris acétate 0,1 M pH 7,5). Celles par le CMCT sont arrêtées par ajout de 20 µg d'ARNt de levure. Un témoin sans agent modifiant est également réalisé. Le culot d'ARN est alors lavé dans 200 µl d'éthanol à 70%, puis séché, repris dans 7 µl d'eau stérile et stocké à -80°C. Comme dans le cas des coupures



enzymatiques, l'analyse des modifications se fait par extension d'amorce avec un oligonucléotide radiomarqué sur 1 µl d'ARN modifié.

#### 1.1.4 Analyse des nucléotides protégés contre l'action des RNases par les protéines d'un extrait nucléaire ou par des protéines recombinantes

Pour la mise en évidence des nucléotides protégés contre l'action des RNases par les composants d'un extrait nucléaire ou par une protéine recombinante, le mélange réactionnel suivant est réalisé :

Transcrit non radiomarqué (1,2 pmol/µl)	1 µl
ARNt de levure (5 µg/µl)	1 µl
Tampon D	3,6 µl
H <sub>2</sub> O	qsp 13 µl
Incubation 10 min à 65°C pour la dénaturation de l'ARN	
MgCl <sub>2</sub> (62,5 mM)	1 µl

Incubation 10 min à température ambiante pour renaturer l'ARN

Le mélange est ensuite incubé 10 min à 30°C en présence d'extrait nucléaire ou de protéines recombinantes en quantité variable dans un volume réactionnel de 20 µl. Ce mélange est ensuite incubé avec l'enzyme selon les conditions décrites ci-avant et les réactions sont arrêtées après 10 min d'incubation à 30 °C. L'arrêt des réactions est suivi par une extraction au phénol/chloroforme et par une précipitation à l'éthanol 96% en présence de glycogène (10 µg) et de 0,3 M d'acétate de sodium final. Le culot d'ARN est alors lavé dans 200 µl d'éthanol à 70%, séché, puis resuspendu dans 7 µl d'eau stérile. Les ARN modifiés sont stockés à -80°C. L'analyse des coupures enzymatiques se fait par extension d'amorce.

#### 1.1.5 Analyse par extension d'amorce

- Marquage de la sonde

ATP [ $\gamma$ <sup>32</sup> P] (Perkin Elmer, 5 mCi/ml, 3000 Ci/mmol)	1 µl
Oligonucléotide à marquer (15 pmol/µl)	1 µl
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 1 mM)	1 µl
T4 polynucléotide kinase (MBI Fermentas 10U/µl)	1 µl
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 300 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 ou modifié	1 µl

Tampon de l'enzyme AMV 10X (Tris-HCl 50 mM pH 8,3 ; KCl 40 mM ; MgCl <sub>2</sub> 6 mM)	0,25 µl
H <sub>2</sub> O stérile	qsp 2,5 µl
Incubation 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 (voir ci-dessus)	0,25 µl
AMV Réverse transcriptase (Quantum Appligene 2U/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 sont ajoutés au mélange de la réaction d'extension à raison d'une concentration de 0,5 mM afin que son incorporation soit statistique. 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.

## 1.1.6 Recherche de modèles de structure secondaire : utilisation du logiciel MFold

Ce logiciel, fourni par le Genetics Computer Group de l'Université de Madison (USA), est basé sur le programme de recherche de structures secondaires et a été établi par Jaeger *et al.* (Jaeger *et al.*, 1989). Ce logiciel tient compte des estimations de la stabilité thermodynamique des appariements de bases de l'ARN à 0°C dans une solution de NaCl 1M. L'intérêt d'utiliser la version Squiggles est de pouvoir imposer, au cours de la recherche d'une structure secondaire potentielle, des contraintes de structuration en simple et double brin, correspondant aux données acquises expérimentalement. Ce logiciel est disponible à l'adresse suivante : <http://mfold.bioinfo.rpi.edu/>.

## 1.2 Technique de FRET utilisée pour apprécier la distance relative séparant 2 zones ARN d'intérêt

Les spectres d'émission de fluorescence sont enregistrés à 4°C à une longueur d'onde d'excitation de 515 nm et s'échelonnant de 500 à 750 nm. Les bandes passantes d'excitation et d'émission sont de 3 nm. La procédure est adaptée de Gavory *et al.*, 2006. L'ARN et l'oligonucléotide couplé au Cy3 sont incubés à un ratio molaire d'1:1 dans 160 µl de tampon Citrate (150 mM NaCl ; 3,25 mM MgCl<sub>2</sub> ; 5 mM Na citrate pH 7,0). La concentration finale est de 0,38 µM. Le mélange est incubé 5 min à 85°C et renaturé par diminution lente de la température (15 min) jusqu'à température ambiante. Après 4 h d'incubation à 4°C, le rendement d'association des 2 composantes est déterminé par électrophorèse en gel de polyacrylamide non dénaturant 7%. La fluorescence dans le gel est mesurée avec un scanner

Typhoon 9410 Healthcare. Lorsque le rendement d'association est satisfaisant, l'émission de la sonde couplée au Cy3 est mesurée par un fluoromètre SAFAS. Dix spectres sont enregistrés et moyennés. Puis, l'oligonucléotide-Cy5 est ajouté à un ratio molaire d'1:1. Le mélange est incubé 4 h à 4°C. Dix spectres sont à nouveau enregistrés et le FRET pour la paire Cy3-Cy5 est calculé en prenant en compte le ratio des oligonucléotides liés et non-liés. Chaque expérience de FRET est répétée au moins 3 fois avec différentes préparations d'ARN et d'oligonucléotides. Les oligonucléotides qui ont servis au cours de cette méthodologie sont présentés dans le **chapitre A** des résultats.

## **2 Méthodes utilisées pour la culture cellulaire**

### 2.1 Culture cellulaire

#### 2.1.1 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 foetal (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>.

#### 2.1.2 Lignée cellulaire ES XX de type HP3-10

Cette lignée est cultivée dans le milieu DMEM glutamax (Dulbecco's Modified Eagle Medium) supplémenté avec 20% de sérum de veau foetal (Invitrogen), 10 U/ml de pénicilline/streptomycine, 2,9 mg/ml de glutamine (Gibco) et en présence de 0,5 mM de 2β-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 une étuve 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.

#### 2.1.3 Entretien des cellules

L'entretien et l'amplification des cellules s'effectuent par culture dans des flasques de 75 cm<sup>2</sup>. La confluence des cellules est vérifiée par microscopie. Le milieu de culture est retiré et 4 ml d'une solution de trypsine (0,5 g/l)/EDTA (0,2 g/l) sont ajoutés. 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 5 min à 800 g. Elles sont ensuite resuspendues dans 10 ml de milieu etensemencées dans une nouvelle flasque de culture avec 1 ml de cette suspension et 15 ml de milieu. Elles sont mises en culture selon les conditions présentées ci-avant jusqu'à confluence (2 à 3 jours)

#### 2.1.4 Technique de transfection

Pour les lignées HEK et HeLa, un nombre nécessaire de plaques de 6 puits estensemencé avec environ 400 000 cellules (3 ml) par puits. Elles sont ensuite incubées dans un incubateur à CO<sub>2</sub> à 37°C pendant une nuit. Le lendemain, les cellules sont transfectées à l'aide du kit de transfection JetPEI (PolyPlus Transfection), selon les recommandations du fabricant. Après 5 h d'incubation avec l'ADN, les cellules sont lavées et remises en culture pendant 24 à 48 h.

## 2.2 Technique d'immunofluorescence

### 2.2.1 Culture sur lames, fixation et perméabilisation des cellules

Au préalable, les lames sont gélatinisées et les cellules en cultures sont rincées au PBS 1X et décollées des supports plastiques par ajout de trypsine selon le protocole décrit ci-avant. Le culot cellulaire est repris dans 2 ml de milieu de culture puis les cellules sont comptées. 300000 cellules par lame sontensemencées et mises en culture durant une nuit.

Puis, les lames sont lavées au PBS 1X puis recouvertes d'ortho-paraformaldéhyde à 3,5% (PFA 3,5%) pendant 20 min sur un lit de glace. La PFA 3,5% est ensuite aspirée et les lames sont rincées au PBS 1X à 3 reprises pendant 5 min. Les cellules sont perméabilisées par leur mise en contact avec une solution de PBS 1X contenant du Triton X100 à 0,1% pendant 4 min à température ambiante. Les lames sont ensuite rincées et lavées à 3 reprises dans du PBS 1X pendant 5 min à température ambiante.

### 2.2.2 RNA-FISH et immunodétection de protéines

- RNA-FISH

#### *Préparation des lames*

Pour les expériences de RNA-FISH, un inhibiteur de RNAses (VRC, Vanadyl ribonucleoside complexes à une dilution au 1000<sup>o</sup>, invitrogen) est ajouté à la solution de perméabilisation. Les lames sont ensuite lavées au tampon SSC 2X à 2 reprises pendant 5 min.

#### *Marquage des sondes*

Les sondes doivent être marquées au préalable par un fluorochrome. Par nick translation, un nucléotide couplé à un groupement fluorophore est incorporé dans le brin d'ADN (approximativement 2 kb en amont du promoteur du gène *Xist* jusqu'au milieu de l'exon 6)

complémentaire à l'ARN à détecter. Le mélange réactionnel réalisé est le suivant selon les recommandations du fournisseur (Abbot) :

Sondes (1 µg/µl)	1 µl
H <sub>2</sub> O	17,5 µl
dTTP-fluorophore (0,2 mM)	2,5 µl
dTTP (0,1 mM)	5 µl
dNTP (0,5 mM ATP/GTP/CTP)	10 µl
Tampon 10X (50 mM Tris HCl pH 7,5 ; 10 mM MgSO <sub>4</sub> ; 0,1mM dithiothreitol, 50 µg/ml BSA)	5 µl
ADN polymérase I (0,5 U/µl)	10 µl
Incubation toute une nuit à 16°C puis conservation à -20°C	

#### *Préparation des sondes pour un spot*

Avant d'hybrider les sondes marquées sur les lames, il est important d'y ajouter de l'ADN de sperme de saumon qui permet de saturer les liaisons non spécifiques et de l'ADN Cot-1 qui masque les régions génomiques répétées. Ce dernier permet également de réduire les liaisons aspécifiques que pourraient avoir les sondes avec les acides nucléiques cellulaires.

Sondes marquées (environ 10 ng/µl)	10 µl
Cot-1 (2,5 µg/µl)	10 µl
ADN de sperme de saumon (10 µg/µl)	3 µl
H <sub>2</sub> O	qsp 50 µl

Le mélange est ensuite précipité à l'éthanol 96% en présence de 0,3 M d'acétate de sodium final. Le culot est ensuite lavé dans 200 µl d'éthanol à 70%, séché, puis resuspendu dans 10 µl de formamide. Le mélange est ensuite dénaturé par une incubation de 10 min à 75°C sous agitation constante. Puis 10 µl de tampon d'hybridation 2X sont ajoutés (SSC 4X ; 20% dextran sulfate ; 2 mg/ml BSA ; 400 mM VRC). Les sondes sont prêtes à être hybridées.

#### *Hybridation*

Les sondes sont spottées sur la lame et incubées dans une chambre humide (SSC 2X/formamide 50%) pendant une nuit à 37°C.

- Immunodétection de protéines

Les lames sont mises en contact avec une solution de pré-blocage (PBS 1X ; Lait 5% ; TritonX100 0,1%) pendant 15 min à température ambiante. Puis, cette solution est supprimée et l'anticorps primaire (dilué dans cette dernière) est spotté sur la lame et incubé pendant 30 min à température ambiante dans une chambre humide. Après 3 lavages successifs (1 rinçage dans la solution de pré-blocage, 20 min à température ambiante dans une solution PBS 1X ; Lait 5% puis un dernier lavage pendant 20 min à température ambiante dans du PBS 1X), l'anticorps secondaire (dilué dans la solution de pré-blocage) est spotté sur la lame pendant 30 min à température ambiante dans une chambre humide. Après 3 nouvelles étapes de lavage précédemment décrites, les lames sont prêtes à être montées.

### 2.2.3 Montage des lames et acquisition des images

Les lames sont préparées selon les recommandations du fournisseur du kit de montage (Vectashield de Vector laboratories) et l'acquisition des images est réalisée avec un microscope de type Zeiss Axioplan.

### 2.3 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 suppression du surnageant, le volume cellulaire culoté (packed cell volume ou pcv) est mesuré. Celui-ci est resuspendu dans environ 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 supprimé 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 de faire gonfler les 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 resuspendus 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 additionné afin de détruire les membranes nucléaires. L'ensemble est homogénéisé avec le dounce. L'échantillon est incubé pendant 30 min à 4°C 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. L'extrait nucléaire recueilli est dosé, aliquoté et conservé à -80°C.

## 3 Méthodes utilisées pour purifier des complexes RNP

### 3.1 Techniques de 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 et de la protéine MS2 (Maltose Binding Protein) L'ARN (100 pmol) est repris dans 50 µl 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 µl avant d'être incubé 20 min dans la glace. Le complexe formé [ARN-MS2/MBP] est ensuite fixé sur 40 µl 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 µl 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 µM d'ARNt de levure. Les billes sont à nouveau isolées par centrifugation d'1 min à 11000 rpm, puis lavées 3 fois avec 500 µl de TpD pendant 10 min. Les complexes RNP sont élués avec 70 µl de tampon D complété en maltose à raison de 10 mM final lors d'une incubation de 30 min sous agitation constante à 4°C.

Puis, 20 µl de bleu SBL (Tris pH 6,8 250 mM ; SDS 8% ; β-mercaptoéthanol 20% ; bleu de bromophénol 0,04%) sont ajoutés à 50% de l'éluat et les protéines sont fractionnées sur gel SDS-PAGE 10% Acryl-bisacryl (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.

### 3.2 Techniques d'immunosélection des complexes RNP

#### 3.2.1 Principe

Différents fragments radiomarqués de la région des A-repeats sont incubés en présence d'un extrait nucléaire de cellules ES dans le but de former des complexes RNP. Puis, les protéines du complexe PRC2 et la protéine PTB sont immunosélectionnées grâce à des anticorps spécifiques : 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). La présence et l'identité des ARN présents dans les complexes RNP immunoprécipités sont détectées par fluorographie.

#### 3.2.2 Préparation des ARN

Pour cette méthode, les ARN sont radiomarqués à leur extrémité 5', le mélange réactionnel suivant est réalisé :

##### *Déphosphorylation de l'extrémité 5' des ARN*

ARN à déphosphyler (1 µg/µl)	1 µl
Tampon CIP 10X (Tris-HCl 0,1 M, pH 7,5 ; MgCl <sub>2</sub> 0,1 M)	1 µl
Calf Intestine Alkaline Phosphatase (Fermentas) (1U/µl)	1 µl
H <sub>2</sub> O	qsp 10 µl

Les ARN sont incubés 1 h à 37°C puis 5 min à 65°C. Ils sont ensuite extraits au phénol/chloroforme, puis précipités comme précédemment et repris dans 10 µl d'eau stérile.

##### *Marquage en 5' des ARN*

ARN déphosphorylé (100 ng/µl)	2,5 µl
ATP [ $\gamma$ <sup>32</sup> P] (Perkin Elmer, 3000 Ci/mmol ; 5 µCi/µl)	1 µl
Tampon PNK (Tris-HCl 0,5 M, pH 7,6 ; MgCl <sub>2</sub> 100 mM ; DTT 50 mM ; Spermidine 1 mM ; EDTA 1 mM)	1 µl
T4 polynucléotide kinase (MBI Fermentas, 10 U/µl)	1 µl
H <sub>2</sub> O	qsp 10 µl

Le mélange est incubé 45 min à 37°C, puis les transcrits sont purifiés par électrophorèse en gel de polyacrylamide dénaturant, élués une nuit à 4°C sous agitation constante dans le tampon d'éluion, et précipités comme décrit précédemment.

### 3.2.3 Préparation des billes

Quarante microlitres de billes G-sépharose sont centrifugés 1 min à 1000 rpm. Elles sont ensuite lavées à 3 reprises avec 500 µl de tampon d'immunosélection contenant NaCl 150 mM ; Tris-HCl 10 mM pH 8,0 ; Igépal 0,1%. Puis, 10 µl d'anticorps, 1,5 µl de BSA (10 mg/ml) sont ajoutés aux billes et le volume est ajusté à 300 µl avec le tampon d'immunosélection. Le mélange est incubé 2 h à 4°C sous agitation constante. Les billes sont ensuite centrifugées 1 min à 1000 rpm et lavées à 3 reprises dans 500 µl de tampon d'immunosélection.

### 3.2.4 Formation des complexes RNP, immunosélection et analyse

Pour la formation des complexes RNP, le mélange réactionnel réalisé est le suivant :

Transcrit radiomarqué en 5' (35 pmol/µl)	2 µl
ARNt de levure (20 µg/µl)	1 µl
Tampon D	2 µl
H <sub>2</sub> O	qsp 9 µl
Incubation 5 min à 65°C pour la dénaturation de l'ARN	
MgCl <sub>2</sub> (62,5 mM)	1 µl
Incubation 5 min à température ambiante pour renaturer l'ARN	
Incubation 5 min à 4°C	
Extrait nucléaire	30 µg
Inhibiteur de RNases (RNasin, 20 U/µl, Amersham Biosciences)	1 µl
Tampon D	qsp 30 µl

Incubation 30 min à température ambiante sous agitation ambiante

Les complexes RNP sont ensuite incubés avec les billes G-sépharose pendant 2 h à 4°C sous agitation dans un volume ajusté à 300 µl avec du tampon d'immunosélection. Puis, les billes sont centrifugées 1 min à 1000 rpm et le surnageant est supprimé. Après 3 lavages de 10 min à 4°C sous agitation avec 750 µl de tampon de lavage contenant NaCl 150 mM ; Tris-HCl 10 mM pH 8,0 ; Igépal 0,5%, les ARN sont extraits au phénol/chloroforme et précipités à l'éthanol 96% en présence de 0,3 M d'acétate de sodium final. Le culot est repris dans 15 µl de bleu de formamide et fractionné en gel de polyacrylamide/urée dénaturant à 7%. Après migration et séchage, le gel est placé en fluorographie une nuit à -80°C.



### 3.3 Analyses des complexes ARN/Protéines par pontage covalent aux UV suivi d'immunosélection

La formation des complexes RNP est réalisée comme décrit précédemment lors des expériences d'immunoprécipitation des complexes RNP. A noter seulement que les transcrits doivent être marqués uniformément selon le protocole présenté dans la section retard sur gel.

#### 3.3.1 Préparation des billes G-sépharose

Les billes sont préparées selon le protocole présenté dans la section précédente.

#### 3.3.2 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 température ambiante est nécessaire, puis les puits sont lavés à 2 reprises en tampon D. Puis, la plaque est placée sur un lit de glace.

#### 3.3.3 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
Incubation 5 min à 65°C pour la dénaturation de l'ARN	
MgCl <sub>2</sub> (62,5 mM)	1 µl
Incubation 5 min à température ambiante pour renaturer l'ARN	
Incubation 5 min à 4°C	
Extrait nucléaire	30 µg
Inhibiteur de RNases (RNasin, 20 U/µl, Amersham Biosciences)	1 µl
Tampon D	qsp 30 µl
Incubation 30 min à température ambiante sous agitation ambiante	

#### 3.3.4 Pontage covalent aux UV entre les protéines et les ARN et immunosélection

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>. 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 µl de bleu SBL 4X. Cet échantillon correspond à l'input avant immunoprécipitation. L'immunosélection est réalisée selon le protocole présenté dans la section précédente. Puis, les ARN sont élués par ajout de 15 µl de bleu SBL 4X sur les billes culotées. Elles sont ensuite vortexées et chauffées 5 min à

96°C. L'échantillon est centrifugé pendant 5 min à 13000 rpm. Le surnageant est fractionné sur gel SDS-PAGE 10% et analysé par fluorographie.

## 4 Production de protéines recombinantes chez *E.coli*

### 4.1 Purification de la protéine PTB

#### 4.1.1 Transformation, culture bactérienne et auto-induction

Des bactéries *E. coli* codon plus (DE3) sont transformées avec 10 à 20 ng de plasmides de type pET28 (étiquette histidine) portant les séquences codant la protéine PTB. Une préculture de 20 mL de LB contenant 50 µg/mL de kanamycine estensemencée par un clone bactérien issu de la transformation et incubée à 37°C sous agitation pendant 18 heures. Une culture de 2 L de milieu ZYM 5052 (Tryptone 10 g ; extrait de levure 5 g ; Na<sub>2</sub>HPO<sub>4</sub> 4.5 g ; KH<sub>2</sub>PO<sub>4</sub> 3.5 g ; NH<sub>4</sub>Cl 2.5 g ; Na<sub>2</sub>SO<sub>4</sub> 1 g ; H<sub>2</sub>O qsp 1 l ; sucres : Glycérol 100 % (v/v) 250 g ; Lactose 100 g ; Glucose 25 g et des métaux : MgSO<sub>4</sub> 1 mM ; MgSO<sub>4</sub> 1 mM ; FeCl<sub>3</sub> 10 µM ; CaCl<sub>2</sub> 2 µM ; 1 µM de MnCl<sub>2</sub> et de ZnSO<sub>4</sub> ; et 0,2 µM de CoCl<sub>2</sub> ; CuCl<sub>2</sub> ; NiCl<sub>2</sub> ; Na<sub>2</sub>MoO<sub>4</sub> ; Na<sub>2</sub>SeO<sub>3</sub> et de H<sub>3</sub>BO<sub>3</sub>) est complétementée avec ces 20 ml de préculture et mise à incuber pendant 18 heures à 37°C sous agitation. L'induction est vérifiée par fractionnement des protéines totales par électrophorèse sur gel SDS-PAGE 10% (Gel de séparation : Acrylamide/Bisacrylamide 40% (p/v) ; SDS 0,1% (v/v) ; Tris-HCl 375 mM ; pH 8,8. Gel de Concentration : Acrylamide/Bisacrylamide 40% (p/v) ; SDS 0,3% (v/v) ; Tris HCl 166,7 mM ; pH 6,8) de 100 µl de culture centrifugés 1 min à 11000 rpm et repris dans 30 µl d'eau et 10 µl de bleu SBL4X.

#### 4.1.2 Préparation des extraits cellulaires

La culture de 2 L est centrifugée pendant 20 min à 4000 rpm à 4°C et le culot est repris dans 20 ml de tampon de sonication (KPi 20 mM pH 7,4 ; 1M KCl). 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 complétementé avec 20 mM d'imidazole et d'un mélange d'inhibiteurs de protéases 1X puis il est filtré à 0,45 µM.

#### 4.1.3 Chromatographie d'affinité sur colonne His-Trap couplée à un système FPLC

Le surnageant de sonication est déposé sur une colonne Sépharose His-trap 20 ml (GE Healthcare) chargée en ions cobalt présentant une forte affinité pour l'histidine. Après lavage avec 5 volumes colonne de tampon de sonication, les protéines sont éluées par ajout de tampon de sonication-imidazole 1M. L'imidazole, qui entre en compétition avec l'histidine pour fixer le cobalt, est ajouté par application d'un gradient découpé en 4 paliers successifs de 50 mM, 250 mM, 500 mM et 1 M. Les fractions contenant la protéine PTB sont analysées sur

gel SDS-PAGE. Les fractions sont ensuite regroupées, dialysées contre du tampon D puis concentrées dans un volume de 500 µl selon le kit « Amicon ultra ».

#### 4.1.3.1.1 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<sup>TM</sup> 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.

## 4.2 Purification des protéines KSRP-4KH

### 4.2.1 Transformation et culture bactérienne

Des bactéries *E. coli* BL21(DE3) sont transformées avec 10 à 20 ng de plasmides de type pET28 (étiquette histidine) portant les séquences codant les protéines à produire. 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) dans 2 L de milieu de culture pendant 4 h à 37°C sous agitation.

### 4.2.2 Préparation des extraits cellulaires et purification des protéines KSRP-4KH

La culture de 2 L est centrifugée pendant 20 min à 4000 rpm à 4°C et le culot est repris dans 20 ml de tampon de sonication (NaCl 200 mM ; Tris-HCl 20 mM ; pH 7,8). Les étapes de sonication, de centrifugation et de préparation de l'échantillon suivent le protocole présenté précédemment. Néanmoins, 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 la colonne sont éluées avec 10 ml de tampon de sonication contenant 500 mM d'imidazole et récupérées par fractions de 500 µl 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. La quantité et la qualité des échantillons protéiques sont appréciées par la méthodologie précédemment décrite pour la protéine PTB. Les aliquots protéiques sont conservés à -80°C.

## 5 Analyses des interactions ARN/Protéines *in vitro*

### 5.1 Analyses des complexes ARN/Protéines par expériences de retard sur gel et empreinte aux RNAses

#### 5.1.1 Préparation des ARN

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

Plasmide linéarisé (500 ng/ $\mu$ l)	1 $\mu$ l
Tampon 5X (MgCl <sub>2</sub> 30 mM ; Spermidine 10 mM ; NaCl 50 mM ; Tris-HCl 200 mM, pH 7,5)	6 $\mu$ l
MgCl <sub>2</sub> (1 M)	0,6 $\mu$ l
rNTP (25 mM UTP/ATP/GTP + 8 mM CTP)	5 $\mu$ l
[ $\alpha$ - <sup>32</sup> P]-CTP (800 Ci/mmol, 10 mCi/mL, Perkin Elmer)	1 $\mu$ l
RNasin (20 UE/ $\mu$ L)	1 $\mu$ l
ARN polymérase T7	3 $\mu$ l
H <sub>2</sub> O	qsp 30 $\mu$ l

Le mélange réactionnel est incubé pendant 3 h à 37°C. La matrice ADN est ensuite dégradée par ajout de 1  $\mu$ l de DNase RQ1 (1 UE/ $\mu$ l, 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 300  $\mu$ L de tampon d'éluion. Les ARN sont alors élués à 4°C pendant 12 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.

#### 5.1.2 Formation des complexes ARN/Protéines

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

Transcrit non radiomarqué (1,2 pmol/ $\mu$ l)	1 $\mu$ l
ARNt de levure (5 $\mu$ g/ $\mu$ l)	1 $\mu$ l
Tampon D	3,6 $\mu$ l
H <sub>2</sub> O	qsp 13 $\mu$ l
Incubation 10 min à 65°C pour la dénaturation de l'ARN	
MgCl <sub>2</sub> (62,5 mM)	1 $\mu$ l

Incubation 10 min à température ambiante pour renaturer l'ARN

Le mélange est ensuite incubé différemment en présence d'extrait nucléaire ou de protéines recombinantes en quantité variable (mais toujours ajusté à un volume réactionnel de 20  $\mu$ l) selon 2 cas de figures.

**Cas 1** : les complexes sont analysés par des expériences de gel retard.

Dans ce mélange est ajouté 1  $\mu$ l du même transcrit à 2500 cpm/ $\mu$ l puis incubé 30 min à 4°C afin de permettre la formation des complexes ARN-protéine. Après avoir ajouté 3  $\mu$ l 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. Ainsi, les complexes ARN/protéine formés sont retardés dans leur migration par rapport aux ARN libres. Le résultat de la réaction de retard sur gel est visualisé après une nuit de fluorographie.

**Cas 2 :** les complexes sont analysés par des expériences d'empreintes aux RNAses.

Le mélange réactionnel est incubé avec l'enzyme selon les conditions décrites ci-avant et les réactions sont arrêtées après 10 min d'incubation à 30 °C. L'arrêt des réactions est suivi par une extraction au phénol/chloroforme et par une précipitation à l'éthanol 96% en présence de glycogène (10 µg). Le culot d'ARN est alors lavé dans 200 µl d'éthanol à 70%, séché, puis resuspendu dans 7 µl d'eau stérile. Les ARN modifiés sont stockés à -80°C. L'analyse des coupures enzymatiques se fait par extension d'amorce.

## 5.2 Analyses des complexes ARN/Protéines par dichroïsme circulaire

Le dichrographe utilisé est un CD6 Jobin Yvon. Les mesures sont réalisées avec un incrément de 0,2 nm/s avec un volume d'échantillon de 550 µl dans chaque compartiment d'une cuve de 2 x 0,437 cm de trajet optique. Le spectre du tampon (Tris-HCl 10 mM pH 7,5; NaCl 150 mM ; filtré à 0,22 µM) est réalisé avant chaque mesure puis il en est soustrait. Deux spectres sont effectués pour chaque mesure puis ils sont moyennés pour minimiser le bruit de fond. Pour la réalisation du spectre de la protéine et de l'ARN, nous avons utilisé une concentration de 0,2 µM. Le résultat est présenté en  $\Delta\epsilon \text{ M}^{-1}.\text{cm}^{-1}$  par résidus selon la loi de Beer-Lambert. La sensibilité de l'appareil est de  $10^{-6} \Delta A$  soit 0,03  $\Delta\epsilon \text{ (M}^{-1}.\text{cm}^{-1}$  par nucléotide).

Pour l'étude de l'interaction, le spectre «avant interaction» a été réalisé en déposant l'ARN et la protéine d'intérêt séparément de chaque côté de la cuve à une concentration de 0,2 µM. Puis le spectre «après interaction» est réalisé après mélange du contenu de la cuve. Le résultat est présenté en  $\Delta\epsilon$  par acide nucléique car le pic à 260 nm caractéristique des ARN est indépendant de l'absorption des protéines.

## 6 Techniques utilisées pour étudier l'épissage de la région des A-repeats

### 6.1 Méthodes spécifiques de l'étude de l'épissage *in vitro*

#### 6.1.1 Préparation de la matrice de transcription

Afin de limiter la transcription à la région étudiée, le plasmide matrice est préalablement hydrolysé par une enzyme de restriction adéquate. L'ADN est ensuite soumis à deux extractions, l'une au phénol/chloroforme (1/1) et la seconde par un mélange chloroforme/alcool isoamylique (24/1). Puis l'ADN est précipité à l'éthanol en présence d'acétate de sodium 0,3 M et finalement repris dans l'eau.

6.1.2 Transcription *in vitro*

Le protocole expérimental utilisé pour la transcription des ARN d'intérêt est le suivant :

Matrice linéarisée (200 ng/μl)	10 μl
Tp 5X (MgCl <sub>2</sub> 30 mM ; Spermidine 10 mM ; NaCl 50 mM ; Tris-HCl 200 mM, pH 7,5)	8 μl
rNTP (A, C, T, G ; 5, 5, 5, 2 mM)	4 μl
Cap 5m7GpppG (10 mM)	2 μl
Inhibiteur de RNases (RNasin, 20 U/μl, Amersham Biosciences)	2 μl
BSA (2 mg/ml)	2 μl
DTT (200 mM)	2 μl
ARN polymérase T7	2 μl
H <sub>2</sub> O	8 μl
Incubation pendant 2 h à 37°C	

Après une étape de dégradation de la matrice ADN comme décrit précédemment, les ARN sont soumis à deux précipitations successives dans l'éthanol absolu, la première se faisant en présence d'acétate d'ammonium 0,25 M et la seconde, en présence de NaCl 0,2 M. Ces étapes de précipitations permettent d'éliminer les nucléotides non incorporés. Le précipité d'ARN est lavé deux fois dans de l'éthanol à 70%, puis séché, repris dans de l'eau stérile et la quantité de transcrits obtenue est déterminée.

6.1.3 Epissage *in vitro*

Les expériences d'épissage *in vitro* sont réalisées avec des extraits nucléaires de cellules HeLa provenant de la société CilBiotech. Le mélange réactionnel est le suivant :

ARN pré-messager (10 fmol/μl)	2 μl
MgCl <sub>2</sub> (62,5 mM)	1 μl
ATP (25 mM)	1 μl
Créatine phosphate (625 mM)	1 μl
Inhibiteur de RNases (RNasin, 20 U/μl, Amersham Biosciences)	0,5 μl
Tampon D	3,5 μl
Alcool polyvinylique 13%	6 μl
Extrait nucléaire (10 mg/ml, traité au préalable 10 min à 30°C)	8 μl

Incubation 5 min à 0°C puis 2h à 30°C

La réaction d'épissage est ensuite arrêtée par ajout de 20 μg de protéinase K (Roche) fonctionnant dans le tampon suivant : EDTA 12,5 mM ; NaCl 150 mM ; SDS 1% ; Tris-HCl 100 mM, pH 7,4. Après une incubation de 30 minutes à 30°C, les protéines sont extraites au phénol/chloroforme (1/1), puis au chloroforme/alcool isoamylique (24/1) et les ARN sont précipités en présence d'éthanol absolu sans ajout de sels. Le précipité est lavé puis séché et repris dans 10 μl d'eau stérile.

## 6.1.4 Conversion des ARN en ADN complémentaires par RT-PCR

*Hybridation de l'oligonucléotide complémentaire à l'extrémité 3' des ARN*

ARN	2 μl
Amorce (5 μM)	1 μl

H <sub>2</sub> O	13 µl
Incubation 5 min à 65 °C puis 4°C	
<i>Réaction de transcription inverse</i>	
dNTP (10 mM de chaque)	3 µl
Tp 5X (Tris/HCl 250 mM pH 8,3 ; KCl 375 mM ; MgCl <sub>2</sub> 15 mM ; DTT 50 mM)	5 µl
Inhibiteur de RNases (RNasin, 20 U/µl, Amersham Biosciences)	1 µl
Transcriptase inverse (MMLV, 200 U/µl, Promega)	1 µl
Incuber 1H à 42 °C puis conserver à -20°C	
<i>Amplification des ADN complémentaires synthétisés</i>	
Matrice	2 µl
dNTP (10 mM)	1 µl
Tp10X (Tris-HCl 10 mM, pH 8,8 ; KCl 50 mM ; Triton X-100 0.1 %)	5 µl
Amorces (25 mM de chaque)	1 µl
GoTaq (Promega, 5 U/µl)	0,5 µl
MgCl <sub>2</sub> (25 mM)	2 µl
H <sub>2</sub> O	qsp 50 µl

Le mélange est soumis au programme PCR suivant :

Dénaturation : 5 min à 96°C	} cycle reproduit 29 fois
Dénaturation : 45 sec à 96°C	
Hybridation : 45 sec à 58°C	
Elongation : 1 min à 72°C	
Terminaison : 10 min à 72°C	

Les produits PCR amplifiés sont ensuite analysés par électrophorèse en gel d'agarose 1% après l'ajout de bleu de dépôt agarose (6X). Les séquences des oligonucléotides utilisés pour les expériences de RT-PCR sont répertoriées dans le tableau suivant :

Oligonucléotide 5'-3'	Utilisation
3899 - GGCTCGAGAATAGCCGTA	RT
3900 - CCACCAAGCCTTATGGCT	PCR
3898 - ACGCAGAAGCCATAATGGC	

## 6.2 Méthodes spécifiques de l'étude de l'épissage *in vivo*

### 6.2.1 Dosage de l'activité luciférase

Les cellules sont transfectées par un plasmide portant le gène de la luciférase « Fire-Fly » sous contrôle du promoteur Xist. La mesure de l'activité luciférase se fait de la manière suivante : les cellules transfectées sont lavées à 2 reprises avec du PBS 1X. Puis, 250 µl de tampon de lyse (Cell Lysis Buffer, Promega) sont ajoutés dans chaque puits et incubés pendant 15 à 20 min à température ambiante. Le lysat est récupéré puis centrifugé 5 min à 13000 rpm et le surnageant est conservé à -20°C. Dans une cuve du luminomètre, 50 µl de la solution du substrat luciférase « fire-fly » (Promega) sont mélangés à 5 µl de lysat cellulaire. La cuve est ensuite placée dans l'appareil et l'émission de la lumière est mesurée durant 10 s.

### 6.2.2 Extraction des ARN totaux et RT-PCR

Les cellules HeLa transfectées sont lavées au PBS 1X et les ARN totaux sont extraits des cellules au Trizol selon les spécifications du fournisseur (Invitrogen). Après une étape de dégradation des ADN par utilisation de la DNase RQ1 (décrit précédemment), les ARN d'intérêt sont convertis en ADN complémentaires selon le protocole de RT-PCR exposé précédemment. Seule la quantité de matrice ARN utilisée pour l'étape de transcription inverse varie et atteint 4 µg. En effet, la population en ARN d'intérêt par rapport à la totalité des ARN cellulaires est beaucoup plus faible que lorsqu'on travaille avec des ARN synthétiques. ). Les séquences des oligonucléotides utilisés pour les expériences de RT-PCR sont répertoriées dans le tableau précédent.



## **RÉSULTATS et DISCUSSION**

### III. Résultats et Discussion

#### A. Détermination expérimentale de la structure 2D de la région des A-repeats des ARN Xist murin et humain

##### 1. Introduction

Nous avons déjà explicité dans le chapitre « Introduction », l'importance d'étudier expérimentalement la structure 2D de la région des A-repeats de l'ARN Xist, ceci afin de mieux comprendre son mécanisme d'action. Si cette structure n'avait pas encore été étudiée expérimentalement, c'est parce que c'était un travail difficile à réaliser du fait de la présence de répétitions de séquences et de longs segments riches en résidus U. Pour y arriver, nous avons dû non seulement utiliser toute l'expertise du laboratoire en matière d'étude de la structure 2D des ARN, mais en plus introduire une nouvelle approche basée sur la technique de FRET en collaboration avec M Blaud (post-doctorante au laboratoire).

Par ailleurs, pour avoir un argument en faveur de la validité de la structure 2D établie pour la région des A-repeats de l'ARN murin, nous avons aussi étudié expérimentalement la structure de la région des A-repeats humaine. Les éléments reliant les motifs répétés ayant des séquences fortement divergentes entre les deux espèces, l'observation de la structure 2D déterminée chez la souris et chez l'homme devait en effet être un argument fort d'une signification biologique de la structure établie.

##### 2. Analyse de la structure 2D des régions des A-repeats murine et humaine par emploi de sondes enzymatiques et chimiques

2.1 Principes de l'emploi de sondes chimiques et enzymatiques pour l'analyse de la structure des ARN en solution

2.1.1 Les sondes enzymatiques et chimiques

Afin d'identifier les régions en simple brin et en double brin des ARN, nous avons utilisé 3 sondes enzymatiques et 2 sondes chimiques. Les sondes enzymatiques étaient :

- La RNase T1, qui coupe spécifiquement les ARN en 5' des résidus guanine et ceci préférentiellement dans les segments non appariés, lorsqu'elle est utilisée en conditions limitantes (**Fig. 37**).
- La RNase T2, qui coupe aussi préférentiellement les régions en simple brin lorsqu'elle est utilisée en conditions limitantes. Cette enzyme agit sans spécificité marquée vis-à-vis des nucléotides encadrant la liaison phosphodiester et libère des fragments avec des extrémités 3' phosphate et 5'-OH comme la RNase T1 (**Fig. 37**).
- La RNase V1 qui permet d'identifier les résidus présents dans les régions en double brin ou les régions structurées par empilement de bases, ceci avec peu de spécificité vis-à-vis de

l'identité des résidus. Cette enzyme libère des fragments avec des extrémités 3'-OH et 5' phosphate (**Fig. 37**).

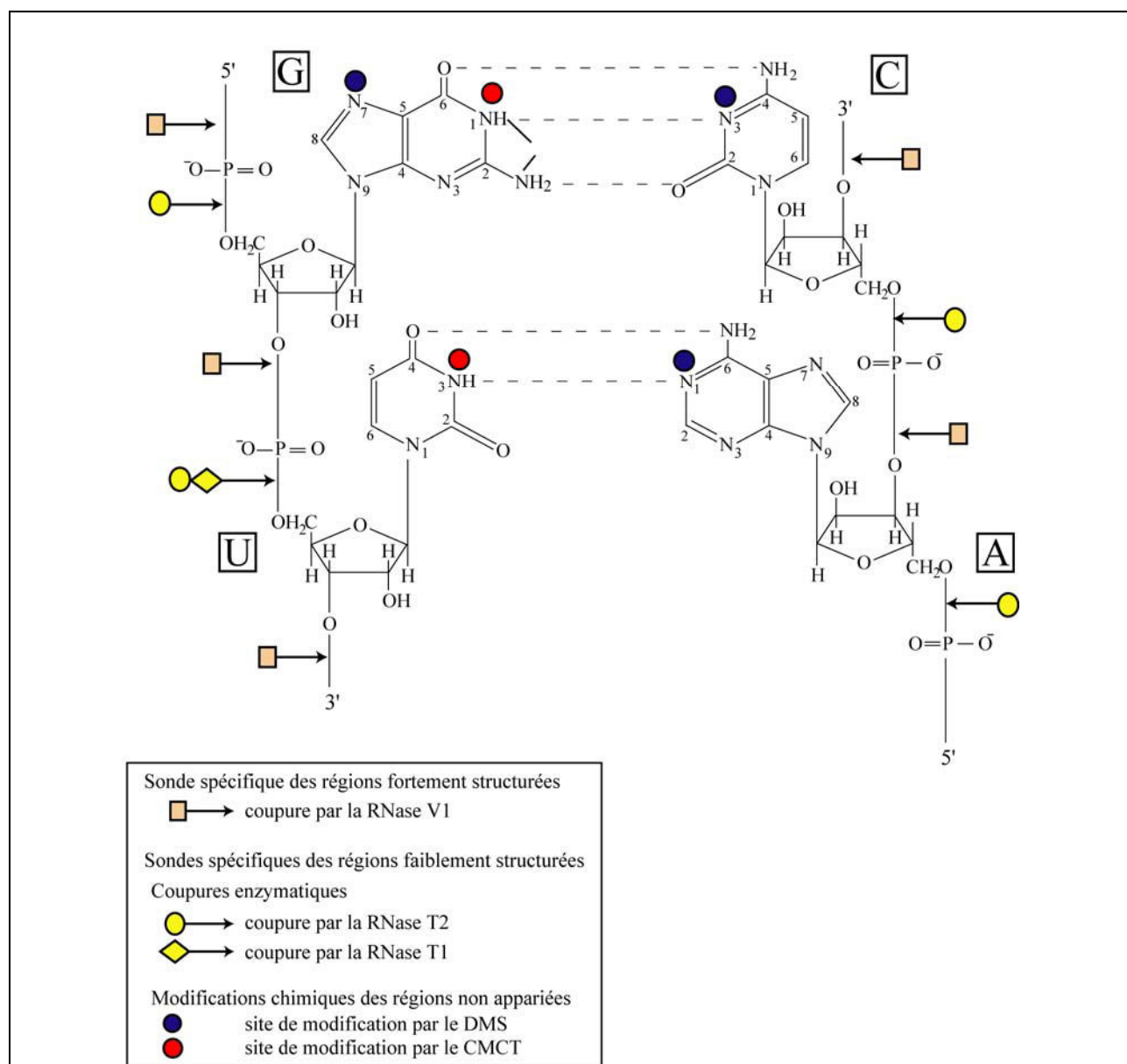
Les sondes chimiques utilisées étaient le diméthylsulfate (DMS) et le N-Cyclohexyl-N'-[2-(N-Méthylmorpholino)-éthyl]Carbodiimid-4-Toluonosulfonate (CMCT).

- Le DMS modifie spécifiquement les résidus A en position N1A, les résidus C en positions N3C et les résidus G en position N7G. Les positions N1A et N3C sont impliquées dans la formation des appariements de bases de type Watson-Crick et de ce fait, seuls les résidus A et C en simple brin sont modifiés et pas les résidus A et C en double brin.

- La 2<sup>ème</sup> sonde était le CMCT qui modifie les positions N3U et à un plus faible degré N1G (**Fig. 37**). Ces positions sont également impliquées dans les appariements Watson-Crick. La modification n'a donc lieu que sur les résidus en simple brin.

Des conditions d'utilisation statistiques de ces sondes avaient été mises au point au laboratoire de façon à ce qu'au maximum une coupure ou une modification ait lieu par molécule. Nous avons dû adapter ces conditions à l'ARN que nous étudions.

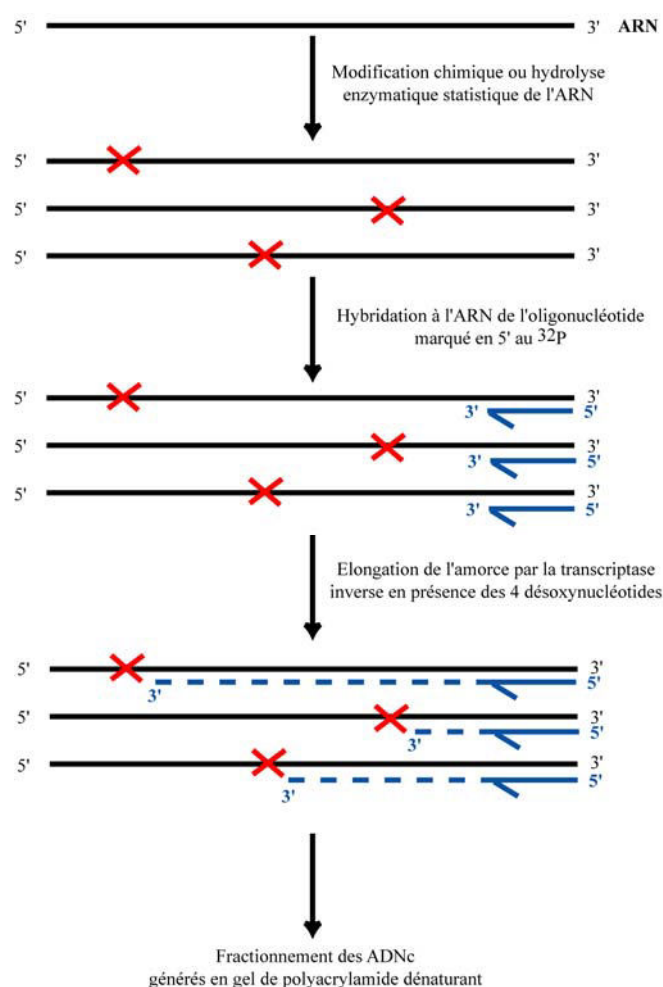
Les conditions sont décrites dans le chapitre Matériel et Méthodes.



**Figure 37 : Représentation schématique des positions modifiées au sein des ARN par l'action des sondes chimiques et des liaisons phosphodiester coupées par l'action des sondes enzymatiques**

### 2.1.2 Identification des positions de clivage et modification

Les positions de coupures par les enzymes et les modifications par le DMS et CMCT ont été identifiées par extension d'amorce par la transcriptase inverse. En effet, l'existence d'un clivage de l'ARN conduit à la formation de produits tronqués lors des expériences d'extension d'amorce (**Fig. 38**). Dans le cas de la RNase V1, la présence d'un phosphate en 5' sur le produit de coupure permet la reconnaissance du dernier résidu par la transcriptase inverse. L'absence de ce phosphate dans le cas des produits de clivage par les RNases T1 et T2 fait que l'arrêt de l'extension a lieu au niveau du nucléotide précédent. Par ailleurs, les méthylations en position N1A et N3C par le DMS et N3U et N1G par le CMCT bloquent la progression de la transcriptase inverse lors des expériences d'extension d'amorce (ce qui n'est pas le cas pour la méthylation en position N7G).

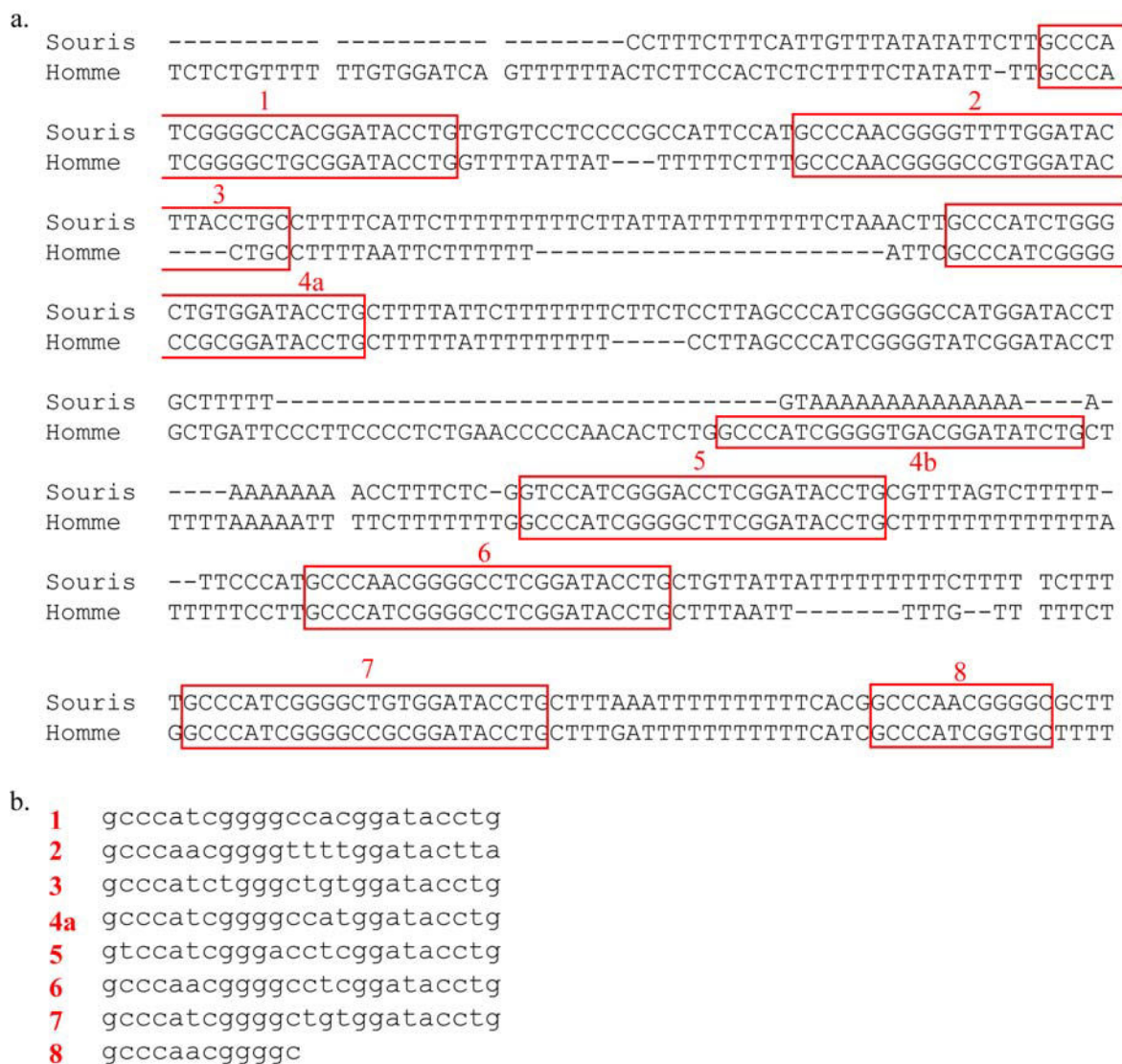


**Figure 38 : Principe de cartographie chimique et enzymatique des ARN et de l'analyse par extension d'amorce**

Identification des coupures et des modifications de l'ARN par élongation d'amorce par la transcriptase inverse. Les croix de couleur rouge représentent les modifications ou les coupures statistiques générées dans l'ARN. L'extrémité 5' des amorces est marquée radioactivement. Les produits d'élongation sont schématiquement représentés par des lignes discontinues.

### 2.1.3 Le choix des oligonucléotides amorce

Les analyses par extension d'amorce nécessitent l'emploi d'oligonucléotides s'hybridant de façon très spécifique à la molécule d'ARN. Dans notre cas, le choix des amorces a été difficile du fait du haut niveau de répétition des séquences (**Fig. 39**). Pour que l'hybridation des oligonucléotides soit spécifique, ils ne doivent pas cibler des séquences présentes plusieurs fois dans l'ARN. De plus, la séquence analysée par extension d'amorce n'excède pas 90 nts. Nous avons donc dû définir un nombre important d'oligonucléotides amorces de façon à analyser l'ensemble de la séquence de la région des A-repeats qui comporte 473 nt chez la souris et 466 chez l'homme. L'efficacité d'extension à partir d'un grand nombre d'oligonucléotides a été testée afin de sélectionner les plus performants pour réaliser les analyses et ceci, aussi bien pour la région des A-repeats murine (10), que pour celle de l'homme (11). Nous avons aussi dû dans certains cas adapter les conditions d'élongation de l'amorce. Ainsi, 6 oligonucléotides ont été retenus dans le but de pouvoir couvrir la totalité de la région des A-repeats murine (3757, 3758, 3760, 3866, 3971, 3973) et 5 pour la région humaine (4242, 4622, 4563, 4564, 4565) (**Fig. 40**).



**Figure 39 : Séquences des régions des A-repeats des ARN Xist murin et humain**

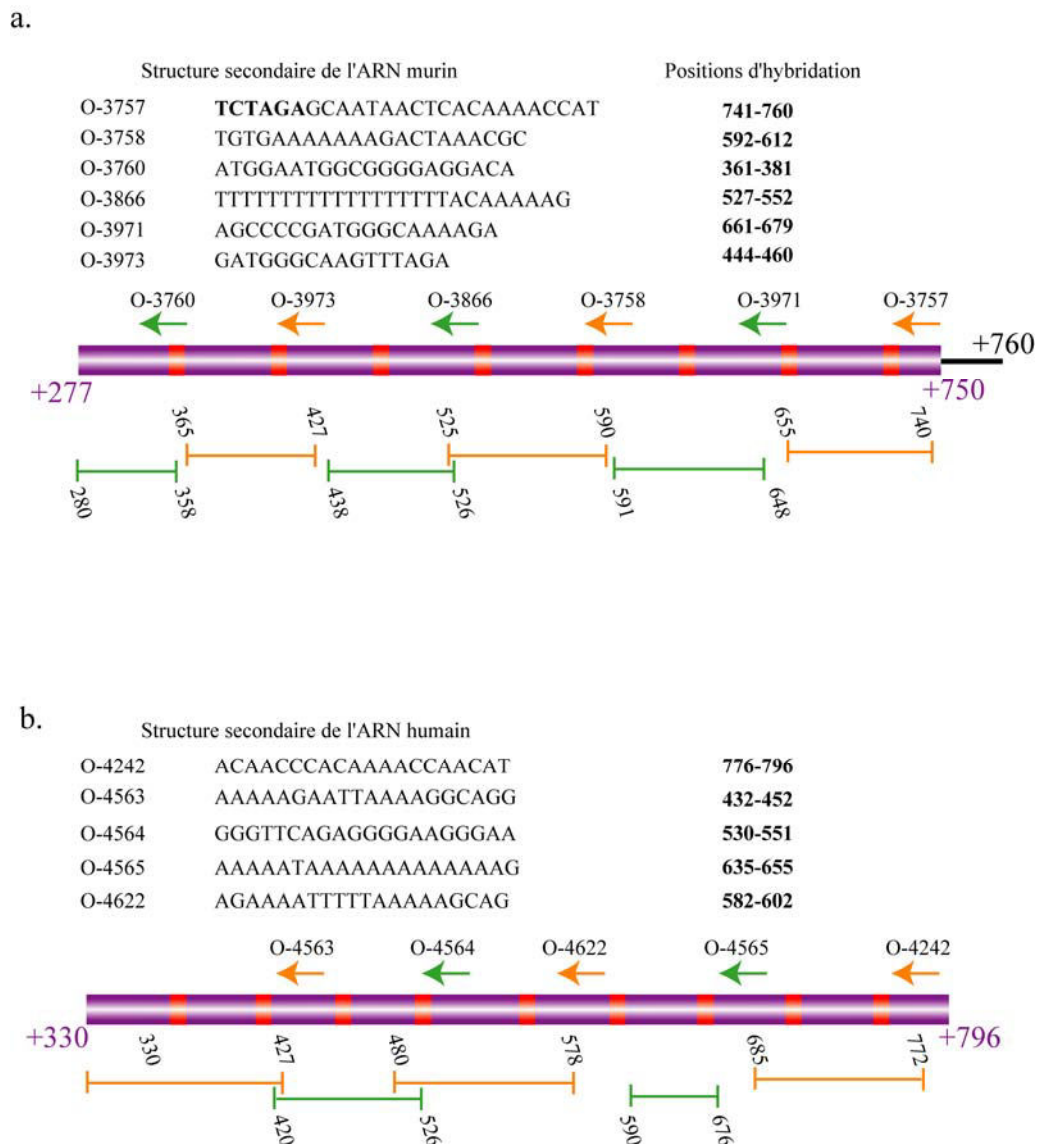
a. Alignement des séquences de la région des A-repeats de souris et d'homme

Souris : no. gil37704378|ref|NR\_001463.2| (Brockdorff *et al.*, 1992)

Homme : no. gil340393|gb|M97168.1| (Brown *et al.*, 1992)

Les répétitions sont encadrées par un rectangle de couleur rouge et leurs numéros sont indiqués de la même couleur.

b. Alignement des éléments répétés de la région des A-repeats murine



**Figure 40 : Schéma représentant pour les régions des A-repeats murine et humaine, la séquence, le numéro, la position d'hybridation de chaque oligonucléotide amorce, le segment analysé est aussi indiqué**

a. Région des A-repeats murine

b. Région des A-repeats humaine

La région des A-repeats est symbolisée par un rectangle de couleur violette et les répétitions par des rectangles de couleur rouge (ce schéma n'est pas à l'échelle).

Le numéro et la séquence de chaque amorce sont donnés. Elles sont représentées par des flèches de couleurs verte et orange au dessus de l'ARN. Les segments analysés avec chacune d'entre elles sont indiqués par des traits de la même couleur. Les positions des extrémités des segments analysés sont indiquées en prenant comme position 1, le nucléotide en 5' de l'ARN Xist.

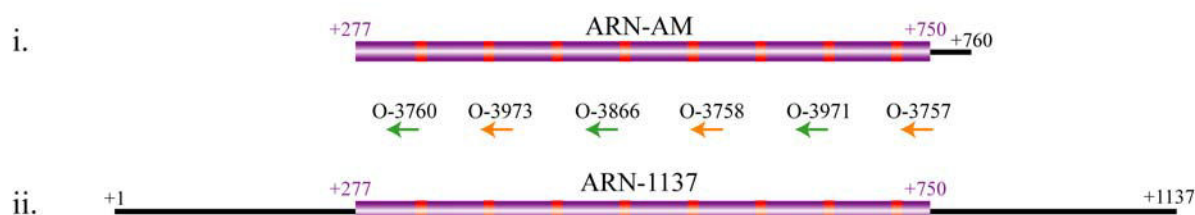
## 2.2 Le choix des ARN utilisés

Lorsque l'on isole un fragment d'ARN de son contexte sans connaître au préalable sa structure 2D, on peut générer ce que l'on appelle des effets de bords. Les segments ayant perdu des partenaires avec lesquels ils sont appariés, peuvent adopter dans le fragment une structure différente de celle qu'ils adoptent dans l'ARN total. De ce fait, il est nécessaire de réaliser les études de structure 2D sur des fragments de l'ARN de longueurs différentes et qui se chevauchent. Seules les structures 2D identiques dans les différents ARN seront retenues pour la construction du modèle.

Nous avons commencé par analyser la structure 2D de la région des A-repeats de l'ARN *Xist* murin. Le premier point dont il fallait s'assurer était de savoir si la région des A-repeats se structure sur elle-même, ou si elle interagit avec d'autres régions la bordant. Une façon de répondre à ce questionnement était d'étudier expérimentalement la structure de la région des A-repeats prise isolément, ainsi que celle d'un très grand fragment la contenant. Bien que l'on ne puisse pas exclure la possibilité que des interactions aient lieu à très longue distance, si la région des A-repeats s'avérait avoir la même structure à l'état isolé qu'au sein du très grand fragment, il était vraisemblable qu'elle se structure sur elle-même. D'où ensuite la possibilité d'affiner l'étude de la structure sur un fragment d'ARN ne contenant que cette région et aussi la possibilité d'étudier la région des A-repeats humaine en utilisant un fragment d'ARN similaire.

Ainsi dans un premier temps, j'ai réalisé une étude de la structure 2D d'un fragment de l'ARN murin de 483 nt (positions +277 à +760) (ARN-AM), qui correspond à la région des A-repeats murine, ainsi que la structure d'un fragment de 1137 nt situé à l'extrémité 5' de ce même ARN. Il correspondait à la région des A-repeats bordée par les 276 nt 5' terminaux de *Xist* et 377 nt en 3' (**Fig. 41**). Pour ces analyses, j'ai uniquement utilisé des sondes enzymatiques. Comme les résultats révélaient très peu de différences d'accessibilité aux enzymes dans la région des A-repeats, selon qu'elle était seule ou au sein d'un grand fragment, nous en avons conclu qu'elle devait se replier sur elle-même. J'ai donc ensuite poursuivi l'analyse sur l'ARN-AM, ceci en employant à la fois des sondes chimiques et enzymatiques. Pour produire les ARN par transcription *in vitro*, j'ai cloné les régions du gène *Xist* murin sous contrôle d'un promoteur T7 dans le plasmide pUC18. Les fragments d'ADN du gène *Xist* murin ont été préparés par amplification (PCR) à partir d'ADN génomique de cellules de souris. La génération des constructions, les conditions de transcription et de purification des transcrits sont données de manière très succincte dans l'Article et en détail dans le chapitre Matériel et Méthodes.





**Figure 41 : Schéma représentant les ARN utilisés pour l'analyse de la structure 2D de la région des A-repeats murine par emploi de sondes enzymatiques et chimiques**

i. Région des A-repeats murine appelée ARN-AM

ii. ARN-1137

La région des A-repeats est symbolisée par un rectangle de couleur violette et les répétitions par des rectangles de couleur rouge (ce schéma n'est pas à l'échelle). Les oligonucléotides amorces utilisés pour les expériences d'extension d'amorce sont indiqués.

Les positions des nucléotides au sein de l'ARN Xist situés aux extrémités de chacun des fragments analysés sont indiquées.

Les ARN ainsi préparés ont été soumis aux hydrolyses et modifications chimiques ménagées dans le tampon classiquement utilisé pour les études d'interaction ARN-protéines et les études fonctionnelles de composants du noyau.

Comme la stabilité de la structure 2D de l'ARN augmente en général en présence de  $Mg^{2+}$  et que le tampon utilisé ne contenait du  $MgCl_2$  qu'à une concentration de 3,125 mM, nous avons testé si le fait d'augmenter la concentration en  $Mg^{2+}$  à 10 mM et 20 mM n'entraînait pas une différence d'accessibilité aux enzymes et aux agents de modification. Comme cela n'était pas le cas, nous en avons conclu que la structure 2D de la région des A-repeats est stable même à une concentration de 3,125 mM en  $Mg^{2+}$ .

Dans ces expériences, il était important que les molécules d'ARN étudiées aient une conformation homogène et que ce soit celle de plus forte stabilité. Il était donc impératif de réaliser une étape de renaturation consistant à incuber l'échantillon d'ARN 10 min à 65°C, puis à le ramener pendant 10 min à température ambiante avant de réaliser les hydrolyses et modifications chimiques. Les conditions d'hydrolyses et de modifications sont données dans le chapitre Matériel et Méthodes.

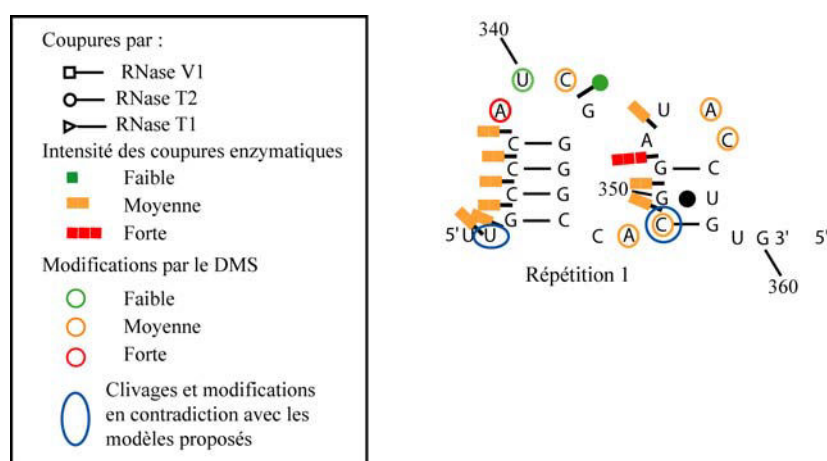
Les positions de clivage et de modifications ont été identifiées par électrophorèse des produits obtenus par extension d'amorce (ADNc), avec en parallèle les produits des réactions de séquençage obtenus avec le même oligonucléotide amorce et une extension d'amorce contrôlée réalisée sur un ARN incubé sans enzyme et sans agent chimique. Ce contrôle permet d'identifier les pauses de la transcriptase inverse. Les analyses de la structure 2D étant très délicates, elles ont toutes été répétées au moins trois fois et ceci sur des lots distincts d'ARN. Nous avons réussi à avoir une grande reproductibilité de nos résultats.

Les électrophorèses, permettant d'illustrer l'accessibilité aux enzymes et modifications de l'ARN-AM sur toute sa longueur, sont présentées dans l'article inséré dans ce chapitre. Par contre, dans cet article, une seule électrophorèse est montrée à titre d'exemple pour l'ARN-1137 ; aussi nous avons inséré en annexe à l'article tous les autoradiogrammes des électrophorèses correspondant à cet ARN avec, à chaque fois en parallèle, l'analyse de la même région réalisée sur l'ARN-AM (**Fig. 1 annexe**). Les électrophorèses qui illustrent l'accessibilité aux enzymes de l'ARN-AM à différentes concentrations salines sont présentées dans la **Figure 2 annexe** jointe en annexe de l'article.

## 2.3 Recherche d'un modèle de structure 2D en accord avec les données obtenues

### 2.3.1 Confrontation des données expérimentales avec le modèle déjà proposé

Comme nous l'avons déjà mentionné, un modèle de structure secondaire avait été proposé pour un motif répété (Wutz *et al.*, 2002). Nos résultats expérimentaux n'étaient pas compatibles avec le modèle proposé comme cela est illustré dans l'article. Nous représenterons seulement ici un exemple de cette discordance (**Fig. 42**). Il s'agit des données obtenues pour le motif situé entre les positions +333 à +360 de la région des A-repeats murine (**Fig. 42**). Nous avons donc recherché un modèle de structure 2D de la région des A-repeats qui soit compatible avec nos données expérimentales et qui soit en plus conforme aux données thermodynamiques sur la structure d'ARN. Pour cela, nous avons utilisé le logiciel MFold.



**Figure 42 : Confrontation de nos données expérimentales sur la structure 2D de la région des A-repeats au modèle proposé par Wutz *et al.*, (2002) pour un des éléments répétés**

Sur la base de prédictions informatiques, Wutz *et al.*, (2002) ont proposé que chaque élément répété de la région des A-repeats se structure individuellement selon 2 petites structures tige-boucles. Les symboles utilisés pour indiquer les coupures et modifications détectées et leurs intensités sont expliqués dans l'encart. A droite est présenté l'exemple de la répétition 1 selon

le modèle proposé par Wutz et al. (2002) avec les modifications et coupures obtenues expérimentalement.

Le nucléotide U334 est hydrolysé par la RNase V1, ce qui démontre qu'il est apparié. Dans le modèle proposé par Wutz et al. (2002), ce nucléotide est non apparié. La modification du nucléotide C349 par les agents chimiques met en évidence qu'il n'est pas apparié. Ce résultat va à nouveau à l'encontre du modèle proposé par Wutz et al. (2002), car dans ce modèle le nucléotide C349 est apparié.

### 2.3.2 Utilisation du logiciel MFold pour la recherche d'un nouveau modèle de structure 2D compatible avec nos données

Le logiciel MFold, fourni par le Genetics Computer Group de l'Université de Madison (USA) est basé sur le programme de recherche de structures secondaires établi par Jaeger *et al.* (Jaeger et al, 1989). Ce logiciel tient compte des estimations de la stabilité thermodynamique des appariements de bases de l'ARN à 0°C dans une solution de NaCl 1M. L'intérêt d'utiliser la version Squiggles est de pouvoir imposer, au cours de la recherche d'une structure secondaire potentielle, des contraintes en état simple ou double brin des différents résidus de l'ARN étudié. Cet état est défini à partir des données expérimentales.

Pour chaque modèle construit par le logiciel, un calcul d'énergie libre est réalisé ( $\Delta G$ ). Plus l'énergie est négative, plus la structure proposée est stable, les structures peuvent aussi être classées selon leur stabilité.

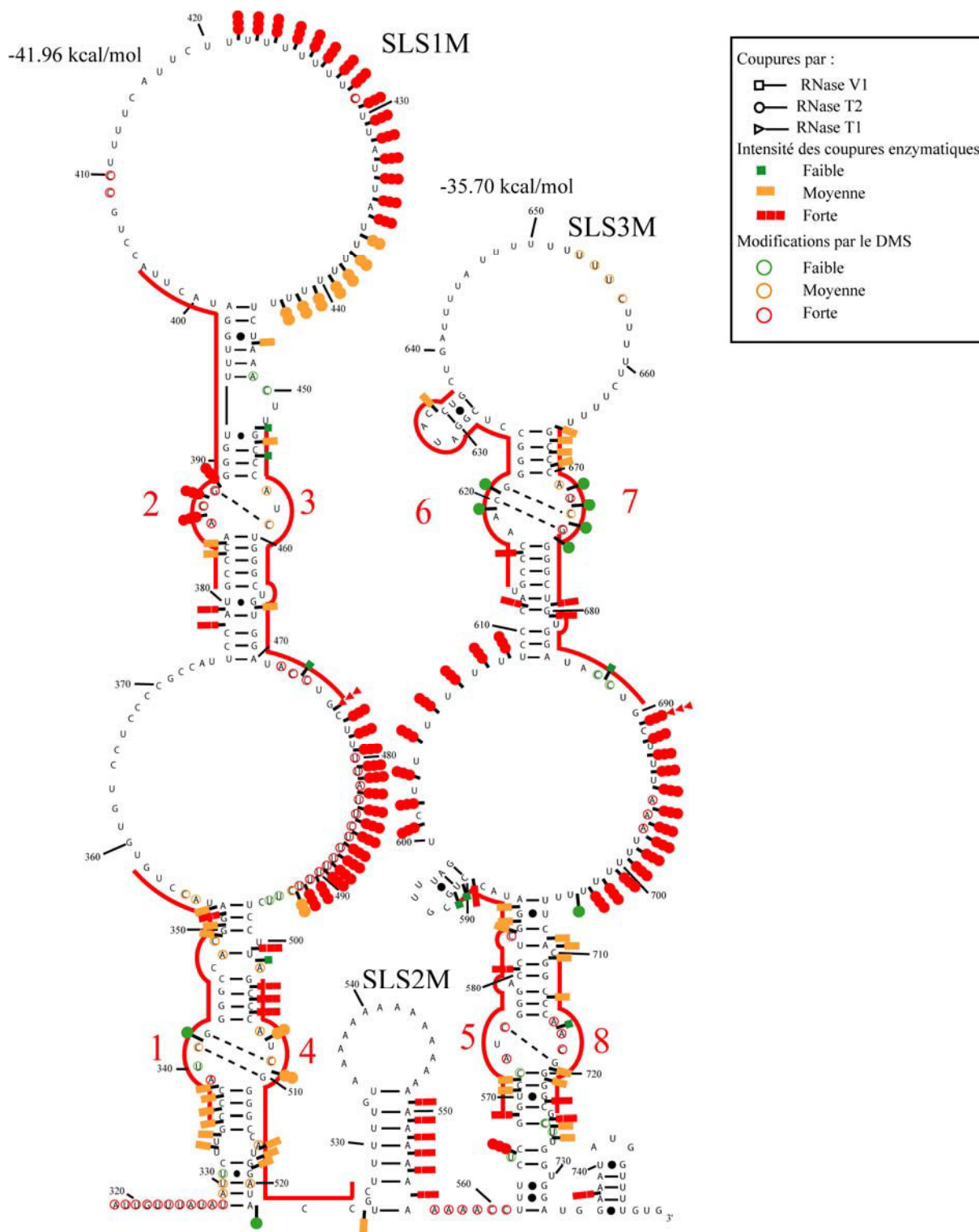
L'application du logiciel MFold à l'ensemble de l'ARN-AM n'a pas conduit à la sélection d'un modèle unique stable et en accord avec toutes les données expérimentales. Néanmoins, dans la plupart des modèles proposés les éléments répétés étaient associés deux à deux, ceci selon différentes combinaisons. L'appariement de deux motifs répétés entre eux conduit à une structuration en meilleur accord avec nos données expérimentales que le repliement des motifs sur eux même, tel que Wutz et al. (2002) l'avaient proposé. Etant donné qu'il existe 7 motifs entiers plus un demi-motif dans l'ARN murin, le nombre de structures 2D potentielles mettant en jeu l'appariement deux à deux de ces 7,5 motifs était grand et nous avons représenté ceux les plus compatibles avec nos données expérimentales dans la **Figure 3** de l'article ci-joint : le modèle 1 dans lequel chaque répétition s'apparie avec la suivante (répétition 1 avec 2...), le modèle 2 dans lequel la répétition 1 s'apparie avec la répétition 8 (la répétition 2 avec la répétition 7, la répétition 3 avec la répétition 6 et la répétition 4 avec la répétition 5) et le modèle 3 dans lequel la région des A-repeats se structure selon 2 grandes structures tige-boucle, chacune formée par l'appariement deux à deux de 4 répétitions : la répétition 1 avec la répétition 4 et la répétition 2 avec la répétition 3 pour la première structure tige-boucle (SLS1M) et la répétition 5 avec la répétition 8 et la répétition 6 avec la répétition 7 pour la deuxième structure tige-boucle (SLS3M).

Une de nos données expérimentales a orienté notre recherche de la structure 2D possible, à savoir que, vu sa grande sensibilité à la RNase V1, l'enchaînement des résidus situé entre les positions 550 et 555 doit être apparié avec une séquence riche en résidus U. Vu l'absence de son clivage ou de sa modification par les enzymes et réactifs spécifiques des simples brins, nous avons supposé que le segment riche en résidus U situé entre les positions 525 et 534, soit le partenaire de l'enchaînement de résidus A, ce qui conduit à la formation d'une structure tige-boucle centrale. En admettant la formation de cette structure tige-boucle, soit les régions l'encadrant pouvaient s'apparier ensemble, soit elles se repliaient sur elles-mêmes (modèle 2 ou 3, respectivement).

J'ai donc utilisé le logiciel MFold et nos données expérimentales pour rechercher si chacune de ces régions pouvait former une structure 2D stable en se repliant sur elle-même. J'ai ainsi observé que la partie 5' de l'ARN-AM (position 318 à 521) peut en effet former la structure 2D stable (SLS1M), dans laquelle la répétition 1 s'apparie avec la répétition 4 et la répétition 2 avec la répétition 3 (**Fig. 43**). D'après le logiciel MFold, cette structure 2D est la plus stable des structures possibles pour cette région, même lorsque les données expérimentales ne sont pas introduites comme contraintes. L'énergie libre calculée est de -41,96 kcal/mol. Une structure similaire peut-être proposée pour la partie 3' terminale. Elle est prédite pour être un peu moins stable. Elle est proposée par le logiciel MFold uniquement si les données expérimentales sont utilisées comme contraintes.

Au travers de cette recherche de structures 2D possibles, nous aboutissons à la proposition de plusieurs modèles (modèle 1, modèle 2, modèle 3) décrits dans l'article. L'un d'entre eux, le modèle 3, semblait être le plus favorable, puisque la moitié 5' était prédite comme correspondant à la structure 2D la plus stable même sans contrainte.

Nous avons alors pensé qu'une comparaison phylogénétique des structures 2D possibles de la région des A-repeats de différentes espèces pourrait apporter des informations sur la validité de chacun des 3 modèles possibles. Nous avons alors entrepris l'étude expérimentale de la région des A-repeats humaine.



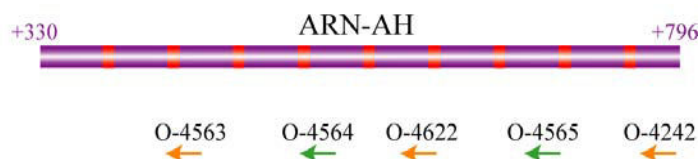
**Figure 43 : Représentation schématique des données expérimentales obtenues pour la région des A-repeats de l'ARN Xist murin (ARN-AM) sur le modèle de structure secondaire 3 qui semblait être le plus vraisemblable**

Les répétitions sont indiquées en rouge et leurs numéros sont indiqués de la même couleur. Les énergies libres des structures tige-boucle calculées avec le logiciel Mfold sont indiquées. Les symboles utilisés pour indiquer les coupures et modifications détectées et leurs intensités sont expliqués à droite du modèle.

## 2.4 Etude phylogénétique de la région des A-repeats

### 2.4.1 Analyse de la structure secondaire de la région des A-repeats humaine

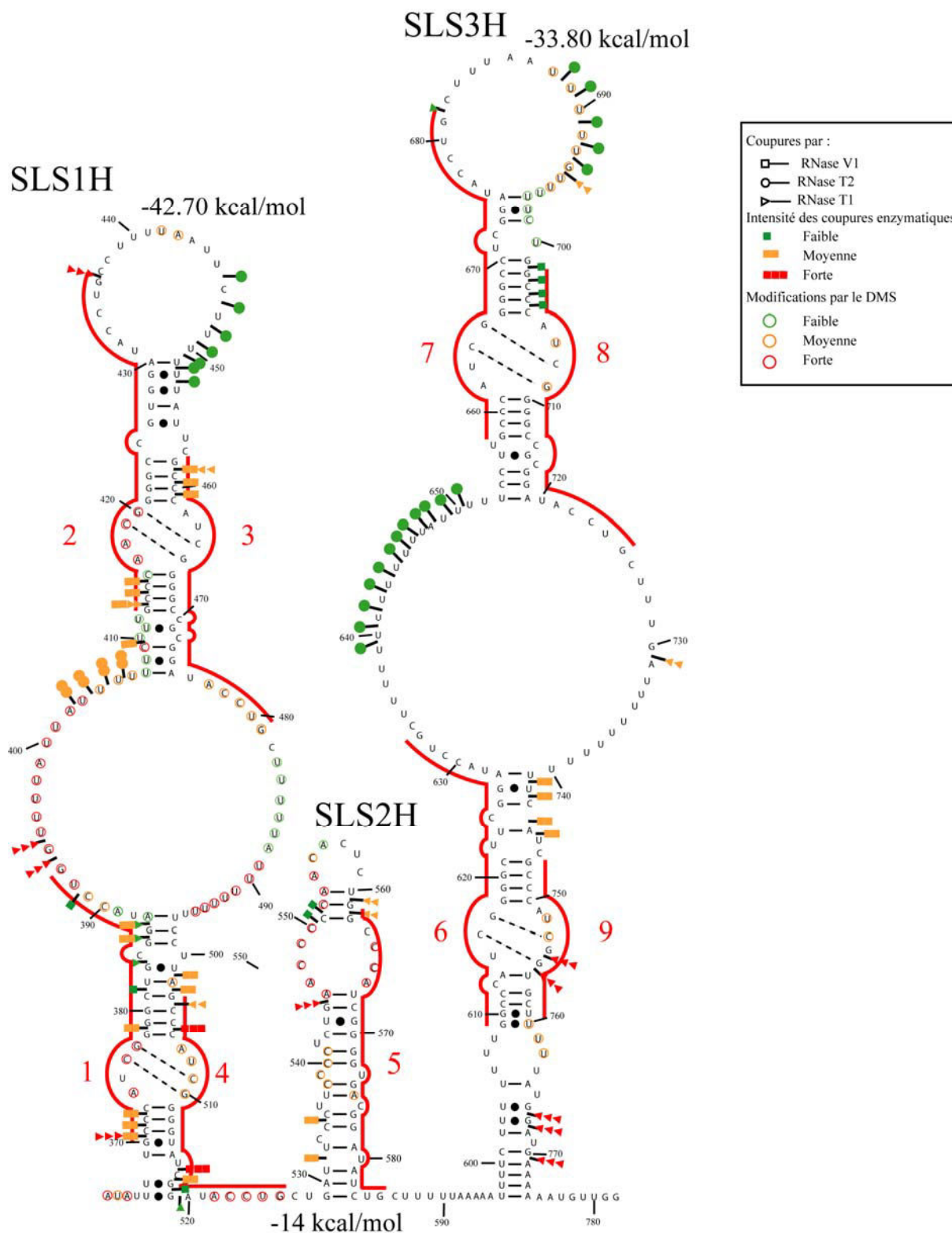
L'analyse de la région des A-repeats de l'ARN Xist humain a été réalisée sur un fragment d'ARN correspondant uniquement à cette région (positions +330 à +796 dans l'ARN Xist humain) (**Fig. 44**). Le fragment d'ADN correspondant a été préparé par amplification (PCR) à partir d'ADNc de cellules HEK293. Nous avons utilisé la même méthode de cartographie des ARN en solution que pour l'étude de l'ARN murin.



**Figure 44 : Schéma représentant l'ARN-AH utilisé pour l'analyse de la structure 2D de la région des A-repeats humaine**

La région des A-repeats est symbolisée par un rectangle de couleur violette et les répétitions par des rectangles de couleur rouge (ce schéma n'est pas à l'échelle). Les oligonucléotides amorceurs utilisés pour les expériences d'extension d'amorce sont indiqués.

Les données expérimentales obtenues pour l'ARN humain étaient compatibles avec les modèles 1, 2 et 3 de structure 2D proposés pour la région des A-repeats murine (**Fig. 3 annexe, Fig. 4 annexe article et Fig. 45**), mais là encore la structure 3 semblait la plus favorable. En effet, la partie 5' de la région des A-repeats humaine peut se structurer selon deux grandes structures tige-boucle irrégulières formées par l'appariement 2 à 2 de 4 motifs répétés. Comme pour la région des A-repeats murine, le logiciel MFold prédit que la formation d'une structure tige-boucle en 5' de l'ARN-AH (position +370 à +530) incluant 4 des 8,5 répétitions est la structure 2D la plus stable possible pour cette région (SLS1H) (**Fig. 45**). Cette structure 2D est prédite comme étant la plus stable avec ou sans introduction des données expérimentales comme contraintes. L'énergie libre calculée est de - 42,70 kcal/mol. Une structure similaire peut aussi être proposée pour la partie 3' terminale. Elle est prédite pour être un peu moins stable (- 33,80 kcal/mol) et est proposée par le logiciel MFold uniquement si les données expérimentales sont utilisées comme contraintes (SLS3H).



**Figure 45 : Représentation schématique des données expérimentales obtenues pour la région des A-repeats de l'ARN Xist humain (ARN-AH) sur le modèle de structure secondaire 3 qui semblait le plus vraisemblable**

Les répétitions sont indiquées en rouge et leurs numéros sont indiqués de la même couleur. Les énergies libres des structures tige-boucle calculées avec le logiciel Mfold sont indiquées. Les symboles utilisés pour indiquer les coupures et modifications détectées et leurs intensités sont expliqués à droite du modèle.



Comme nous l'avons déjà mentionné, la région des A-repeats humaine diffère de son homologue murine par la présence d'un élément répété supplémentaire. Les hydrolyses enzymatiques et les modifications chimiques de cette région suggéraient que la répétition n°5 se structure différemment des autres éléments répétés et forme une structure tige-boucle centrale (SLS2H). En d'autres termes, la structure tige-boucle centrale formée par les enchaînements de résidus U et de résidus A, qui est absente dans l'ARN humain, est remplacée par une structure tige-boucle formée par la répétition supplémentaire. L'ensemble des résultats qui seront détaillés dans l'article, montrait donc que la structure correspondant au modèle 3 pouvait être formée dans les ARN de souris et d'homme.

Afin de vérifier que la moitié 5' de la région des A-repeats humaine était bien capable de se structurer indépendamment du reste de la région, j'ai réalisé des analyses expérimentales de la structure secondaire en solution de cette région produite de manière individuelle. Pour cela, j'ai utilisé un fragment d'ARN allant des positions +330 à +562 (ARN-SLS1H) (**Fig. 46**). Les résultats décrits dans l'article montrent que les profils d'hydrolyses des ARN-AH et ARN-SLS1H sont très similaires, ce qui confirmait que la moitié 5' de la région des A-repeats peut se structurer indépendamment de la moitié 3' et qu'elle a sans doute une structure similaire prise isolément et au sein de la région entière. Ceci renforçait la validité du modèle 3.



**Figure 46 : Schéma représentant l'ARN-SLS1H utilisé pour l'analyse de la structure 2D de la région des A-repeats humaine**

La région des A-repeats est symbolisée par un rectangle de couleur violette et les répétitions par des rectangles de couleur rouge (ce schéma n'est pas à l'échelle). Les oligonucléotides amorces utilisés pour les expériences d'extension d'amorce sont indiqués.

#### 2.4.2 Comparaison de séquences de la région des A-repeats chez les différents mammifères dont le gène *Xist* a été séquencé

Comme nous l'avons mentionné ci-dessus, lorsqu'une structure 2D d'ARN a une importance biologique, la possibilité de la former est souvent conservée au cours de l'évolution. Nous avons donc élargi la comparaison de séquences de la région des A-repeats à l'ensemble des mammifères dont le gène *Xist* avait été séquencé. L'alignement de séquences obtenu avec le logiciel ClustalW pour les différentes régions des A-repeats étudiées est présenté dans la **Figure S5 Article** (programme consultable sur <http://align.genome.jp/>). La **Figure S6 Article** montre la conservation de la possibilité de former les structures tige-boucle



SLS1 et SLS3 du modèle 3 contenant les 4 premières répétitions de la région des A-repeats chez les différents mammifères étudiés. L'ensemble de nos données suggérait donc que la région des A-repeats adopte la structure 3 en solution, il était néanmoins nécessaire d'en apporter la preuve formelle. Pour cela, nous avons mis en œuvre une méthode novatrice d'étude de la structure 2D des ARN, basée sur le transfert d'énergie entre deux fluorophores (FRET).

## 2.5 Analyse physico-chimique par FRET de la structure 2D de la région des A-repeats murine

Les expériences de FRET ont été mises au point et réalisées en collaboration avec M Blaud (post-doctorante au laboratoire) et A Savoye (actuellement doctorante au sein du groupe).

### 2.5.1 Stratégie expérimentale

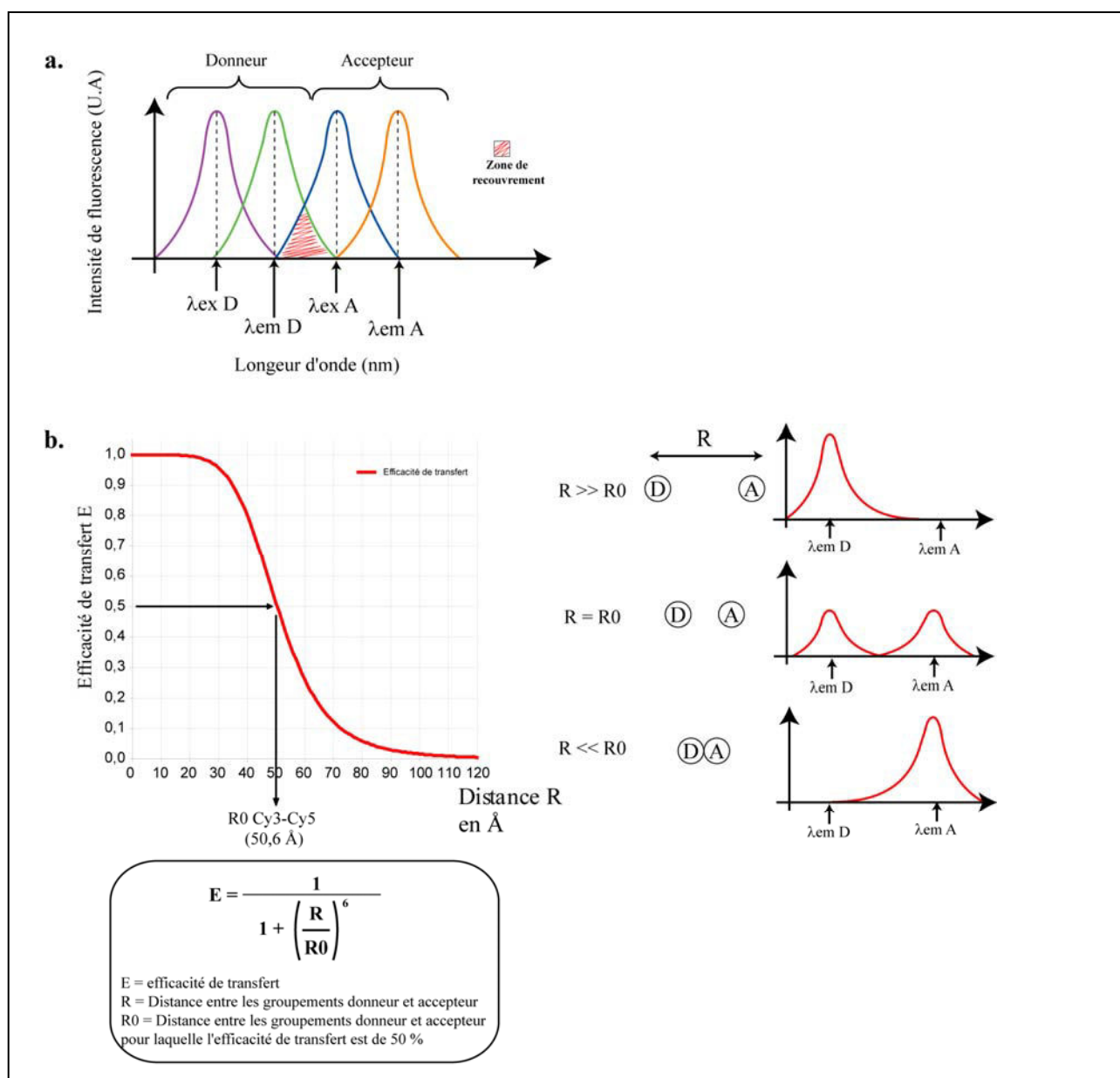
Le FRET (Förster Resonance Energy Transfer) est une méthode de choix pour définir des distances au sein de complexes ou macromolécules biologiques. Le phénomène de FRET résulte d'un transfert d'énergie entre un fluorophore dit « donneur » (D) et un fluorophore dit « accepteur » (A). En pratique, l'excitation du « donneur d'énergie » conduit à une atténuation de l'émission du « donneur d'énergie », qui s'accompagne d'une apparition d'une émission de l'« accepteur d'énergie ». Le choix du couple de fluorophores donneur-accepteur est essentiel, en effet, pour que le transfert d'énergie s'effectue, il faut que les spectres d'émission du « donneur » et d'excitation de l'« accepteur » se recouvrent (**Fig. 47a**). D'après la théorie de Förster, l'intensité du phénomène de FRET dépend de la distance séparant le « donneur » de l'« accepteur », selon la formule suivante :  $E = 1/[1+(R/R_0)^6]$ , où E correspond à l'efficacité de transfert (%), R est la distance qui sépare les deux fluorophores et  $R_0$  est le rayon de Förster. Ce dernier correspond à la distance pour laquelle l'efficacité du transfert d'énergie au sein d'un couple de fluorophores donneur-accepteur est de 50% (**Fig. 47b**). Cette distance, qui dépend de la nature du couple utilisé, est généralement comprise entre 10 et 100 Å (Stryer et Haugland, 1967). Par exemple, le couple Cyanine3-Cyanine5 que nous avons utilisé dans cette étude, a un rayon  $R_0$  de 50,6 Å.

Au-delà de cette distance, l'efficacité du transfert d'énergie entre les 2 fluorophores chute très rapidement. Donc, lorsque R est très supérieur à  $R_0$ , l'efficacité de FRET est faible et l'émission de fluorescence du donneur est peu affectée par la présence de l'accepteur. Lorsque R est égal à  $R_0$ , le transfert d'énergie est de 50%. Lorsque R est très inférieur à  $R_0$ , le transfert d'énergie est maximal (**Figure 47b**).

Les applications déjà faites du FRET en biologie sont nombreuses et concernent aussi bien des études portant sur les interactions intermoléculaires (récepteur/ligand), que sur les

interactions intramoléculaires (repliement des acides nucléiques, structure quaternaire d'une protéine) (Gavory *et al.*, 2006). Jusque là, pour mesurer des distances dans un ARN ou entre ARN, les fluorophores étaient liés de façon covalente aux extrémités de l'ARN. Pour des raisons techniques, cette stratégie n'est applicable qu'à des ARN de petites tailles. Comme la région des A-repeats est grande, nous avons pensé qu'il serait possible de placer les fluorophores aux extrémités d'oligonucléotides complémentaires à cette région. Nous avons donc utilisé différents oligonucléotides complémentaires à la région des A-repeats portant à leurs extrémités un fluorophore donneur ou accepteur. Lors de ces travaux, un article portant sur l'analyse de la structure 2D de l'ARN de la télomérase par le même type d'approche a été publié (Gavory *et al.*, 2006).

Nous allons décrire dans le paragraphe suivant les paramètres que nous avons dû optimiser pour que la méthode soit performante. Nous verrons la stratégie de sélection des fluorophores et des oligonucléotides, la préparation des hybrides ARN-oligonucléotides et les conditions de prise de mesures.



**Figure 47 : Principe du FRET**

a. Spectres normalisés caractéristiques de l'émission et de l'excitation de chaque fluorophore d'un couple donneur-accepteur. Ces spectres mettent en évidence la zone de recouvrement du spectre d'émission du groupement donneur et d'excitation de l'accepteur. Ce recouvrement rend possible le transfert d'énergie entre le donneur et l'accepteur d'énergie.  $\lambda_{ex D}$  : longueur d'onde d'excitation du donneur d'énergie ;  $\lambda_{em D}$  : longueur d'onde d'émission du donneur d'énergie ;  $\lambda_{ex A}$  : longueur d'onde d'excitation de l'accepteur d'énergie ;  $\lambda_{em A}$  : longueur d'onde d'émission de l'accepteur d'énergie.

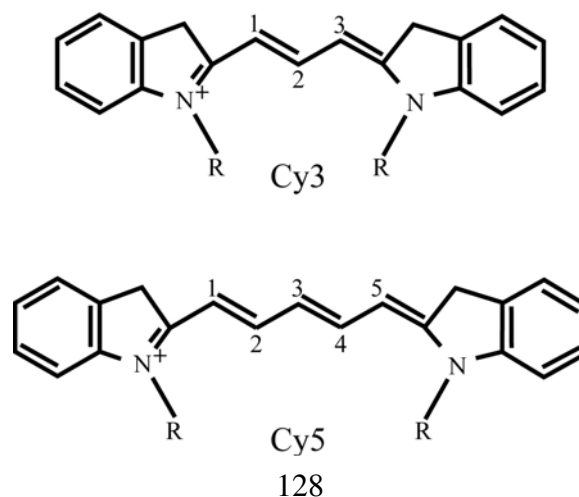
b. Courbe d'efficacité de transfert (E) en fonction de la distance (R) séparant les groupements donneur et accepteur. (E) répond à la formule présentée sous la courbe. A droite sont illustrés différents cas de figures qui peuvent se présenter en fonction de la valeur du paramètre R. Se référer au texte pour les détails.

## 2.5.2 Mise au point du protocole expérimental

- Choix des fluorophores

Nous avons choisi le couple de fluorophores Cyanine3-Cyanine5 (Cy3-Cy5) (**Fig. 48**). Il s'agit de molécules constituées de cycles aromatiques reliés par une chaîne de 3 (Cy3) ou 5 (Cy5) atomes de carbone. Ce couple est actuellement l'un des plus utilisés car le spectre d'émission du donneur (Cy3) recouvre largement celui d'excitation de l'accepteur (Cy5). De plus, les fluorophores Cy3 et Cy5 sont plus stables que les autres fluorophores disponibles, tel que le couple FITC/Rhodamine. Le  $R_0$  caractéristique de ce couple est de 50,6 Å. Les longueurs d'onde caractéristiques d'excitation et d'émission du groupement Cy3 sont respectivement de 548 ( $\lambda_{ex Cy3}$ ) et 564 nm ( $\lambda_{em Cy3}$ ), celles du groupement Cy5 sont respectivement de 646 ( $\lambda_{ex Cy5}$ ) et 662 nm ( $\lambda_{em Cy5}$ ).

En théorie, lors des expériences de FRET, la mesure de la fluorescence doit être réalisée à la longueur d'onde d'excitation du donneur, c'est-à-dire 548 nm dans le cas de Cy3. Toutefois, lorsque le groupement Cy3 est excité à 548 nm, un pic d'excitation résiduel est détecté à cette même longueur d'onde. Ce pic d'excitation résiduel perturbe les mesures du signal d'émission à 564 nm puisque ces 2 pics sont chevauchants. Pour éviter ce phénomène, l'excitation du groupement donneur a été réalisée à une longueur d'onde de 515 nm.



**Figure 48 : Structures chimiques des fluorophores Cyanine3 (Cy3) et Cyanine5 (Cy5) utilisés dans cette étude**

- Choix des oligonucléotides

Nous retrouvons dans cette approche les mêmes problèmes de redondance de séquence que pour les analyses par transcription inverse. Nous avons donc dû définir les oligonucléotides avec soin, car ils devaient répondre à différents critères importants.

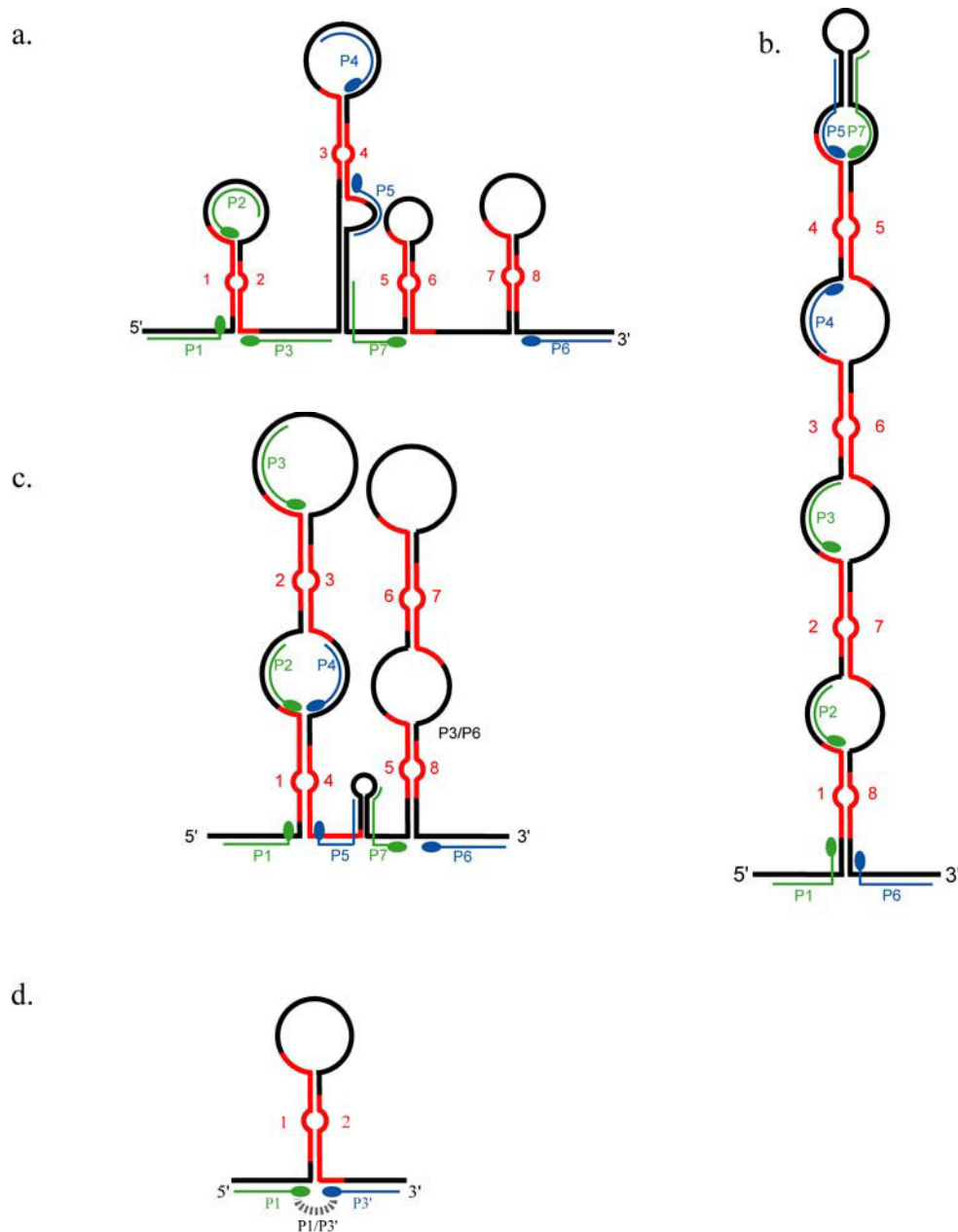
Tout d'abord, pour obtenir une hybridation à un site unique sur l'ARN, leurs séquences ne devaient pas être complémentaires à une région répétée de l'ARN. De plus, il était essentiel que la région complémentaire dans l'ARN soit en simple brin, sous peine d'altérer la structure 2D par appariement avec les oligonucléotides.

Comme nous l'avons mentionné ci-dessus, différents modèles de structure 2D étaient possibles : le modèle 1, dans lequel la région des A-repeats se structure selon 4 structures tige-boucle formées chacune de l'appariement 2 à 2 de 2 motifs répétés successifs (**Fig. 49a**); le modèle 2 correspondant à une grande structure tige-boucle formée par l'appariement 2 à 2 de tous les éléments répétés (**Fig. 49b**) ; et le modèle 3, dans lequel la région des A-repeats se structure en 2 grandes structures tige-boucle formées par l'appariement 2 à 2 de 4 éléments répétés (**Figure 49c**). Nous avons utilisé des couples d'oligonucléotides qui nous permettaient de confirmer ou d'invalider chacun des 3 modèles.

Nous avons défini 7 oligonucléotides (P1 à P7) répondant à ces critères. L'utilisation des couples P1-P5, P2-P4 et P6-P7 devait nous permettre de confirmer le modèle 3. En effet, d'après ce modèle, les oligonucléotides s'hybrident à des régions très proches l'une de l'autre (à la base des structures SLS1M et SLS3M pour les couples P1P5 et P6P7 et dans la boucle interne de la structure SLS1M pour le couple P2P4, **Fig. 49c**). Si la structure adoptée correspondait au modèle 3, nous devions observer un fort transfert d'énergie. Par contre, le transfert d'énergie de ces couples devait être faible, si la région des A-repeats se structurait selon les modèles 1 ou 2 (**Fig. 49c**). Nous avons aussi utilisé le couple P1-P6 qui devait conduire à un fort transfert d'énergie si la région des A-repeats se repliait selon le modèle 2, et à un faible transfert si elle se repliait selon les modèles 1 et 3 (**Fig. 49c**).

Afin de définir quel était le transfert d'énergie attendu pour 2 oligonucléotides couplés à des séquences simple brin bordant une hélice, nous avons utilisé un ARN modèle (ARN-2R). Cet ARN se structure en une hélice bordée par 2 séquences en simple brin complémentaires à la paire d'oligonucléotides utilisée (**Fig. 49d**).

Les différents oligonucléotides couplés aux fluorophores ont été synthétisés chimiquement et purifiés par Eurogentec<sup>®</sup>. Leurs séquences ainsi que leurs régions d'hybridation sont indiquées dans la **Figure 49e**.



e.

Oligonucléotides définis pour les expériences de FRET (5'-3')	Positions d'hybridation
P1 GAATATATAAACAATGAAAGAAAGG-Cy3	319-332
P2 Cy3-GCGGGGAGGACACACAGGTA	355-372
P3 Cy3-AATGAAAAGGCAGGTAAGTA	399-418
P4 GAAGAAAAAAGAATAAAAG-Cy5	473-496
P5 Cy5-ACAAAAAGCAGGTATCCATG	515-534
P6 GCAATAACTCACAAAACCAT-Cy3	741-761
P7 Cy5-GGTTTTTTTTTTTTTTTTT	542-561
P3' Cy5-AATGAAAAGGCAGGTAAGTA	399-418

**Figure 49 : Stratégie utilisée pour évaluer la validité de 3 modèles possibles de la région des A-repeats murine par l'approche de FRET**

a. Modèle 1

b. Modèle 2

c. Modèle 3

d. Représentation schématique de l'ARN-2R utilisé pour définir le taux de transfert d'énergie attendu pour un couple d'oligonucléotides s'hybridant à la base d'une structure tige-boucle.

Les répétitions et leurs numéros sont indiqués en rouge. Le transfert d'énergie est représenté par des lignes pointillées d'épaisseur proportionnelle à l'intensité de FRET attendue dans le cas du modèle 3.

d. Oligonucléotides utilisés pour les expériences de FRET.

Cy3 correspond au fluorophore Cyanine3 représenté en vert.

Cy5 correspond au fluorophore Cyanine5 représenté en bleu.

Le nom, la séquence et la position d'hybridation des oligonucléotides couplés aux fluorophores utilisés pour les expériences de FRET sont donnés.

- Préparation des hybrides ARN/oligonucléotides

Une fois les oligonucléotides sélectionnés, il fallait s'assurer de leur appariement à l'ARN et en particulier, définir les conditions optimales d'hybridation.

Dans tous les tests réalisés, l'ARN était soumis à une étape de renaturation explicitée dans le chapitre Matériel et Méthodes. Une fois l'ARN renaturé, nous avons testé l'effet de la nature du tampon (Hepes, Citrate de sodum, Tris-borate), de la nature des sels monovalents (KCl ou NaCl) et de la concentration en sel divalent  $MgCl_2$  (3,125 à 25 mM), du temps (4 h ou 12 h) et de la température d'incubation (4°C ou 20°C), sur l'efficacité de formation des hybrides ; ceci pour une concentration définie d'ARN (0,38  $\mu M$ ) et différentes concentrations d'oligonucléotides (0,38 à 3,8  $\mu M$ ). Le taux d'hybridation était testé par électrophorèse en gel d'acrylamide non dénaturant à 7%, en tampon Tris-borate à pH 7,5, suivie d'une mesure de la fluorescence avec un appareil Typhoon (9410, Healthcare). L'ensemble de ces essais nous a amené à sélectionner un tampon contenant : 150 mM de chlorure de sodum ; 3,125 mM de  $MgCl_2$  et 15 mM de citrate de sodum (pH 7,0), un temps d'incubation de 4 h à une température de 4°C. L'utilisation d'un excès d'oligonucléotides imposait la purification des hybrides afin d'éliminer les oligonucléotides libres. Plusieurs types de purification ont été testés, électrophorèse suivie d'une électro-élution en conditions non dénaturantes, ou gel filtration sur billes de sépharose (Microspin Illustra S400, Healthcare). Dans tous les cas, ces étapes de purification ont conduit à une forte dissociation des hybrides. Nous avons essayé alors de nous en affranchir en nous plaçant dans des conditions stœchiométriques d'ARN et d'oligonucléotides et ceci, aussi bien pour l'oligonucléotide portant le groupement donneur, que pour l'oligonucléotide portant le groupement accepteur. Ceci a été possible car nous avons montré une forte efficacité de fixation des oligonucléotides dans ces conditions stœchiométriques, toujours par électrophorèse des hybrides formés (**Fig. 50**).

- Paramètres de mesures

Les mesures ont été réalisées avec un spectrofluoromètre flx (SAFAS). Le premier paramètre à évaluer était la quantité de fluorophores nécessaires pour obtenir des mesures fiables et reproductibles. A partir des différents essais réalisés, une concentration en oligonucléotide de 0,38  $\mu M$  s'est avérée satisfaisante. Etant donné que les hybrides étaient

formés en conditions stœchiométriques, cela signifiait que chaque oligonucléotide et l'ARN soient utilisés à cette concentration de 0,38  $\mu\text{M}$ .

Le spectrofluoromètre flux (SAFAS) permet de prendre des mesures à différentes températures. Nous avons testé la reproductibilité du signal obtenu à 4°C et 20°C. La température de 4°C s'est avérée plus satisfaisante, probablement du fait d'une meilleure stabilité des hybrides.

Des tests nous ont permis de conclure que la bande-passante devait être de 3 nm pour l'excitation et pour l'émission, le réglage du photomultiplicateur à 750 V (permet d'amplifier le signal émis pour atteindre des valeurs quantifiables sans pour autant saturer l'appareil). Les spectres d'émission et d'excitation ont été enregistrés entre 500 et 750 nm.

- Protocole adopté pour les mesures

En contrôle, nous avons vérifié l'absence d'émission de fluorescence par le tampon seul excité à 515 nm, ceci afin d'évaluer le bruit de fond inhérent au tampon. Puis, le spectre d'émission de fluorescence de chaque oligonucléotide libre, qu'il soit couplé au Cy3 (PCy3) ou au Cy5 (PCy5), a été enregistré entre 500 et 750 nm pour des longueurs d'onde d'excitation de 515 nm pour Cy3 et 646 nm pour Cy5. Ces enregistrements nous ont permis de vérifier la longueur d'émission ( $\lambda_{em}$ ) de chaque oligonucléotide (**Fig. 50**). Nous avons aussi vérifié que le mélange des oligonucléotides portant un groupement Cy3 et l'autre un groupement Cy5 ne modifiait pas les longueurs d'émission ( $\lambda_{em}$ ) de chaque fluorophore et qu'aucun transfert d'énergie notable n'était observé.

L'étape suivante consistait à former l'hybride [PCy3-ARN]. Pour cela, après renaturation, l'ARN était incubé avec l'oligonucléotide portant le groupement donneur pendant 4 h à 4°C, ceci à une concentration d'oligonucléotide et d'ARN de 0,38  $\mu\text{M}$ . Après incubation, un aliquot de l'échantillon était déposé sur gel natif à 7%, afin de vérifier la formation de l'hybride et de déterminer la quantité d'oligonucléotide libre par rapport à celle de l'oligonucléotide hybridé. Puis, le spectre de fluorescence de l'hybride [PCy3-ARN] excité à 515 nm était enregistré dans les conditions décrites ci-dessus. La valeur de l'intensité de fluorescence de l'hybride [PCy3-ARN] à 564 nm ( $\lambda_{em}$  Cy3, pic d'émission maximale du donneur) nous permettait de déterminer la valeur d'émission maximale du groupement donneur fixé à l'ARN en absence du groupement accepteur (**Fig. 50**).

Puis l'hybride ternaire [PCy3-ARN-PCy5] était formé. Dans ce but, l'oligonucléotide PCy5 était ajouté à l'échantillon [PCy3-ARN] toujours en quantités stœchiométriques (1 : 1 : 1) et l'ensemble était incubé pendant 4 h à 4°C. Un aliquot était déposé sur gel natif à 7% afin de déterminer la proportion d'oligonucléotide PCy5 libre à la suite de l'incubation. Ces mesures ont révélé que pour chaque oligonucléotide PCy5, le taux d'hybrides ternaires [PCy3-ARN-PCy5] formés était de 97%. Le taux d'oligonucléotide libre PCy5 était donc

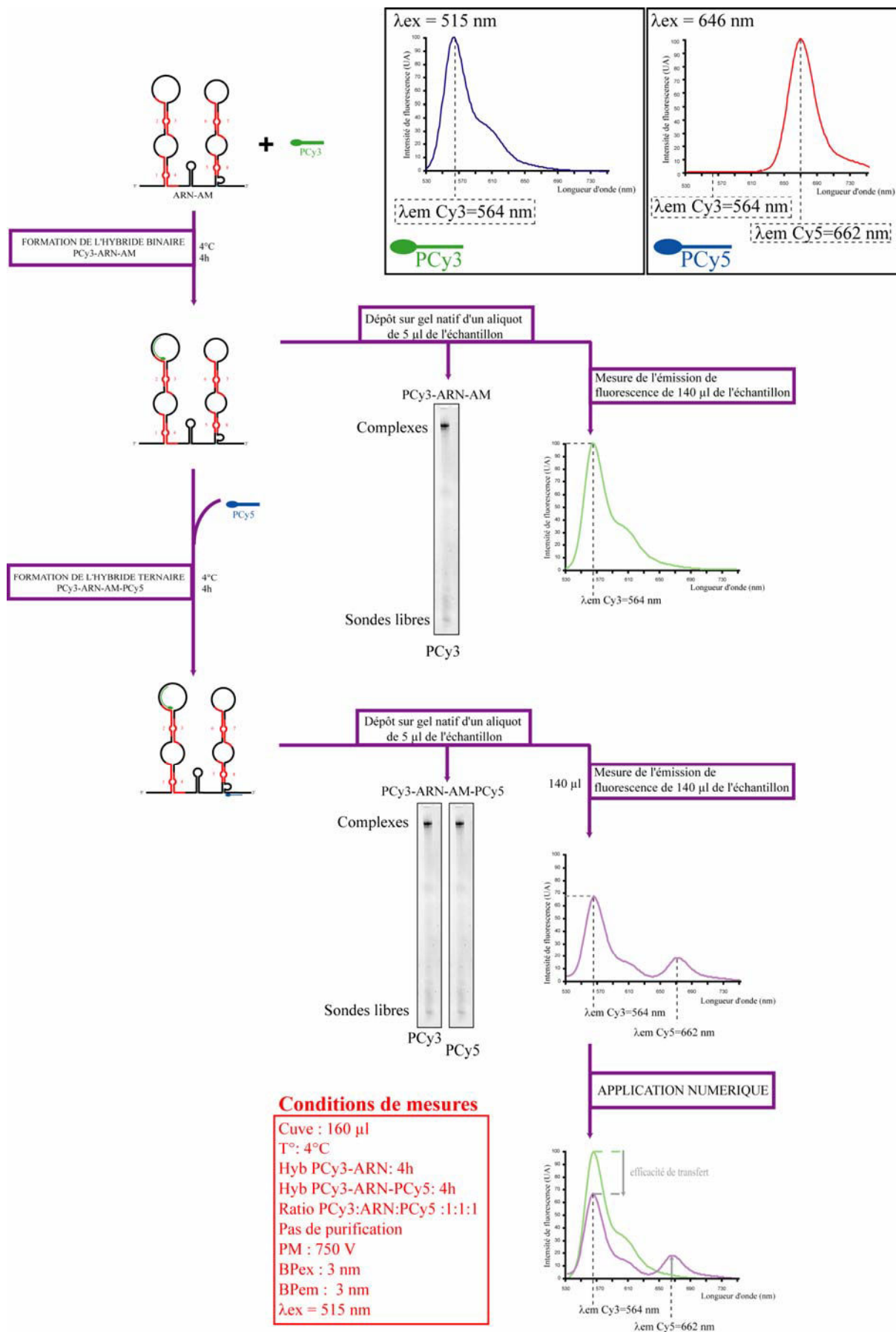
négligeable et n'a pas été pris en compte dans la détermination de l'efficacité de transfert. Ensuite, un spectre de fluorescence de l'échantillon était enregistré entre les longueurs d'onde 500 et 750 nm, ceci en utilisant toujours pour excitation la longueur d'onde de 515 nm. La diminution du pic à 564 nm (correspondant à la longueur d'onde à laquelle l'émission du donneur est maximale) et l'augmentation du pic d'émission à 662 nm (correspondant à la longueur d'onde à laquelle l'émission de l'accepteur est maximale) reflétaient l'existence d'un transfert d'énergie. Ainsi, la comparaison des intensités de fluorescence du groupement donneur à sa longueur d'onde d'émission maximale (564 nm,  $\lambda_{em}$  Cy3) dans l'hybride binaire [PCy3-ARN] et dans l'hybride ternaire [PCy3-ARN-PCy5] nous permettait d'estimer l'efficacité de transfert pour le couple d'oligonucléotides utilisé.

Malgré le fait que nous ayons travaillé à des concentrations bien supérieures à la constante de dissociation ( $K_d$ ) entre les oligonucléotides et l'ARN, l'efficacité d'hybridation à l'ARN n'était pas tout à fait identique pour tous les oligonucléotides PCy3. C'est pourquoi, afin de comparer les résultats obtenus pour chaque couple d'oligonucléotides, il nous a fallu tenir compte de cette différence d'efficacité de fixation. Nous avons donc mesuré les taux relatifs d'oligonucléotides libres/liés par mesure de l'intensité de fluorescence des bandes correspondant à l'hybride et à l'oligonucléotide libre sur le gel natif à 7% (Imagequant). Nous avons ensuite normalisé chaque efficacité de transfert déterminée pour un couple d'oligonucléotides donné par rapport au taux d'oligonucléotides libres/liés de ce même couple.

Pour un échantillon, 10 spectres étaient enregistrés à chaque expérience et chaque expérience était réalisée au moins 3 fois avec des lots d'ARN différents et des aliquots d'oligonucléotides différents.

Le protocole adopté pour les expériences de FRET est présenté dans la **Figure 50**. Les résultats obtenus sont détaillés dans l'article. L'ensemble des spectres enregistrés au cours des expériences de FRET est fourni dans la **Figure 4 annexe**.





**Figure 50 : Protocole utilisé pour les expériences de FRET**

Cette figure représente le protocole utilisé pour les expériences de FRET. Les conditions expérimentales sont indiquées dans l'encart rouge : l'ARN (0,38 µM) a été incubé 4 h à une température de 4°C avec l'oligonucléotide PCy3 (stœchiométrie de 1:1) dans un volume de 160 µl. Après vérification de la formation des complexes et estimation de la proportion

d'oligonucléotide libre et fixé en gel d'acrylamide natif (7%), le spectre d'émission a été enregistré (500 à 750 nm) avec une longueur d'onde d'excitation de 515 nm. Les hybrides [PCy3-ARN] ont ensuite été incubés avec l'oligonucléotide PCy5 (stœchiométrie de 1:1) pendant 4 h à 4°C. Un aliquot a été déposé sur gel natif à 7% afin de déterminer la proportion d'oligonucléotide PCy5 libre. Le spectre d'émission de fluorescence a été mesuré et l'efficacité de transfert d'énergie (E) entre le groupement donneur et accepteur a pu être déterminée.

Même si la structure 2D correspondant au modèle 3 est stable d'après les calculs thermodynamiques, il était aussi possible que sa stabilité soit renforcée par la fixation des protéines nucléaires. Une autre possibilité que nous ne pouvions pas exclure était qu'au contraire, la fixation des protéines puisse favoriser une des autres structures possibles. Pour tester cette hypothèse, j'ai réalisé des études de la structure 2D et l'accessibilité de la région des A-repeats en extrait nucléaire de cellules HeLa.

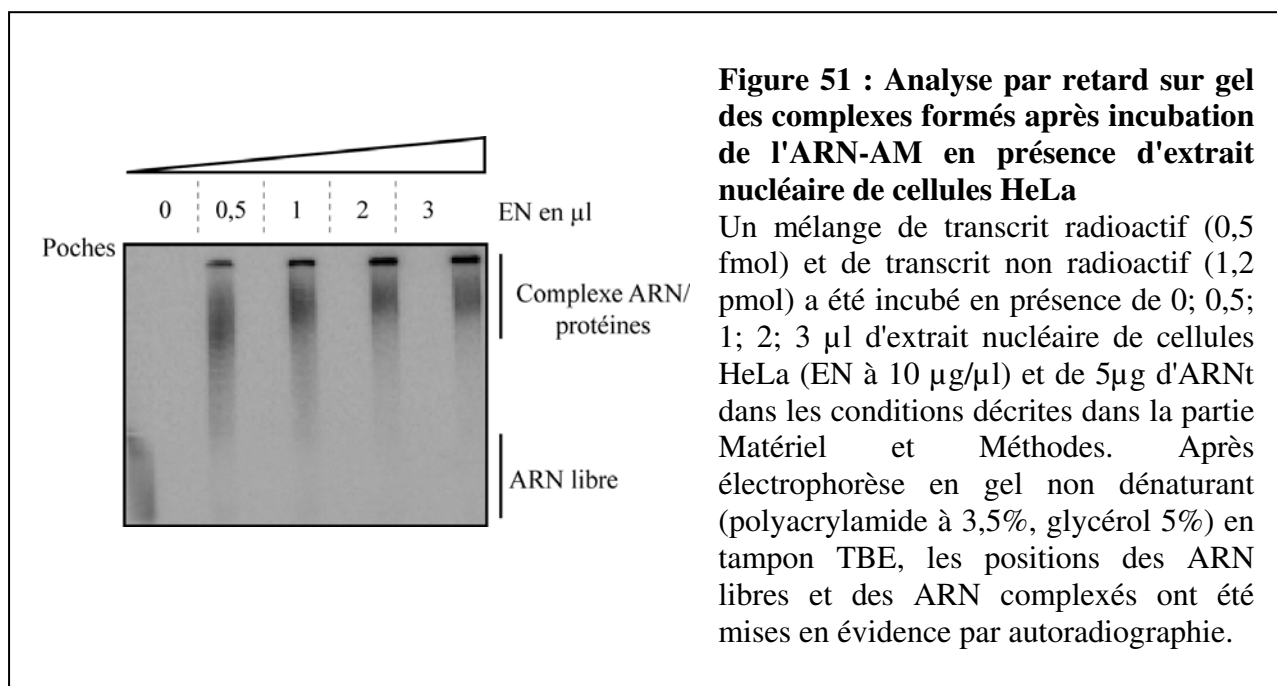
### 3. Expérience d'empreinte en extrait nucléaire

#### 3.1 Objectifs et stratégie expérimentale

Les expériences d'empreinte par des RNases ont été réalisées dans des conditions similaires à celles utilisées pour l'étude de la structure 2D des ARN. Elles devaient nous permettre de savoir si de grands segments de la région des A-repeats étaient protégés par des protéines de l'extrait nucléaire et donc constituaient des sites de fixation pour ces protéines. L'apparition de grands changements d'accessibilité aux enzymes de certains segments pouvait aussi refléter des changements de structure de l'ARN.

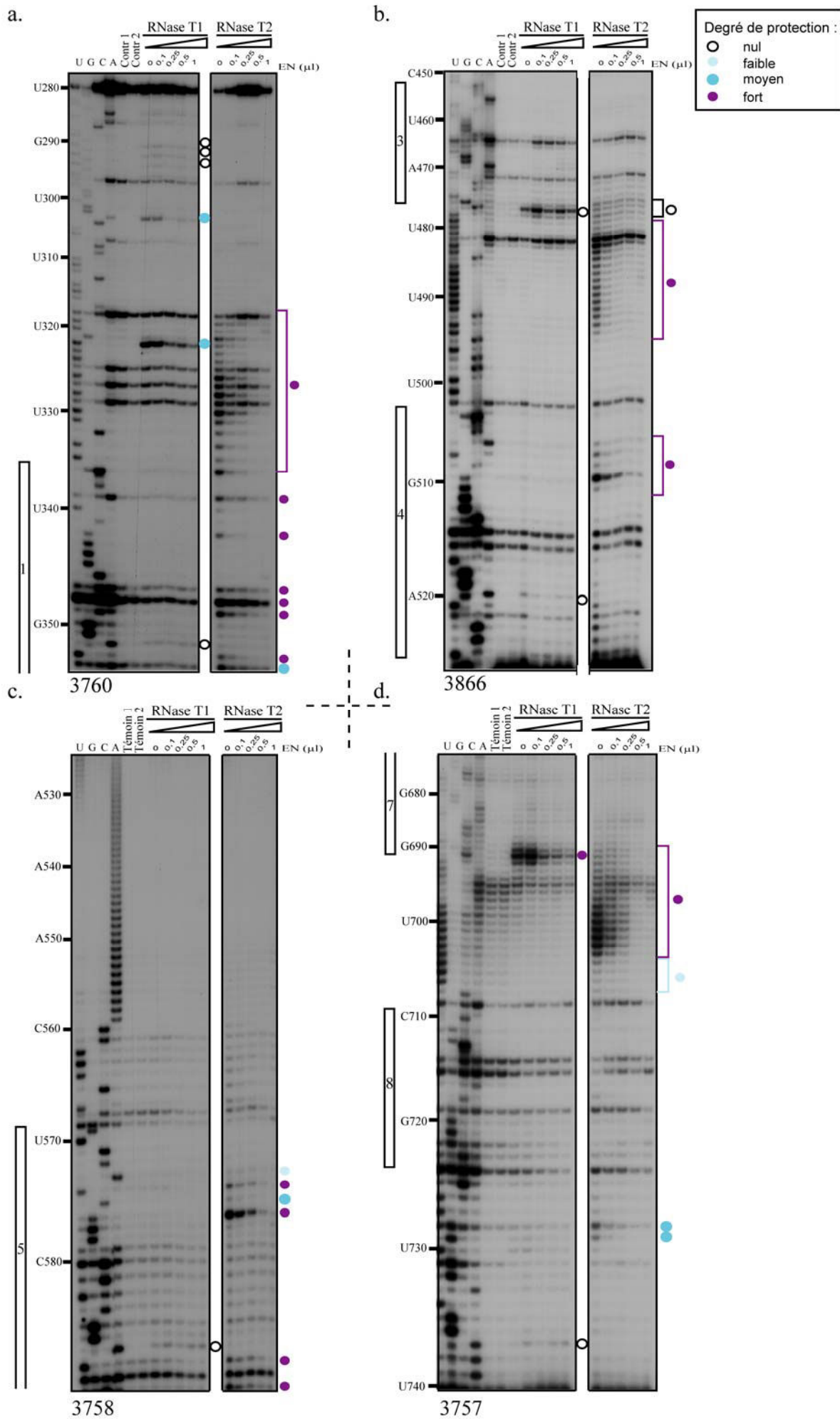
Etant donné que la région des A-repeats est très conservée entre l'homme et la souris et bien que les études aient été réalisées sur l'ARN de souris (ARN-AM), nous avons utilisé des extraits nucléaires de cellules HeLa, qui étaient disponibles au laboratoire, pour réaliser ces expériences. Tout d'abord, il a fallu mettre au point les conditions de formation des complexes. Le point important était de définir les quantités d'extrait nucléaire à utiliser pour que les complexes soient formés, sans pour autant que des agrégats se forment du fait d'un excès de protéines nucléaires. Par des expériences de retard sur gel, nous avons défini les concentrations relatives en ARN et en extrait nucléaire requises pour obtenir une complexation totale de l'ARN. Un mélange d'ARN-AM non marqué à une concentration de 0,06  $\mu\text{M}$  et d'une très faible quantité de ce même ARN uniformément marqué par incorporation d'UTP[ $\alpha$ - $^{32}\text{P}$ ] lors de la transcription (25 pM) a été incubé dans les conditions décrites dans la partie Matériel et Méthodes, ceci en présence de différentes quantités d'extrait nucléaire préparé par la société CilBiotech (0,5 à 3  $\mu\text{l}$  à 10  $\mu\text{g}/\mu\text{l}$ ). Les complexes ARN/protéines formés ont été fractionnés par électrophorèse en gel de polyacrylamide non dénaturant à 3,5%. L'ARN était déjà totalement complexé en présence d'1  $\mu\text{l}$  d'extrait nucléaire (**Fig. 51**). Pour réaliser les expériences d'empreinte, nous avons donc choisi

d'utiliser 4 quantités différentes d'extrait nucléaire inférieures ou égales à ce seuil : 0,1 ; 0,25 ; 0,5 et 1  $\mu$ l.



### 3.2 Résultats obtenus

La formation des complexes entre l'ARN-AM et les composants de l'extrait nucléaire a été réalisée dans des conditions où les extraits nucléaires sont actifs, et notre référence d'activité au laboratoire est leur capacité à catalyser des réactions d'épissage *in vitro* (voir la partie Matériel et Méthodes). Des hydrolyses par les RNases T1 et T2 ont ensuite été réalisées sur les complexes, ceci en employant les conditions d'hydrolyse utilisées pour les études de structure 2D. Les positions de clivage ont été identifiées par extension par la transcriptase inverse des amorces déjà définies pour les analyses de structure 2D (3757, 3758, 3760 et 3866) (**Fig. 40**). Des exemples représentatifs de fractionnements des ADNc produits sont donnés dans la **Figure 52a-d** et les résultats obtenus sont représentés de manière schématique dans la **Figure 53**. Par souci de clarté, les **Figures 52 (Fig. 5 annexe) et 53 (Fig. 6 annexe)** sont également données en annexe de l'article.

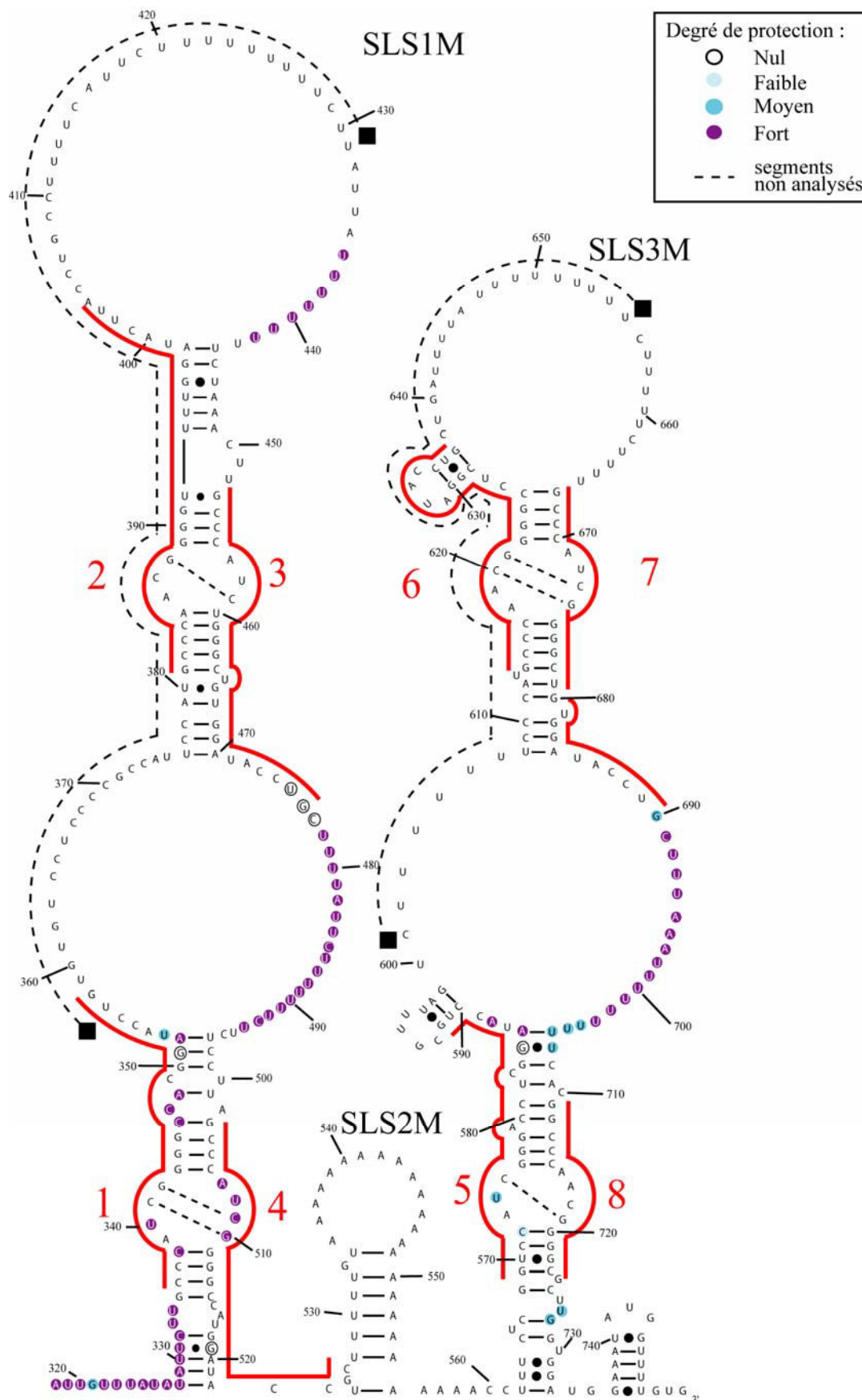


**Figure 52 : Analyse de la protection contre l'action des RNases générée sur l'ARN-AM par fixation des protéines d'un extrait nucléaire de cellules HeLa**

Le transcrit a été soumis à une hydrolyse ménagée par les RNases T1 et T2 (0,5 U/1,2 pmol d'ARN) après incubation en présence ou en absence d'extrait nucléaire de cellules HeLa. Les conditions sont décrites dans la partie Matériel et Méthodes. Les positions des coupures ont été identifiées par extension des oligonucléotides 3760 (a.), 3866 (b.), 3758 (c.) et 3757 (d.) marqués en 5'. Les arrêts d'extension ont été positionnés dans la séquence grâce à des séquençages du transcrit réalisés avec les mêmes oligonucléotides (pistes U, G, C, A). Les ADNc ont été fractionnés par électrophorèse en gel de polyacrylamide dénaturant à 7%. Les positions des nucléotides sont indiquées à gauche de l'autoradiographie. La formation de complexes et les hydrolyses ont été réalisées en présence de  $MgCl_2$  à une concentration de 3,125 mM.

Le contrôle 1 (piste Contr 1) a été réalisé en absence de RNase et en absence d'extrait nucléaire (EN). Le contrôle 2 (Contr 2) a été réalisé en absence de RNase et en présence de 1  $\mu$ l d'extrait nucléaire. Les zones protégées sont indiquées à droite des autoradiographies par des cercles dont l'intensité est proportionnelle au degré de protection.

Les répétitions présentes dans la région des A-repeats sont indiquées par des rectangles à droite des autoradiographies.



**Figure 53 : Représentation des protections contre l'hydrolyse par les RNases T1 et T2 générées par fixation des protéines d'un extrait nucléaire de cellules HeLa sur la région des A-repeats**

D'après les résultats obtenus présentés dans la **Figure 52**, les zones protégées ont été indiquées sur le modèle 3 de structure 2D proposé pour l'ARN-AM par des cercles avec le code couleur indiqué dans l'encart en haut à droite. Le numéro des répétitions est indiqué en rouge.

Seulement 50% de l'ARN-AM ont été analysés au cours de ces expériences mais les segments analysés incluent les répétitions 1, 3, 4, 5, 7 et 8, et des informations intéressantes peuvent être tirées des données obtenues :

- du fait de leur protection en extrait nucléaire, une grande partie des segments en simple brin devait être associée à des protéines de l'extrait. C'est en particulier le cas des très grandes boucles (régions 317 à 334, 436 à 442, 478 à 494, 690 à 703) riches en résidus U. Notons que le niveau de protection varie le long de ces segments en simple brin, avec une forte protection des parties très riches en résidus U et une protection beaucoup plus faible au niveau des extrémités 3'. L'absence de protection des extrémités 3' est un fort argument en faveur du maintien de l'appariement des répétitions 2 à 2. En effet, le repliement des répétitions sur elles-mêmes selon le modèle de Wutz (**Fig. 2 article**) aurait conduit à une protection des extrémités 3' des répétitions du fait de leur implication dans une hélice. Plusieurs protéines contenues dans les extraits nucléaires de cellules HeLa sont connues pour se fixer sur des régions riches en résidus U (PTB, KSRP, hnRNP C1/C2 par exemple). La dissymétrie de protection observée pour les grands segments en simple brin reliant les répétitions 3 et 4 d'une part et les répétitions 7 et 8 d'autre part, régions que nous avons entièrement analysées, suggérait : d'une part que la structure 3 est conservée dans les complexes formés en extraits nucléaires et d'autre part, que les protéines associées à ces segments ont une forte affinité pour les enchaînements de résidus U.

- en ce qui concerne la partie des répétitions qui est impliquée dans la formation des hélices, nous notons, pour celles qui ont été analysées (1, 4, 5 et 8), une protection des résidus proposés pour être en boucle. Comme ces boucles sont présentes dans tous les types de structure 2D proposés, cette protection ne reflète pas un changement de structure 2D mais une protection générée par la fixation de protéines. Il doit donc exister des protéines capables de se lier à ces petites boucles internes.

- par contre, au sein des hélices, les résidus G des répétitions 1, 4 et 5 (G351, G519 et G585) sont accessibles aux RNases, ce qui suggère une absence de fixation de protéine sur les hélices correspondantes.

L'examen des segments protégés suggérait donc l'implication de protéines ayant une forte affinité pour les segments riches en résidus U ; ce qui est le cas pour la protéine PTB car les séquences optimales pour sa fixation sont UCUU(C), UUCUCU ou CUCUCU (Perez et al., 1997). C'est également le cas de la protéine KSRP qui se lie aux séquences AUUUA. Des études de pontage *in vitro* aux UV avaient permis à d'autres auteurs (Brown et Baldry, 1996) de mettre en évidence l'interaction d'une protéine de 40 kDa avec la région 5' de l'ARN Xist. Ces auteurs avaient proposé qu'il s'agisse de la protéine hnRNP C1/C2 qui se lie aux séquences poly(U), AGUA(U)<sub>5</sub>GUGGA. Cette protéine hnRNP C1/C2 était donc aussi un partenaire possible de la région des A-repeats.

De manière générale, les fortes protections de la région des A-repeats que nous avons observées en extrait nucléaire pouvaient correspondre à trois catégories de protéines :

1. Des protéines pouvant renforcer la stabilité de la structure 2D. Cela pouvait être le cas pour PTB dans la mesure où cette protéine possède 4 domaines de fixation à l'ARN fixant chacun des séquences riches en résidus U. Nous pouvions ainsi envisager qu'en se liant par ses domaines de liaison à l'ARN aux brins d'une grande structure tige-boucle SLS1 ou SLS3, PTB puisse stabiliser ces grandes structures tige-boucle.
2. Des protéines plateformes pouvant recruter des protéines fonctionnelles ayant un rôle dans l'inactivation.
3. Des protéines pouvant être impliquées dans l'inactivation, soit en se liant directement à l'ARN, soit en se liant par l'intermédiaire de protéines plateformes.

Comme nous l'avons décrit dans l'introduction, les données sur les partenaires protéiques de la région des A-repeats étaient très disparates et les études des interactions possibles peu approfondies. Les expériences de RIP réalisées récemment suggèrent fortement que deux des protéines du complexe PRC2, Ezh2 et Suz12 interagissent avec la région des A-repeats (Zhao *et al.*, 2008), mais les démonstrations expérimentales ne sont pas irréfutables. Comme nous l'avons mentionné dans l'introduction, les protéines SATB1/B2 pourraient aussi être associées à la région des A-repeats mais aucune preuve expérimentale n'est disponible. De plus, la protéine SAF-A/hnRNP U co-localise avec le chromosome X inactif et la fixation de cette protéine sur le Xi dépend de son site de fixation aux ARN. D'après ces données, SAF-A/hnRNP-U pourrait être également un partenaire de la région des A-repeats. Il faut souligner que lorsque nous avons abordé la thématique au laboratoire, aucune preuve directe de la fixation de protéine sur la région des A-repeats n'avait été obtenue. Pour mieux comprendre le mécanisme d'action de la région des A-repeats, il était crucial d'identifier ses partenaires.

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

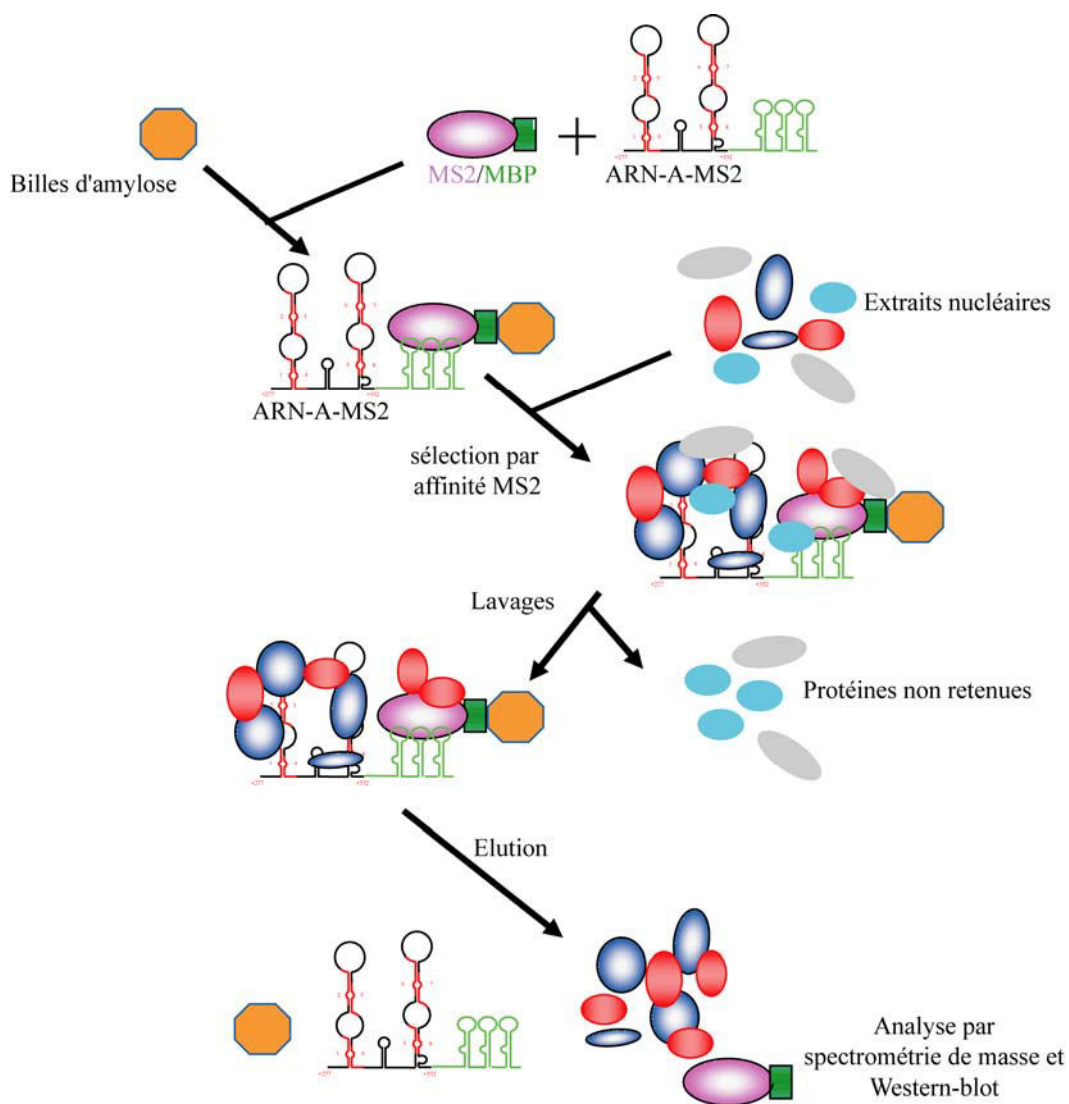
Dans ce but, nous avons utilisé une méthode de chromatographie d'affinité dite « MS2/MBP », mise au point par l'équipe de R Lührmann pour purifier des complexes pré-spliceosomaux. Elle avait été adaptée au laboratoire par V Marchand, C Aigueperse et I Motorin dans l'objectif de purifier les complexes RNP formés avec un segment d'ARN placé dans un extrait nucléaire. Dans un premier temps, nous avons testé si cette méthode pouvait être appliquée à la recherche de partenaires de la région des A-repeats en utilisant un extrait nucléaire de cellules HeLa. Nous avons ainsi pu mettre au point les conditions de mise en œuvre de cette stratégie afin ensuite d'appliquer l'approche à l'étude des complexes formés dans des extraits de cellules ES de souris. Les expériences sur les extraits de cellules HeLa



nous ont permis de définir les quantités relatives d'ARN, de protéine de fusion MS2/MBP et d'extrait nucléaire à utiliser dans le cas de la région des A-repeats. Cet ARN était en effet plus grand que celui que V Marchand et I Motorin avaient utilisé. Nous commencerons par présenter le principe de la méthode utilisée.

### **1. Principe de la méthode de chromatographie d'affinité basée sur le système MS2/MBP**

Cette méthode repose sur l'insertion de façon co-transcriptionnelle de trois segments de l'ARN du bactériophage MS2 d'*E. coli* à l'extrémité 3' de l'ARN d'intérêt. Ces motifs se lient avec une forte affinité ( $K_D = 38$  pM) à la protéine de l'enveloppe de ce bactériophage. Cette protéine est produite sous forme de fusion avec la protéine MBP (Maltose Binding Protein), ce qui permet l'immobilisation de l'ARN « hybride » sur des billes d'amylose auxquelles la protéine de fusion MS2/MBP est capable de se fixer (**Fig. 54**). L'ARN ainsi immobilisé peut alors être incubé dans un échantillon d'extrait nucléaire afin de former des complexes ribonucléoprotéiques (RNP). Après plusieurs étapes de lavage destinées à éliminer les protéines liées de manière non spécifique, les complexes sont élués en conditions natives par ajout de maltose. Les protéines des complexes sont purifiées par électrophorèse en gel SDS-PAGE. Après coloration au bleu de Coomassie colloïdal, les bandes sont découpées et soumises à des analyses par spectrométrie de masse et par Western-blot, ce qui permet d'identifier et de quantifier les partenaires protéiques de l'ARN.



**Figure 54 : Principe de la purification des complexes ribonucléoprotéiques par chromatographie d'affinité basée sur l'emploi de la protéine de l'enveloppe du phage MS2 et de la protéine MBP**

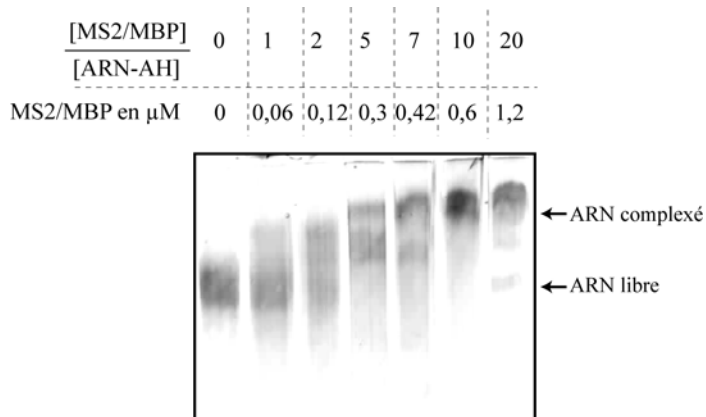
L'ARN d'intérêt fusionné co-transcriptionnellement à 3 motifs MS2 est incubé en présence de la protéine de fusion MS2/MBP. Ce complexe est ensuite mis en contact avec des billes d'amylose. Le complexe ternaire ainsi formé est ensuite incubé en présence d'extraits nucléaires. Après plusieurs lavages successifs des complexes formés, l'ajout de maltose permet leur élution. Environ 50% des protéines contenues dans ces complexes sont fractionnées par électrophorèse en gel SDS-PAGE. Après coloration au bleu colloïdal, les bandes du gel sont découpées et analysées par spectrométrie de masse. Le reste de l'éluat est analysé par Western-blot.

## 2. Mise au point des conditions de purification des RNP basée sur l'emploi d'extrait nucléaire de cellules HeLa

Dans un premier temps, nous avons dû :

- 1.) vérifier la capacité de fixation de la protéine de fusion MS2/MBP produite sous forme recombinante chez *E. coli* sur les ARN fusionnés au motif MS2 et,

2.) définir les concentrations respectives d'ARN et de protéines à utiliser pour avoir une fixation efficace et spécifique de l'ARN. Pour cela, nous avons réalisé des expériences de retard sur gel avec l'ARN-AH-MS2 et la protéine MS2/MBP, ceci en utilisant le protocole présenté dans le chapitre Matériel et Méthodes. Un retard sur gel partiel était observé lorsque des concentrations égales en ARN et en protéine (0,06  $\mu\text{M}$ ) étaient utilisées et tous les sites de fixation de la protéine MS2/MBP portés par l'ARN étaient saturés pour un rapport molaire [protéine]/[ARN] de 10 (**Fig. 55**). Dans des expériences contrôles réalisées avec l'ARN-AH sans le motif MS2, aucun retard sur gel n'était observé même à un ratio molaire [protéine]/[ARN] de 10 (expériences non montrées).



**Figure 55 : Analyse par retard sur gel des complexes formés avec l'ARN-AH-MS2 en présence de concentrations croissantes de protéines MS2/MBP**

Un mélange de transcrite radioactif (0,5 fmol) et de transcrite non radioactif (1,2 pmol) a été incubé en présence de protéine MS2/MBP à des concentrations croissantes de 0; 0,06; 0,12; 0,3; 0,42; 0,6 et 1,2  $\mu\text{M}$ , ceci dans les conditions décrites dans la partie Matériel et Méthodes. Après électrophorèse en gel non dénaturant (polyacrylamide à 3,5%, glycérol 5%), les positions de l'ARN libre et complexé ont été mises en évidence par autoradiographie. Elles sont indiquées par des flèches.

Nous avons sélectionné un rapport molaire [protéine]/[ARN] de 5/1 pour purifier les complexes RNP. Dans ces conditions, tous les ARN sont complexés mais tous les sites de fixation pour la protéine MS2/MBP ne sont pas occupés puisque les ARN sont présents dans plusieurs bandes. Ce choix du rapport 5/1 se justifiait par le fait qu'il est important d'utiliser le moins de protéine MS2/MBP possible, car elle se retrouve dans les fractions d'élution et donc dans le gel SDS-PAGE, et lorsqu'elle est présente en grande quantité elle masque les protéines des complexes ayant des mobilités électrophorétiques voisines.

Nous avons d'abord testé s'il y avait bien formation de RNP entre l'ARN-AH-MS2 humain fixé sur les billes d'amylose et les protéines d'un extrait nucléaire de cellules HeLa. Une expérience contrôle a été réalisée dans les mêmes conditions avec l'ARN-MS2. Cent pmoles de chacun de ces ARN ont été fixées sur les billes d'amylose par l'intermédiaire de la

protéine MS2/MBP. L'ensemble a alors été incubé pendant 15 min avec 100 µl d'extrait nucléaire de cellules HeLa, préalablement dialysé contre du tampon D, et ramené à une concentration de 10 mg/ml. Les composants non fixés de l'extrait nucléaire ont été éliminés par centrifugation, puis la matrice a été lavée trois fois avec du tampon D 1X. Les complexes nucléoprotéiques ont alors été élués par ajout de maltose à une concentration de 10 mM.

Vu l'existence de grandes boucles dans la région des A-repeats, on pouvait se demander si l'ARN résisterait à une incubation dans l'extrait nucléaire. Aussi, l'intégrité de chacun des deux ARN a donc été vérifiée à la fin de l'expérience par fractionnement sur gel de polyacrylamide dénaturant 5% et révélation par coloration au BET. Les résultats mettaient en évidence que la quantité d'ARN-MS2 éluee correspondait à celle mise en œuvre au début de l'expérience. Par contre, la quantité d'ARN-AH-MS2 éluee diminuait par rapport à celle mise en œuvre au début de l'expérience. I Motorin, V Marchand et C Aigueperse, lors de l'optimisation du protocole qu'ils avaient faite, considéraient que cette perte de matériel avait lieu lors des étapes de lavage, la stabilité du complexe ARN-AH-MS2/MBP-MS2 étant vraisemblablement plus faible que celle du complexe ARN-MS2/MS2-MBP. Nous avons néanmoins considéré que cette perte de matériel n'était pas gênante pour l'objectif de purification des RNP que nous nous étions fixé.

Pour que l'approche soit pertinente, il fallait que les contenus en protéines des RNP formées avec les ARN-MS2 contrôle et l'ARN hybride soient différents. Le contenu en protéines des fractions d'éluion a donc été analysé par électrophorèse en conditions dénaturantes (SDS-PAGE) suivie d'une coloration du gel au bleu de Coomassie. Les profils de fractionnement des protéines obtenus après éluion des complexes formés avec chacun des 2 ARN étaient bien différents. Un faible nombre de protéines était détecté pour l'ARN-MS2, ce nombre étant beaucoup plus élevé pour la région des A-repeats, ce qui reflétait la spécificité de la méthode. Nous avons donc conclu que nous pouvions l'utiliser pour rechercher les partenaires de la région des A-repeats.

Par ailleurs, étant donné qu'une coloration au bleu de Coomassie permettait de détecter efficacement les protéines éluées, nous avons conclu que les quantités protéiques purifiées devaient être suffisantes pour leur identification par spectrométrie de masse.

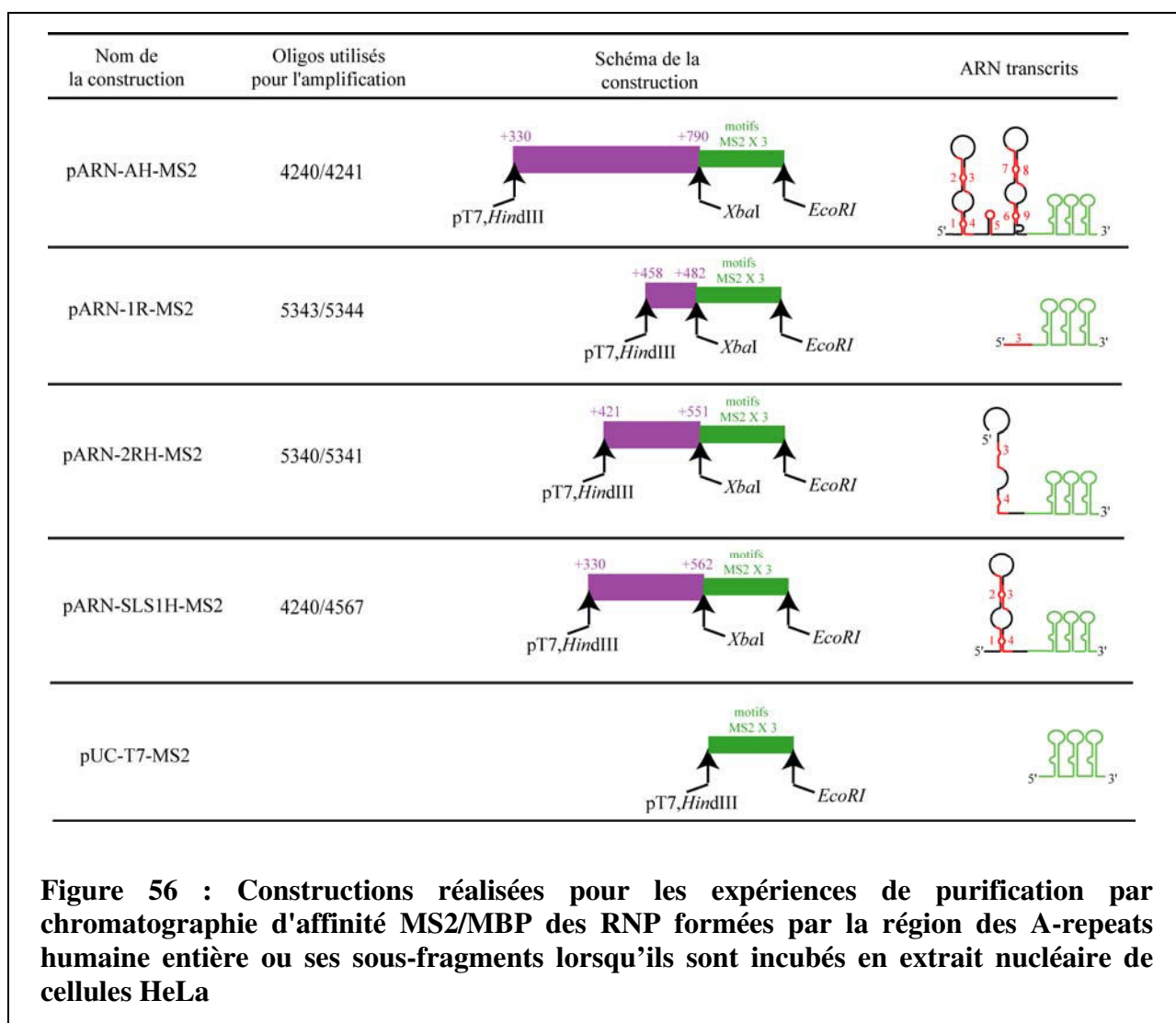
Ayant mis au point les conditions expérimentales, il nous a paru intéressant de réaliser une comparaison des protéines fixées sur la région des A-repeats entière et différents sous-fragments.

### **3. Les différents ARN utilisés pour la purification des RNP formées en extrait nucléaire**

Sur la base de nos données structurales et d'empreinte réalisées sur les RNP formées avec la région des A-repeats murine entière, nous pouvions nous attendre à avoir une composition différente en protéines des complexes selon qu'ils étaient formés avec des ARN

contenant 1, 2 ou 4 répétitions ou avec la région des A-repeats entière. C'est ce que nous avons testé dans un premier temps avec l'extrait nucléaire de cellules HeLa, que nous avons en grande quantité et la région des A-repeats humaine entière ou fragmentée. Pour cela, nous avons produit différents fragments de la région des A-repeats humaine fusionnés en 3' à la séquence MS2. L'ADN correspondant à ces segments, amplifiés par PCR, a été inséré dans le vecteur pUC18 comme décrit dans la partie Matériel et Méthodes. Les couples d'oligonucléotides utilisés pour l'amplification permettaient d'ajouter les sites *HindIII*/promoteur T7 et le site *XbaI* respectivement aux extrémités 5' et 3' des produits d'amplification.

Les constructions pUC-1R-MS2, pUC-2RH-MS2, pUC-SLS1H-MS2 et pUC-AH-MS2 permettant de produire des ARN renfermant 1, 2 et 4 répétitions ou la région des A-repeats entière, ont été réalisées à la fois pour la région des A-repeats humaine et pour celle murine. La **Figure 56** présente les constructions contenant les séquences humaines, les constructions contenant les séquences murines seront explicitées dans l'article.

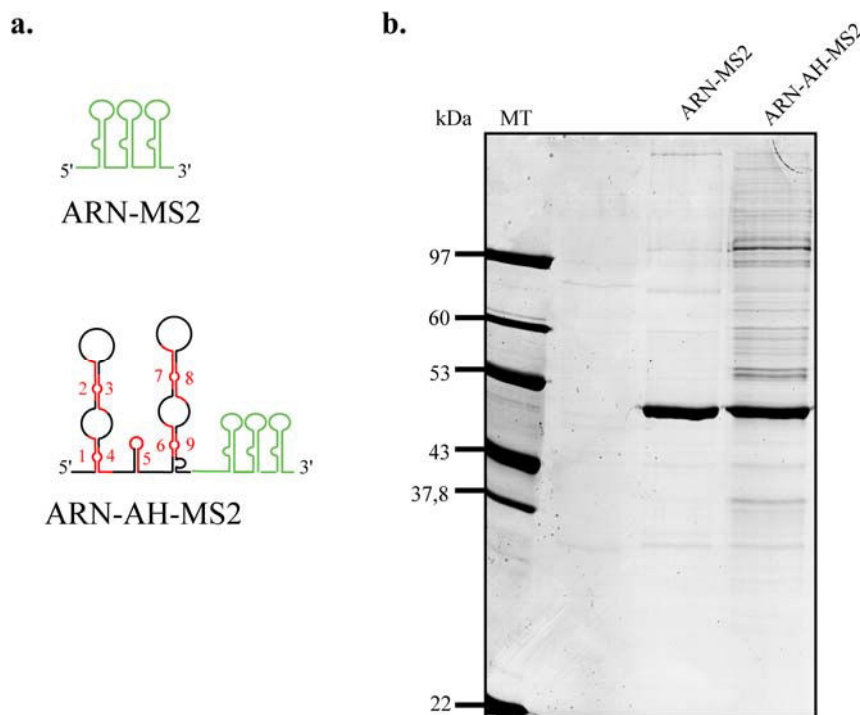


Dans un premier temps, les ARN hybrides correspondant à la région des A-repeats humaine ont été utilisés pour former des complexes avec un extrait nucléaire de cellules HeLa, ceci afin de vérifier que des complexes de compositions distinctes en protéines étaient bien obtenus suivant la taille du fragment de la région des A-repeats présent dans l'ARN. De plus, les données attendues pouvaient nous indiquer quelles protéines nucléaires issues de cellules où l'inactivation du chromosome X est déjà établie sont capables de se fixer sur la région des A-repeats. Comme nous allons le voir, les résultats se sont avérés intéressants.

#### **4. Des RNP ayant des compositions en protéines différentes sont obtenues en utilisant la région des A-repeats entière ou seulement 1, 2 ou 4 répétitions**

Nous avons d'abord réalisé deux expériences successives de purification des RNP formées avec la région des A-repeats humaine entière, selon le protocole décrit plus haut, ceci afin de vérifier la reproductibilité de ce type d'expérience. Dans les 2 cas, un très grand nombre de bandes a été obtenu après séparation des protéines contenues dans les RNP par électrophorèse en gel SDS-PAGE (**Fig. 57**) ; 34 bandes majoritaires ont été découpées et leur contenu en protéines a été analysé par spectrométrie de masse (nanoLC-MS/MS) par l'équipe d'A Van Dorsselaer. Des analyses ont été réalisées en parallèle sur 9 bandes faibles obtenues dans l'expérience témoin réalisée avec un ARN ne contenant que les motifs MS2.

Les analyses de spectrométrie de masse réalisées ne permettent pas d'avoir des données quantitatives sur les protéines détectées. Aussi, un très grand nombre de protéines a été identifié (205). Certaines (43) correspondaient à des contaminations par exemple par la kératine. Lorsque le contenu en protéines des complexes formés d'une part par la région des A-repeats et par l'ARN-MS2 seul d'autre part était comparé, 76 protéines uniquement présentes en présence de la région des A-repeats étaient mises en évidence. Trois de ces protéines étaient attendues : SAF-A/hnRNP-U, PTB et KSRP, ce qui validait la méthode. Néanmoins, il était difficile à ce stade d'avoir une idée précise sur la signification de la présence des 73 autres protéines. Le fait de comparer les protéines fixées à différents fragments de la région des A-repeats devait nous apporter des informations à cet égard.



**Figure 57 : Analyse par électrophorèse en SDS-PAGE des protéines contenues dans les complexes ribonucléoprotéiques formés avec les ARN-MS2 et AH-MS2 purifiés par chromatographie d'affinité reposant sur l'emploi de la protéine de fusion MS2/MBP**

a. Représentation schématique des ARN utilisés lors de cette expérience. L'ARN-MS2 correspond aux motifs MS2, utilisé comme contrôle. L'ARN-AH-MS2 correspond à la région des A-repeats humaine complète.

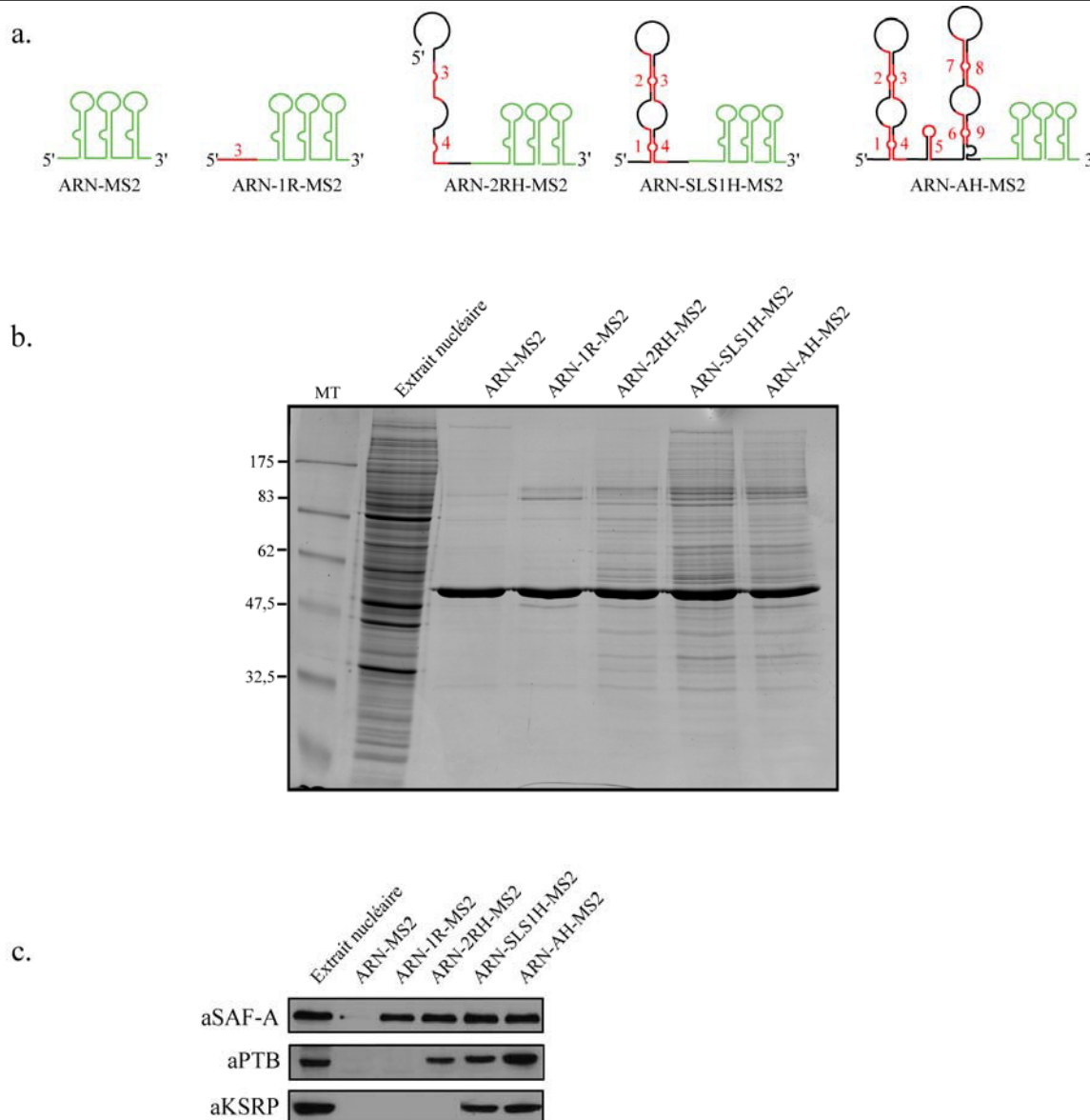
b. Environ 100 pmoles d'ARN ont été incubées en présence de 500 pmoles de protéine MS2/MBP, puis avec 50 µl de billes et enfin avec 100 µl d'extrait nucléaire de cellules HeLa (10 mg/ml). Après lavage, les RNP ont été éluées par ajout de maltose 10 mM et fractionnées en gel SDS-PAGE 10%. Le gel a ensuite été coloré par une solution de bleu de Coomassie colloïdal. Un marqueur de poids moléculaire a été déposé à gauche du gel (MT).

Nous avons donc essayé de savoir si le nombre de répétitions influençait le recrutement des protéines, en particulier, si l'architecture des structures tige-boucle à 4 motifs répétés SLS1H et SLS3H était importante pour la fixation de certaines des protéines. Nous avons donc répété les expériences avec l'ARN contrôle MS2 et les ARN-1R-MS2, ARN-2RH-MS2, ARN-SLS1H-MS2, et ARN-AH-MS2 qui contiennent respectivement 1, 2, 4 répétitions ou l'ensemble de la région des A-repeats (**Fig. 58a**). Les profils de fractionnement des protéines contenues dans les complexes formés avec ces ARN révélèrent une forte augmentation du nombre de protéines fixées lorsqu'on passait de 1 à 2 répétitions, mais aussi lorsqu'on passait de 2 à 4 répétitions (**Fig. 58b**). Par contre, les profils obtenus pour 4 répétitions et la région des A-repeats entière étaient très similaires. Ce constat a été vérifié par une analyse systématique par spectrométrie de masse des protéines contenues dans chaque complexe. Pour cela, chaque piste du gel a été découpée systématiquement tous les 2 mm et le contenu en protéines de chaque bande a été analysé et ceci dans 2 expériences distinctes, réalisées avec chacun des 5 ARN. Les données obtenues seront discutées en comparaison

avec celles obtenues en utilisant des extraits nucléaires de cellules ES dans un chapitre que nous consacrerons à l'examen de l'ensemble des données de protéomique, à la lumière des connaissances disponibles dans la littérature. Ce chapitre est placé derrière le manuscrit de l'article soumis à PLoS Biology. Etant donné que la réalisation de toutes ces analyses par spectrométrie de masse par L Fouillen et S Sanglier étaient très longues (plus de 3 mois), et que cette approche ne permettait pas d'avoir des données quantitatives, nous avons utilisé en parallèle l'approche Western-blot pour tester si les protéines telles que SAF-A, PTB et KSRP que nous avons montré se fixer sur la région des A-repeats entière se fixaient aussi sur les différents fragments. Ainsi, pour chaque expérience de purification des RNP, le volume d'élution des complexes était de 70  $\mu$ l, la moitié était consacrée à l'identification par spectrométrie de masse et le 1/10 était utilisé par expérience de Western-blot.

Les résultats obtenus par Western-blot montraient que SAF-A se liait à tous les ARN testés à l'exception de l'ARN-MS2. PTB se liait sur tous les ARN à l'exception des ARN-MS2 et 1R-MS2. Ces résultats étaient attendus étant donné qu'aucun site de fixation potentiel de PTB n'est présent dans une séquence correspondant à une seule répétition. Les sites potentiels de fixation de PTB sont situés dans les séquences liant les répétitions. De manière intéressante, KSRP ne se liait ni sur l'ARN-1R-MS2, ni sur l'ARN-2RH-MS2 mais se liait aux ARN-SLS1H-MS2 et ARN-AH-MS2, c'est-à-dire lorsque la séquence permettait de former au moins une grande structure tige-boucle à 4 motifs répétés (**Fig. 58c**). Ces observations étaient donc en faveur d'un rôle possible de ces structures pour la fixation de KSRP. Ces données suggéraient donc qu'à la fois la séquence et la structure secondaire de la région des A-repeats sont importantes pour la fixation des partenaires, ce qui validait l'intérêt de nos études structurales.





**Figure 58 : Analyse des protéines contenues dans les complexes ribonucléoprotéiques formés avec différents fragments de la région des A-repeats humaine et un extrait nucléaire de cellules HeLa et purifiés par chromatographie d'affinité MS2/MBP**

a. Représentation schématique des ARN utilisés lors de cette expérience. L'ARN-MS2 correspond aux motifs MS2. Les ARN-1R-MS2, ARN-2RH-MS2, ARN-SLS1H-MS2 et ARN-AH-MS2 correspondent respectivement à 1, 2, 4 et 9 éléments répétés.

b. Fractionnement en gel SDS-PAGE à 10% des protéines des RNP purifiées suivi d'une coloration au bleu de Coomassie colloïdal. Un marqueur de poids moléculaire a été déposé dans la première piste (MT) et 2  $\mu$ g d'extrait nucléaire de cellules HeLa ont été déposés dans la deuxième piste (piste extrait nucléaire).

c. Mise en évidence par l'approche de Western-blot de la présence des protéines SAF-A, PTB et KSRP dans les RNP purifiées.

## 5. Purification des RNP formées avec la région des A-repeats entière et des sous-fragments dans un extrait nucléaire de cellules ES non différenciées (J0) de souris

Nous avons alors décidé d'identifier les RNP formées dans des extraits nucléaires de cellules ES de souris. Comme nous l'avons mentionné dans l'introduction, l'intérêt d'utiliser des extraits nucléaires de cellules ES en cours de différenciation était de pouvoir étudier la formation des RNP formées avec la région des A-repeats dans un extrait de cellules dans lesquelles l'initiation de l'inactivation du chromosome X peut être déclenchée (J0). L'emploi d'extrait de cellules ES non différenciées (J0) de souris femelles est d'autant plus justifié que la sur-expression de *Xist* dans des cellules ES avant différenciation est suffisante pour engager le processus d'inactivation aléatoire (Wutz et Jaenisch, 2000). Cela suggère que tous les facteurs impliqués dans l'inactivation sont présents avant la différenciation et que seule la sur-expression de *Xist* est importante pour déclencher l'inactivation. Dans ce paragraphe nous introduirons les résultats qui sont présentés de manière détaillée dans le manuscrit de l'article qui suit.

Depuis plusieurs années, la préparation d'extrait nucléaire avait été abandonnée au laboratoire puisqu'ils sont achetés à la société Cilbiotech. Avant de préparer des extraits nucléaires de cellules ES de souris femelles qui sont difficiles à cultiver, nous avons testé l'efficacité du protocole (Dignam *et al.*, 1983) de préparation d'extraits nucléaires sur des cellules humaines disponibles au laboratoire, les cellules HEK293. Nous avons en particulier testé plusieurs concentrations salines et une concentration de 1M en KCl a été retenue sur la base des essais réalisés. Afin d'apprécier la qualité des extraits préparés, nous avons réalisé des expériences de Western-blot permettant de détecter certaines protéines nucléaires connues (hnRNP A1 et hnRNP C1/C2 pour lesquelles nous disposons d'anticorps) et nous avons comparé leur abondance dans l'extrait nucléaire que nous avons préparé, par rapport à celle d'un extrait de cellules HeLa fourni par la société Cilbiotech. Pour cela après fractionnement par SDS-PAGE de 1 µg d'extrait nucléaire, des tests ont été réalisés par Western-blot.

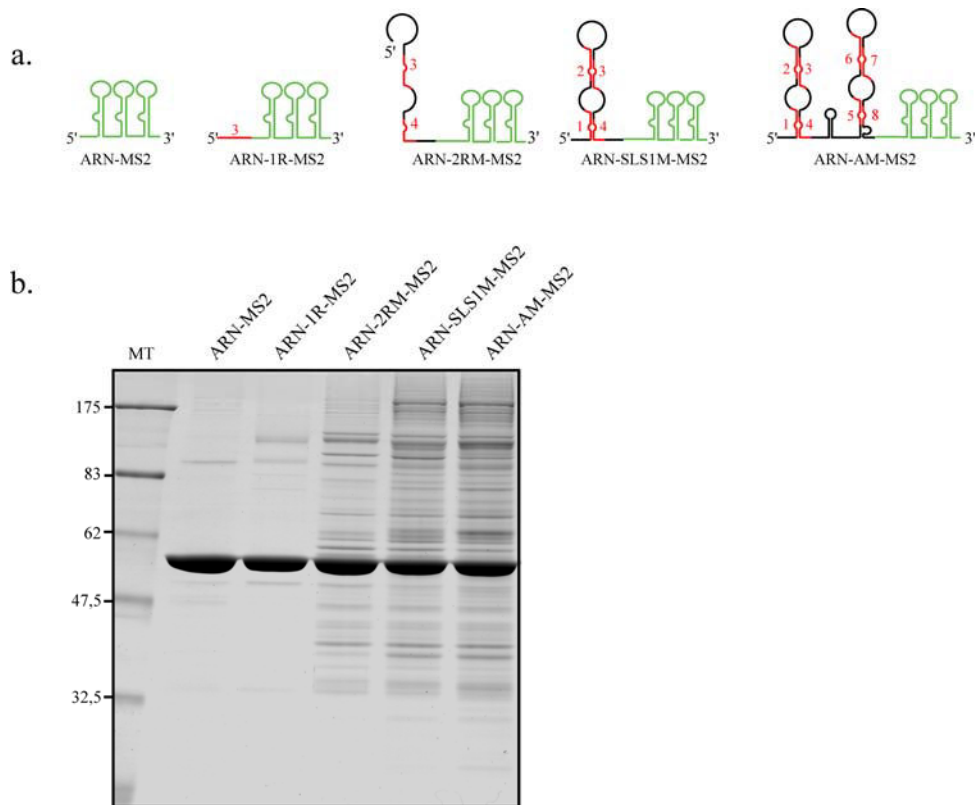
Ayant ainsi appris à préparer des extraits nucléaires, j'ai effectué un séjour de plusieurs semaines dans le laboratoire de P Avner (Institut Pasteur, Paris) afin de m'initier à la culture des cellules ES. Pendant ce séjour, j'ai pu préparer des extraits nucléaires de cellules ES. Pour cela, nous avons utilisé la lignée cellulaire ES murine LF2 femelle qui présente l'avantage de croître sans cellules nourricières (fibroblastes), ce qui permettait de ne pas avoir de contamination par des fibroblastes. Le protocole utilisé pour la préparation des extraits nucléaires est détaillé dans la partie Matériel et Méthodes.

Nous avons réalisé les expériences de purification des RNP avec des extraits nucléaires de cellules ES non différenciées (J0) avec l'ARN contrôle MS2 et les ARN-1R-MS2, ARN-2RM-MS2, ARN-SLS1M-MS2, et ARN-AM-MS2 qui contiennent respectivement 1, 2, 4 répétitions ou l'ensemble de la région des A-repeats (**Fig. 59a**). Les

profils de fractionnement par SDS-PAGE des protéines contenues dans les éluats sont présentés dans la **Figure 59b**.

L'examen de ces profils permettait de constater que les compositions en protéines des complexes formés avec les différents ARN variaient. Comme observé avec l'extrait nucléaire de cellules HeLa, le nombre de protéines augmentait avec le nombre d'éléments répétés, et les profils obtenus en présence de 4 répétitions et pour la région des A-repeats entière étaient très similaires. Cette observation a été vérifiée par une analyse par spectrométrie de masse. Les pistes du gel SDS-PAGE ont été découpées de façon systématique tous les 2 mm et les protéines contenues dans les bandes ont été analysées par spectrométrie de masse. Ceci a permis d'identifier 7 protéines pour l'ARN-MS2, 6 protéines pour l'ARN-1R-MS2, 33 protéines pour l'ARN-2RM-MS2, 95 protéines pour l'ARN-SLS1M-MS2, 93 protéines pour l'ARN-AM-MS2. Parmi les protéines identifiées, nous avons trouvé dans les RNP formées avec la région des A-repeats murine entière ou avec 4 répétitions, séquence suffisante pour former la structure SLS1M, 4 des 5 protéines du complexe PRC2 (Ezh2, Suz12, RbAp46 et RbAp48). Etant donné l'importance de ce complexe dans l'inactivation du chromosome X, nous avons alors concentré nos efforts sur l'étude des déterminants structuraux requis pour la fixation de ces composants. L'ensemble des résultats concernant la structure secondaire de la région des A-repeats et son implication dans le recrutement du complexe PRC2 a permis la rédaction d'un article soumis à PLoS qui est présenté dans les pages suivantes de ce manuscrit.

L'ensemble des autres résultats obtenus par purification des RNP sera explicité dans une partie qui fait suite à l'article et qui est consacrée à l'examen de l'ensemble des données de protéomique.



**Figure 59 : Analyse des protéines contenues dans les complexes ribonucléoproteiques purifiés par chromatographie d'affinité MBP/MS2 formés dans un extrait nucléaire de cellules ES non différenciées avec différents fragments de la région des A-repeats de l'ARN Xist murin**

a. Représentation schématique des ARN utilisés lors de cette expérience. L'ARN-MS2 correspond au motif MS2. Il est utilisé comme contrôle. Les ARN-1R-MS2, ARN-2RM-MS2, ARN-SLS1M-MS2 et ARN-AM-MS2 correspondent à 1, 2, 4 et 8 éléments répétés, respectivement.

b. Fractionnement sur gel SDS-PAGE 10% des protéines des complexes RNP et coloration par bleu de Coomassie colloïdal. Deux marqueurs de poids moléculaire ont été déposés de chaque côté du gel (MT).

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

Article soumis à PLoS Biology

## 2D 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>3</sup>, Anne Savoye<sup>1</sup>, Virginie Marchand<sup>1\*</sup>, Agnès Dubois<sup>2</sup>, Sarah Sanglier-Cianférani<sup>3</sup>, Alain Van Dorsselaer<sup>3</sup>, Philippe Clerc<sup>2</sup>, Philip Avner<sup>2</sup>, Athanase Visvikis<sup>1</sup> and Christiane Branlant<sup>1</sup>

*1 AREMS, Nancy université, UMR 7214 CNRS-UHP 1, Faculté des sciences et techniques, BP 70239, 54506 Vandoeuvre-lès-Nancy Cedex, France*

*2 Génétique Moléculaire Murine, Institut Pasteur and CNRS2578 25 Rue du Docteur Roux, 75015 Paris, France*

*3 Laboratoire de Spectrométrie de Masse BioOrganique, Institut Pluridisciplinaire Hubert Curien, Département des Sciences Analytiques, Université de Strasbourg, CNRS UMR 7178, ECPM, 25, rue Becquerel, 67087 Strasbourg, France*

*\* Present adress : Developmental Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse, 1, 69117 Heidelberg, Germ*

+ Corresponding author

Christiane BRANLANT

ARN, RNP, structure-fonction-maturation, Enzymologie Moléculaire et Structurale, Nancy université, UMR 7214 CNRS-UHP 1, Faculté des sciences et techniques, BP 70239

54506 Vandoeuvre-lès-Nancy Cedex, France

Tel: 33 3 83 68 43 03 Fax: 33 3 83 68 43 07

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

**ABSTRACT*****Background***

In placental mammals, inactivation of one of the X chromosomes in female cells ensures sex chromosome dosage compensation. The 17kb 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.

***Methodology/Principal findings***

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 2D structure for the A region that contains two long stem-loop structures each including 4 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 4-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.

***Conclusions***

Our data with its emphasis on the importance of inter-repeat pairing changes fundamentally our conception of the 2D structure of the A region of Xist RNA and supports its possible implication in recruitment of the PRC2 complex.

## 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 H3K27me3, H3K9me3, 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 embryos carrying a mutated Xist<sup>ΔA</sup> gene inherited from males are selectively lost during embryogenesis underlines the importance of this element [9]. Recent data has 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>ΔA</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 tri-methylation of histone H3 is catalysed by Ezh2 [12,14] and both Eed and Suz12 are required for this activity [20]. Recently a short 1600 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 2D structure, studies aimed at establishing the 2D 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 2D 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 2D 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-E.

### M-fold assisted modelling of the mouse A region 2D 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 [11] (Figure 2). 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 2D 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. 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 3C). 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 3A), since in this case, the entire polyA region would interact with the polyU track located upstream of repeat 3, which was not in agreement with the probing data (Figure 3A). 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 stem-loop structure with an A rich terminal loop (Model 2, Figure 3b). 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 3C). 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 (Figure 3A and B) 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 (Figure 4 and 5 and Figure S3 A-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: i) A 2D 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 3C), 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, 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 2D 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 4).

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, elephant A region sequences (Figure S5 and S6). Nucleotide sequence conservation extends out beyond the repeats themselves for the majority of the repeats.

### **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 6). This Model predicts that the P1-P5 and P2-P4 pairs of 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 3 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 6). 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. 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 adopts a single unique 2D 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 3 proposed structures (Figure S6). The oligonucleotide pairs used are shown in Figure 6, along with examples of typical fluorescence intensity spectra recorded in FRET experiments for the P2-P4 and P3-P6 pairs (Figure 6D). 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 5 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.

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 7A). 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 4 repeats and Suz12 when the entire A region was used (Figure 7C). The control HIV-1 RNA bound none of these proteins (Figure 7B).

To explore further 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 7 D and E) 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 this complex.

## 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 4 repeats and long U rich terminal and internal loops. Our proteomic analyses suggests that these 4-repeat structures may correspond to functional modules initiating the assembly of the PRC2 complex.

### A new conception of the 2D structure of the A repeat region

Until now, both computer [11] and experimental analyses [22] of the possible 2D 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 2D 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 2D 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 2D 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 2D 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 4 repeats (4-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 polyA sequence forming a short stem loop with a polyU sequence. Interestingly, nucleotide sequence conservation in the A region extends to the spacer extremities, which contribute to the possibility of forming the 4-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 4-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 4-repeat structures may be stabilized *in cellulo* by RNA-protein interactions. Interestingly, protein PTB which contains 4 RNA Recognition Motifs (RRMs), that are each able to interact with UCUU(C), UUCUCU or CUCUCU sequences, showed high affinity for the A region in nuclear extract binding experiments [25] (Figure 7). As UCUU motifs are present on each side of the large internal loops in the 4-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 4-repeat

structure as suggested by previously proposed models for protein PTB-RNA interaction [26].

### **Possible functional implication of the 4-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 4-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 repeat region in either binding Suz12 or in stabilising the binding of Suz12 to the 4 repeat structure.

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 (data not show). 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 results show that Suz12 has distinct association properties compared to other components of the PRC2 complex and suggest that it may require the presence of an additional RNA motif, which is formed only by the entire A region, or more simply require more than 4 repeats for efficient association with the RNA. The observation of Wutz and colleagues, (2002) that the presence of at least 5.5 repeats was needed to initiate inactivation [11] may be related to this finding. Additional support for the functional significance of the 4-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 4-repeat structure, whereas the two inactive variants (XS1 and XNX) cannot (Figure S9).

### **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 events 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 2D structures in solution were probed according 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 addition of 1  $\mu$ l of 62.5 mM MgCl<sub>2</sub> to a final concentration of 3.25 mM MgCl<sub>2</sub>. After 10 min preincubation at room temperature, RNase T1 (0.02 or 0.0375 U/ $\mu$ l) or T2 (0.025 or 0.0375 U/ $\mu$ l) were added under conditions such that they cleaved single-stranded segments. V1 RNase (2.5x10<sup>-3</sup> or 5x10<sup>-3</sup> U/ $\mu$ l) was used to cleave double stranded and stacked residues. DMS [1  $\mu$ l 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  $\mu$ l 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  $\mu$ l 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  $\mu$ M) 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 flx 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 account of 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 2h at 4°C. After three washes with 500 µl of buffer D, 1 mg of nuclear extract in 150 µl of buffer D containing 5 µM 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 µl 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 (3000 Ci/mmol), purified and quantified according to [34]. About 70 pmol of the RNA were denatured, renatured as described above and incubated with 30 µg of nuclear extract for 30 min at room temperature with constant agitation. About 40 µl of Protein G-sepharose beads suspension blocked with BSA (2 µg) and coated for 2 h at 4°C with 10 µl of each antibodies were incubated with the RNP complexes for 2 h at 4°C in 300 µl 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 µl of immunoselection buffer containing 0.5% NP40. RNAs were phenol extracted, ethanol precipitated, fractionated on 7% polyacrylamide gel and analysed by autoradiography.

**References**

1. Lyon MF (1961) Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* 190: 372-373.
2. 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.
3. 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.
4. Cohen HR, Panning B (2007) *XIST* RNA exhibits nuclear retention and exhibits reduced association with the export factor TAP/NXF1. *Chromosoma* 116: 373-383.
5. 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.
6. 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.
7. 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.
8. Brockdorff N (2002) X-chromosome inactivation: closing in on proteins that bind *Xist* RNA. *Trends Genet* 18: 352-358.
9. 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.
10. 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.
11. Wutz A, Rasmussen TP, Jaenisch R (2002) Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nat Genet* 30: 167-174.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.

19. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, et al. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129: 1311-1323.
20. 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.
21. 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.
22. Duszczyc 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.
23. 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.
24. 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.
25. Perez I, McAfee JG, Patton JG (1997) Multiple RRM contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. *Biochemistry* 36: 11881-11890.
26. Auweter SD, Allain FH (2008) Structure-function relationships of the polypyrimidine tract binding protein. *Cell Mol Life Sci* 65: 516-527.
27. 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.
28. 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.
29. Mouglin 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.
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.

**Acknowledgments**

Professor R. Lührman (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. Professors I Motorine and C Aigueperse (AREMS, Nancy) are acknowledged for their advice in implanting this approach at Nancy. V Ségault (AREMS, Nancy) is thanked for advice concerning the immunopurification assays. S Mazerès (IPBS, Toulouse) is thanked for helping us to define the FRET experimental protocol. S Maenner, A Savoye and V Marchand were PhD fellows of the French Ministry “Enseignement Supérieur et la Recherche” and M Blaud was supported by a post-doctoral fellowship from the “Centre National de la Recherche Scientifique” (CNRS). L Fouillen acknowledges the CNRS, Région Alsace and AliX for financial support. Research funds to CB were obtained from the CNRS and the French Ministry of “Enseignement Supérieur et la Recherche”. PA was supported by recurrent funding from the CNRS and the Institut Pasteur and by an EU contract from the Epigenome NoE. Both CB and PA were supported by the French National Agency for Research (ANR) (contract N°ANR 07 BLAN 0047 102).

## Figure legends

### 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 lefthand 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.

### Figure 2 : Representation of experimental data on the previously proposed 2D structure of the Xist A region

Each of the 7 repeats as well as the 8<sup>th</sup> half repeat from the mouse A repeat region were 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. Squared base correspond to nucleotides which are not reactive toward chemical probes. The V1 RNase cleavages and chemical modifications that cannot be explained by the two stem-loop structure models are encircled by blue lanes.

### Figure 3: Representation of experimental data on possible 2D structures of the Xist A region involving inter-repeat interaction

In Model 1, the various stem-loop structures all involve two successive repeats, in Model 2 repeats in the 5' half of the A region interact with repeats in the 3' half of this region, in Model 3 two stem-loop structures containing 4 repeats are separated by a small stem-loop corresponding to polyA and polyU sequences. 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. The free energies of each stem-loop structures at 0°C and in 3M NaCl were calculated with the M-fold software.

### Figure 4: Probing of the 2D 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.

### Figure 5: 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 contains 4 repeats. Representation of the enzymatic cleavages and chemical modifications are as in Figures 2 and 3. The free energies of the stem-loop structures were calculated with the M-fold software.

### Figure 6: Steady-state fluorescence studies provides additional support to the A region Model 3.

In Panel (A, B and C), the binding sites of oligonucleotides P1 to P6 used in the FRET experiments are shown for the three 2D 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 Panel (D), the emission fluorescence spectra from 530 to 745 nm of the donor oligonucleotide bound alone to the RNA (green curve) was collected, as well as the emission spectra obtained in the presence of the donor and acceptor oligonucleotides (violin 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 Panel (D). The FRET efficiencies measured for the six pairs of oligonucleotides are provided in Panel (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 Panel (A, B and C) by lines joining the oligonucleotides. The thickness of the lines reflects the efficiency of the FRET effect.

**Figure 7: The PRC2 complex assembles on fragments of the A region containing at least 4 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 panel (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 Panel (D). RNAs bound to the G sepharose beads were fractionated by electrophoresis on 7% denaturing gels. Autoradiograms of the gels are shown in Panel (E). Input corresponds to 10% of the material incubated with the beads.

## Supporting Figures legends

### 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| [2] (Mouse *Xist* gene) and no. gi|340393|gb|M97168.1| [5] (Human *XIST* gene). Restriction sites and fluorescent dyes introduced by the oligonucleotides are indicated.

### 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 1) 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 are indicated by vertical bars on the right hand side of the autoradiograms.

### Figure S2 : Representation of experimental data on a 2D structure in which repeats form individual stem-loop structures

Each of the 7 repeats as well as the 8<sup>th</sup> 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.

### 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), 4242 (E) (Table S1).

### 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 are as in Figure 2. The free energies of each stem-loop structures at 0°C and in 3M NaCl were calculated with the M-fold software.

### Figure S5: Conservation of sequences surrounding the repeats in vertebrate A regions

Sequences 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 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)

**Figure S6 : The possibility to form 4-repeats stem-loop (SLS1 and SLS3) structure is conserved in vertebrates.**

SLS1's and SLS3's 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.

**Figure S7 : Control FRET experiment performed with two oligonucleotides bordering one helix.**

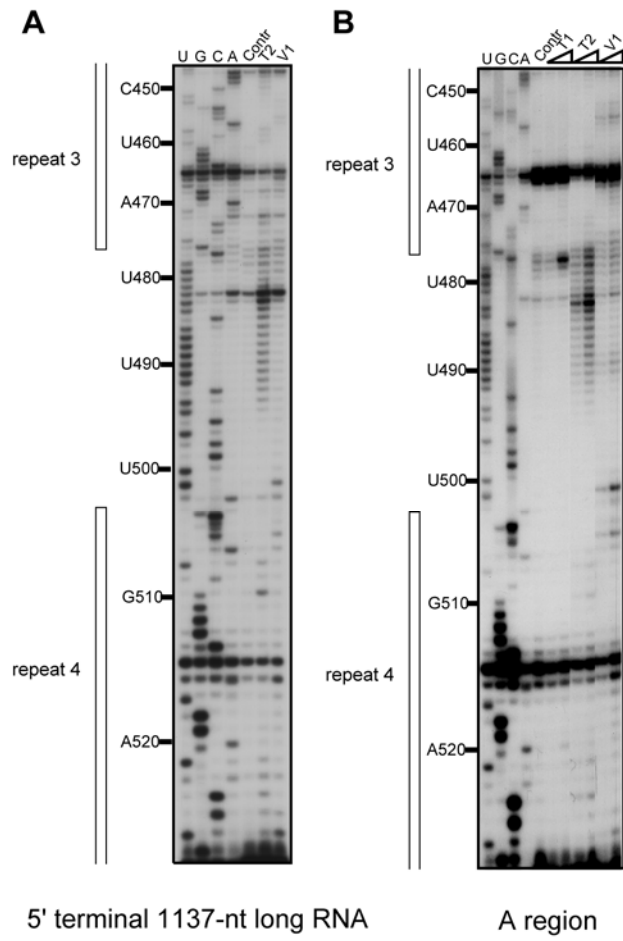
(A) Schematic presentation of the transcripts used in the control experiment.  
 (B) Fluorescence spectra obtained with the donor P1 oligonucleotide bound to naked 2R/RNA (green curve) and with oligonucleotides P1/P3' bound to the RNA (violet curve). See Legend to Figure 6 for details.

**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.

**Fig 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

**Figure 1**

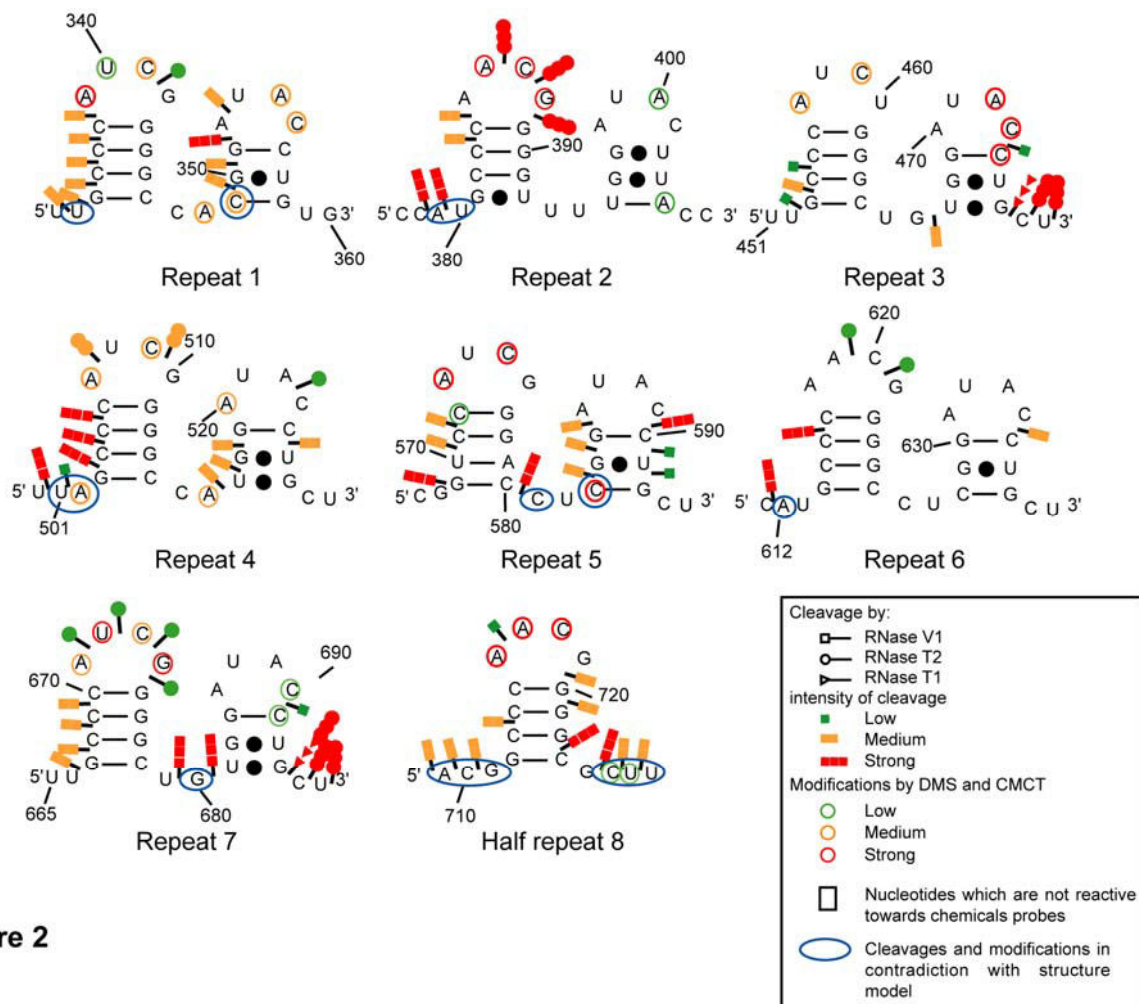


Figure 2



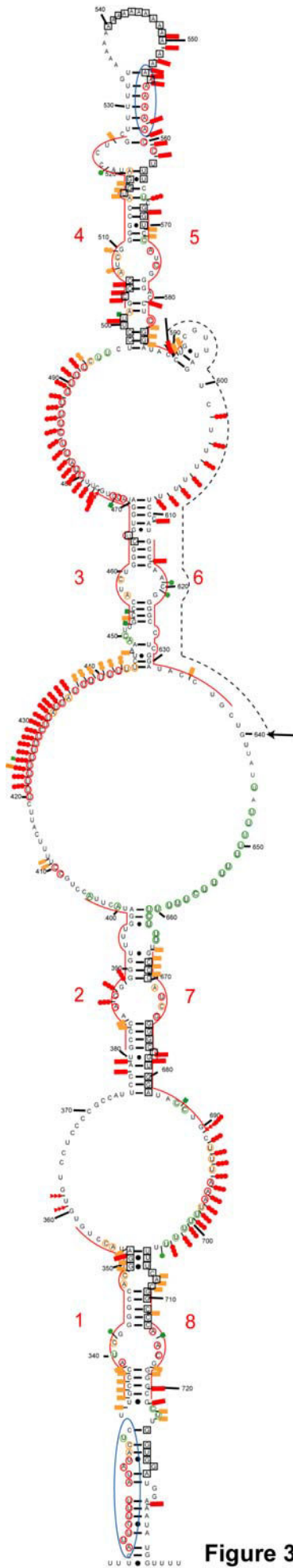


Figure 3B

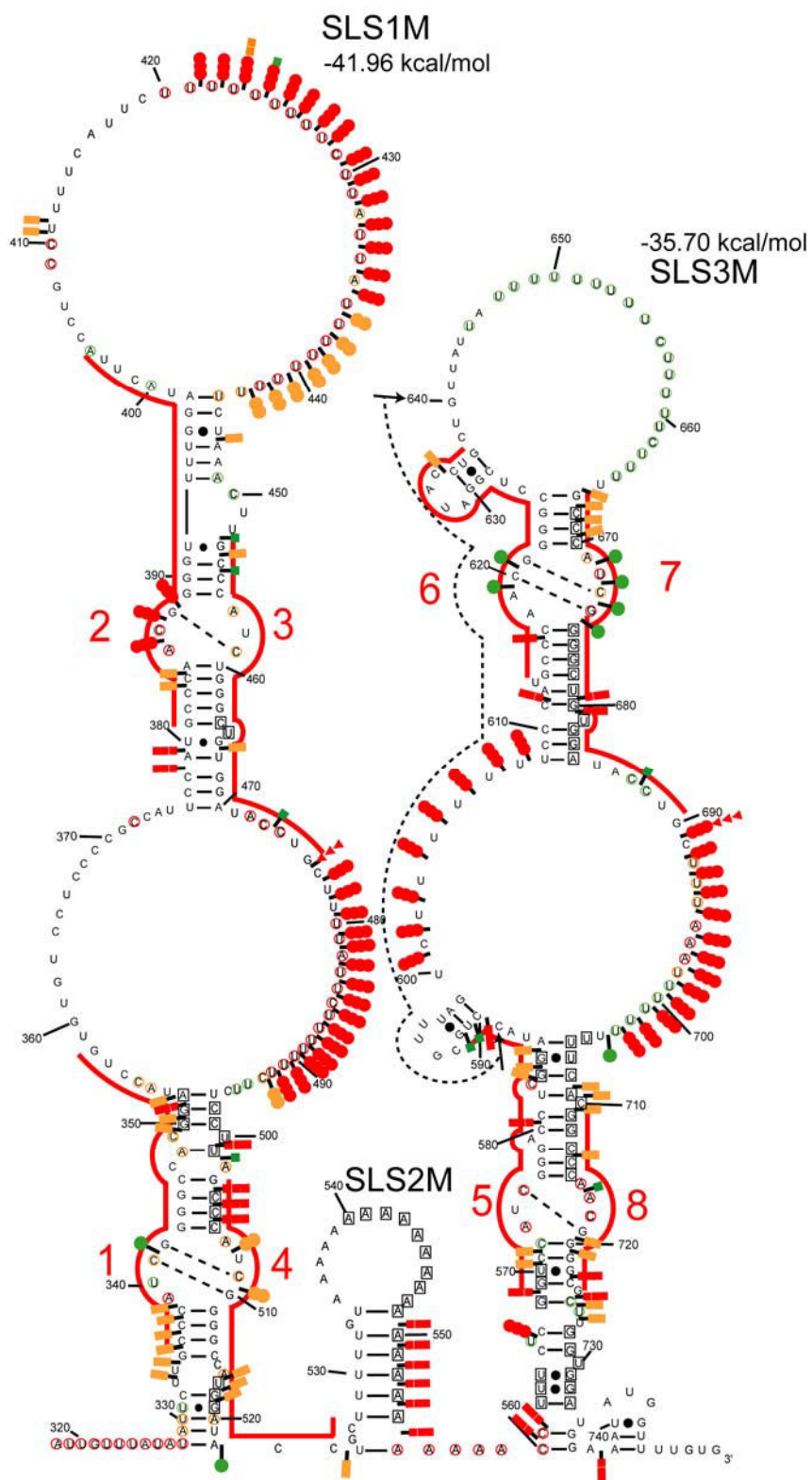
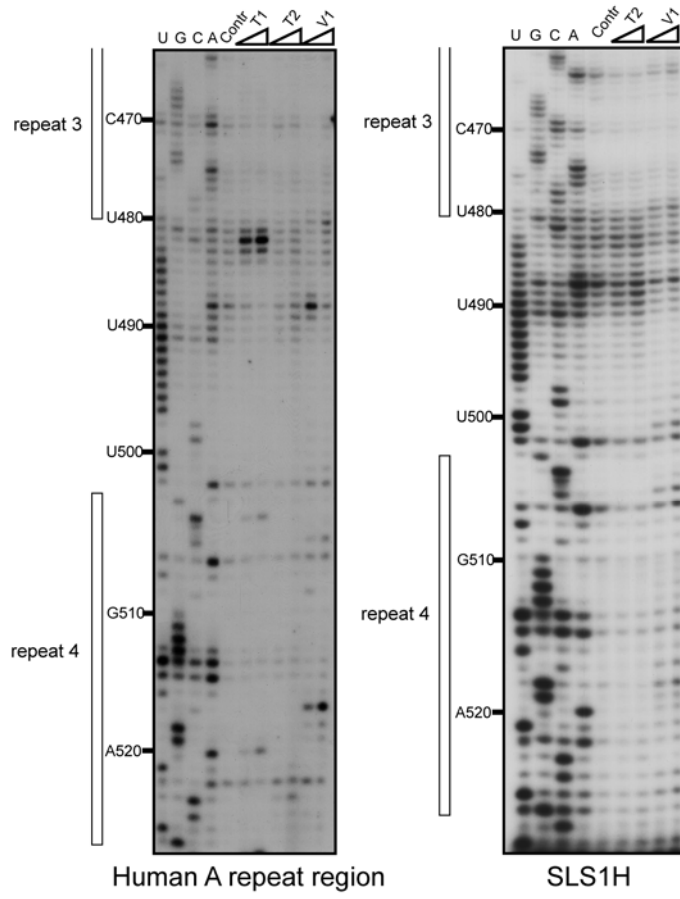


Figure 3C

**Figure 4**

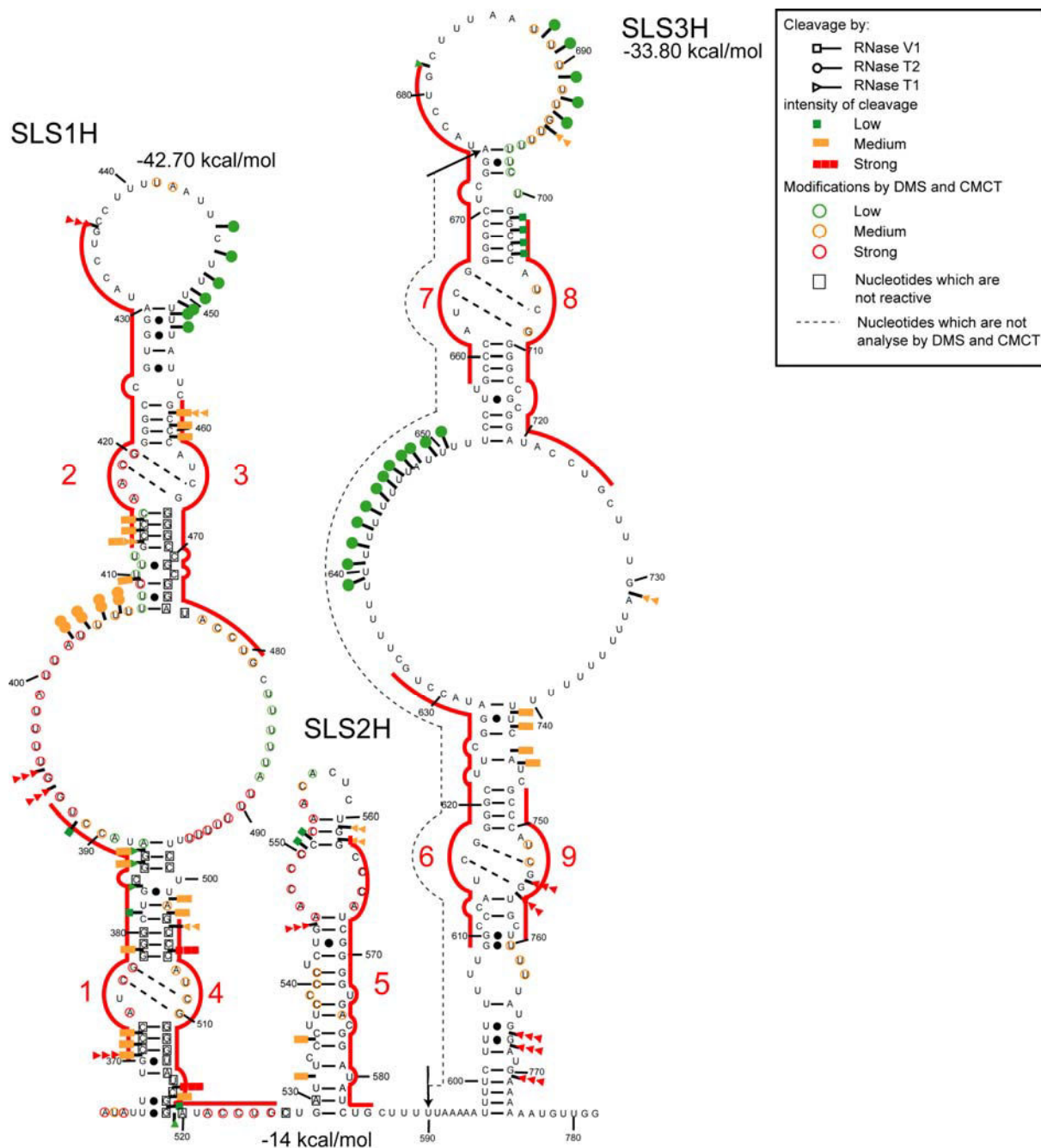


Figure 5



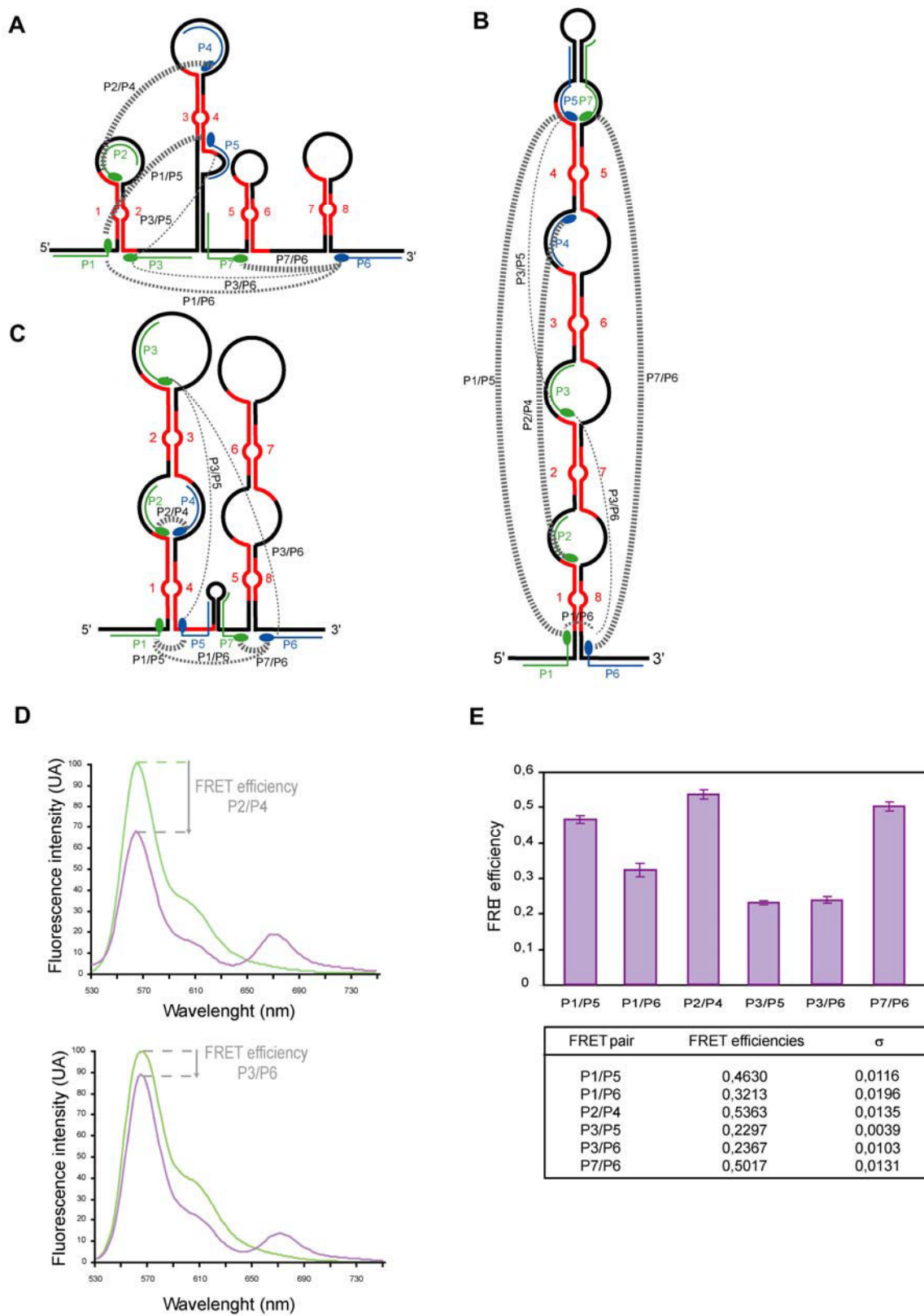


Figure 6

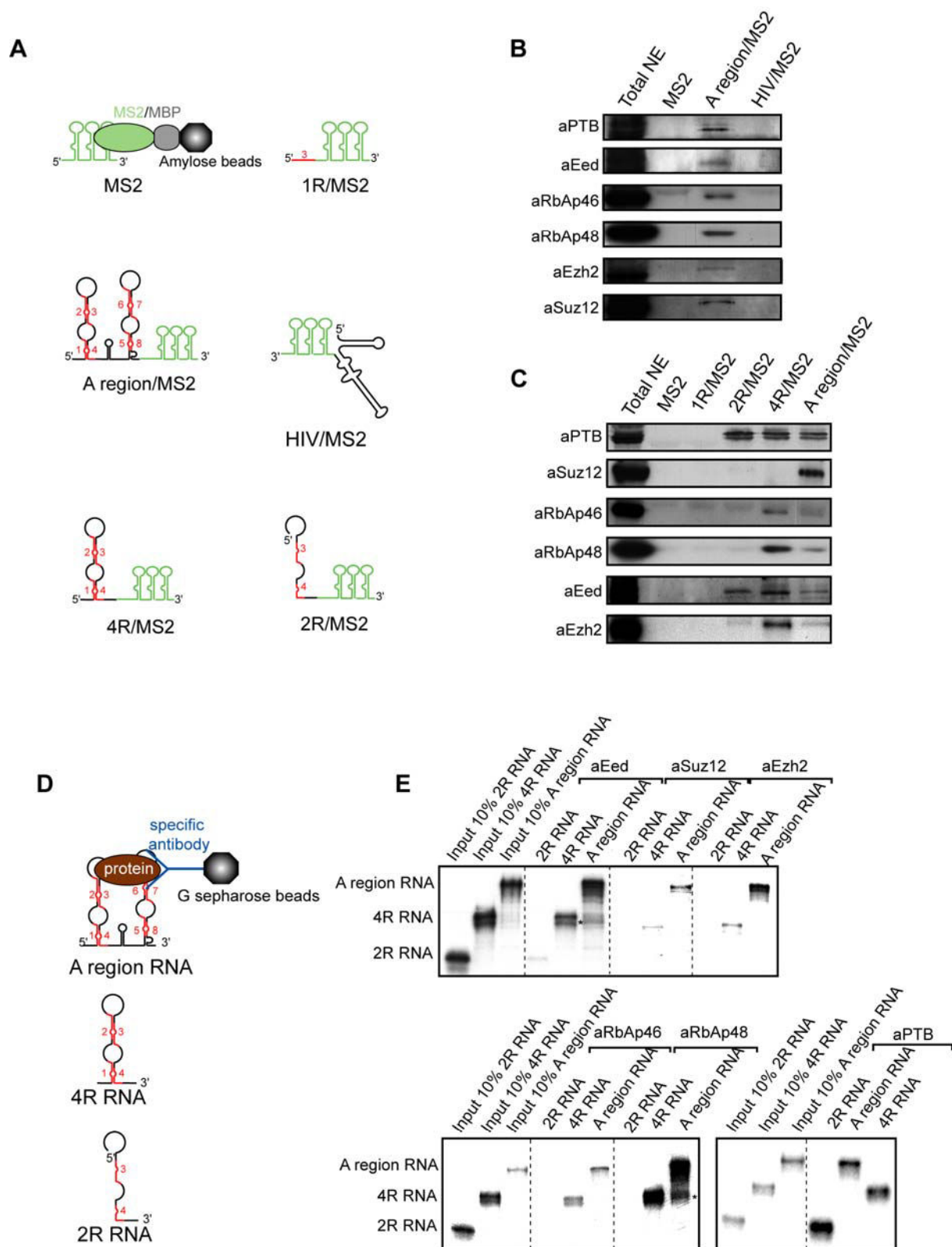


Figure 7

## SUPPORTING INFORMATION

### 2D 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>3</sup>, Anne Savoye<sup>1</sup>, Virginie Marchand<sup>1\*</sup>, Agnès Dubois<sup>2</sup>, Sarah Sanglier-Cianférani<sup>3</sup>, Alain Van Dorsselaer<sup>3</sup>, Philippe Clerc<sup>2</sup>, Philip Avner<sup>2</sup>, Athanase Visvikis<sup>1</sup> and Christiane Branlant<sup>1</sup>

*1 AREMS, Nancy université, UMR 7214 CNRS-UHP 1, Faculté des sciences et techniques, BP 70239, 54506 Vandoeuvre-lès-Nancy Cedex, France*

*2 Génétique Moléculaire Murine, Institut Pasteur and CNRS2578 25 Rue du Docteur Roux, 75015 Paris, France*

*3 Laboratoire de Spectrométrie de Masse BioOrganique, Institut Pluridisciplinaire Hubert Curien, Département des Sciences Analytiques, Université de Strasbourg, CNRS UMR 7178, ECPM, 25, rue Becquerel, 67087 Strasbourg, France*

*\* Present adress : Developmental Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse, 1, 69117 Heidelberg, Germ*

+ Corresponding author

Christiane BRANLANT

ARN, RNP, structure-fonction-maturation, Enzymologie Moléculaire et Structurale, Nancy université, UMR 7214 CNRS-UHP 1, Faculté des sciences et techniques, BP 70239

54506 Vandoeuvre-lès-Nancy Cedex, France

Tel: 33 3 83 68 43 03 Fax: 33 3 83 68 43 07

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

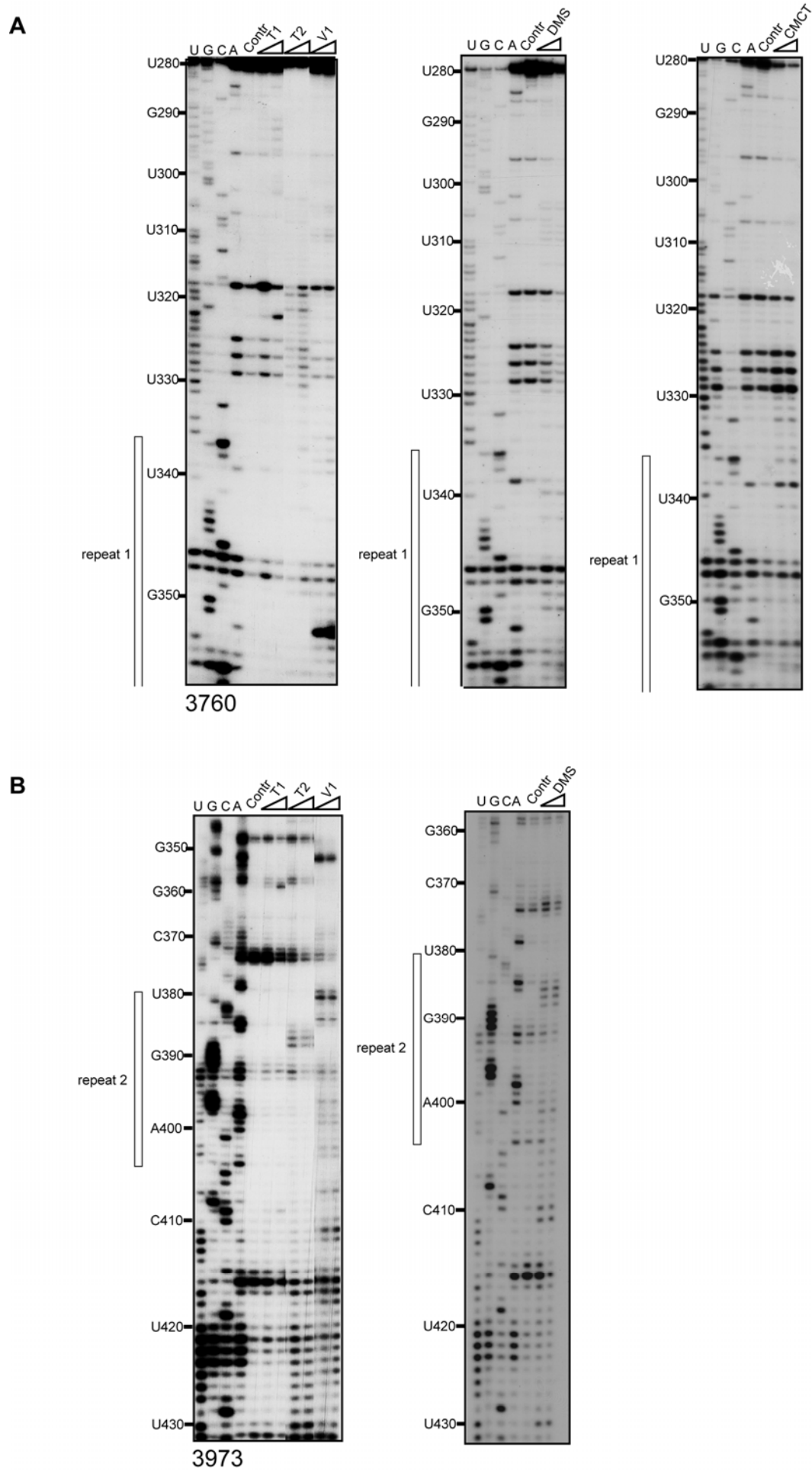
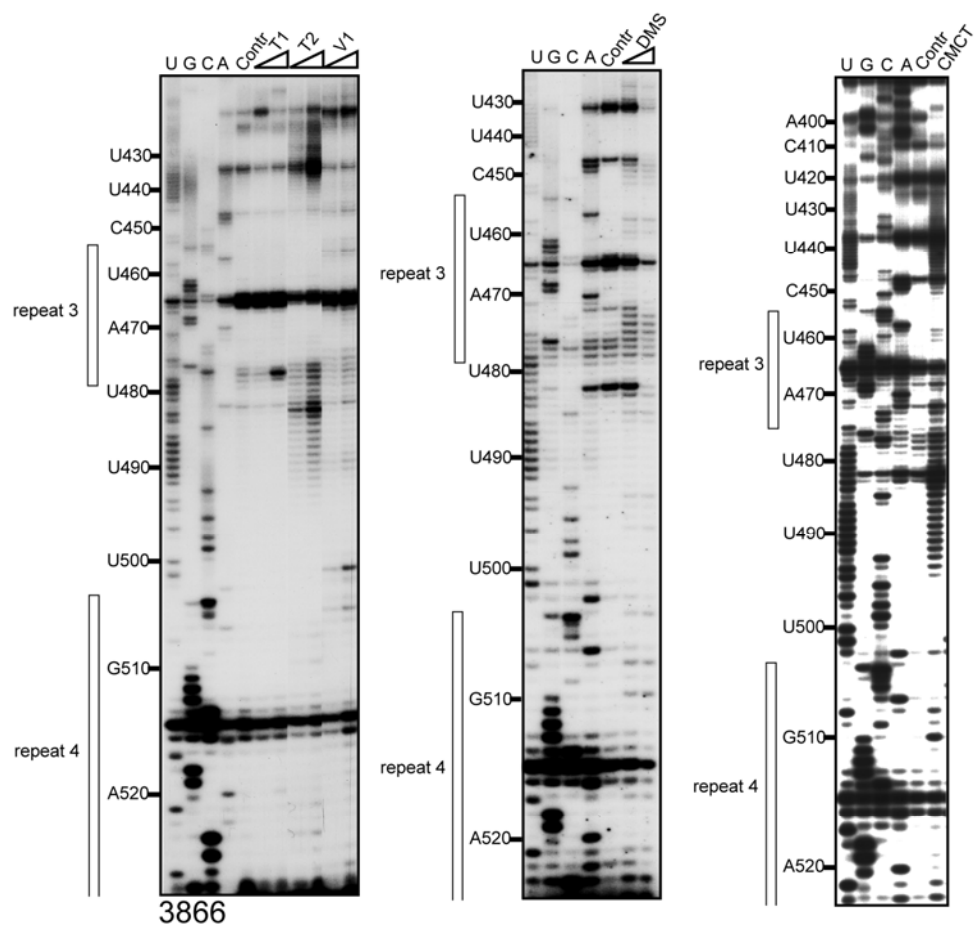


Figure S1

C



D

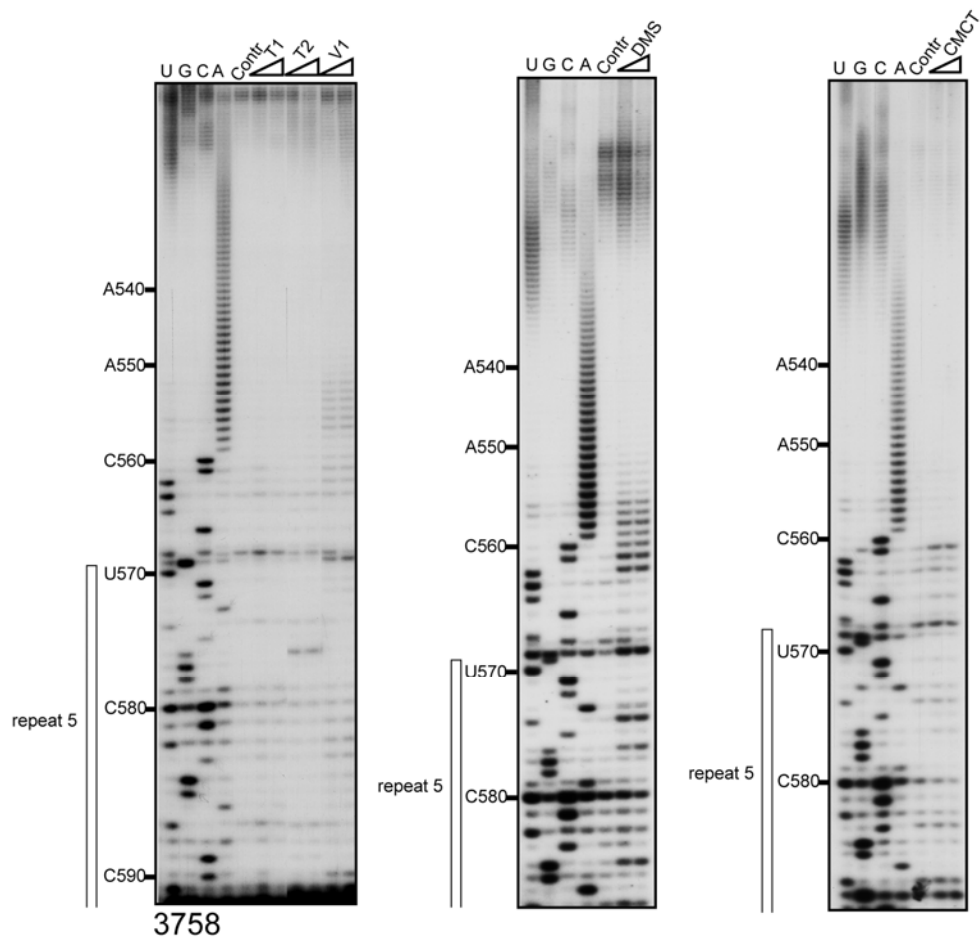
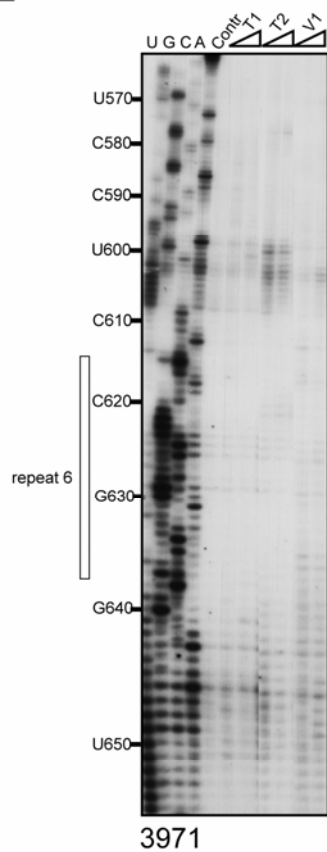


Figure S1

E



F

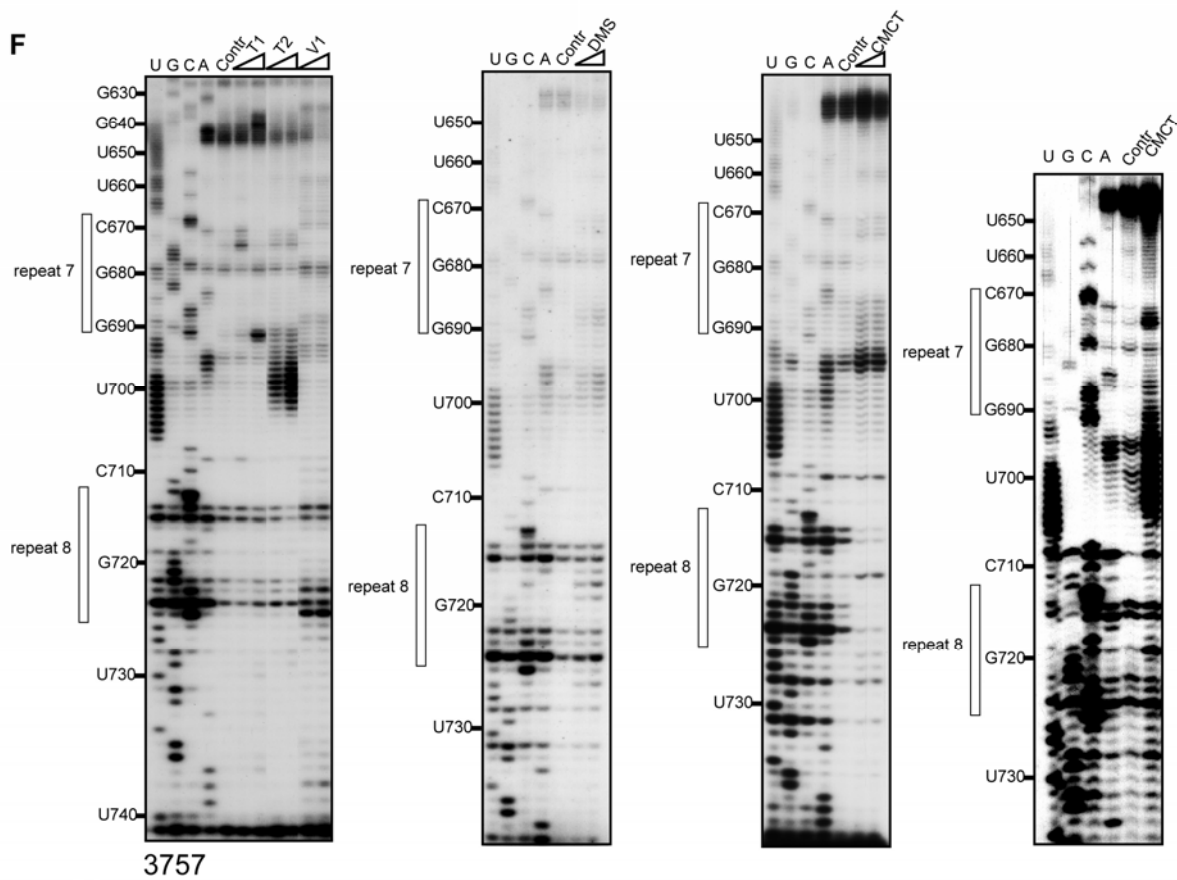


Figure S1



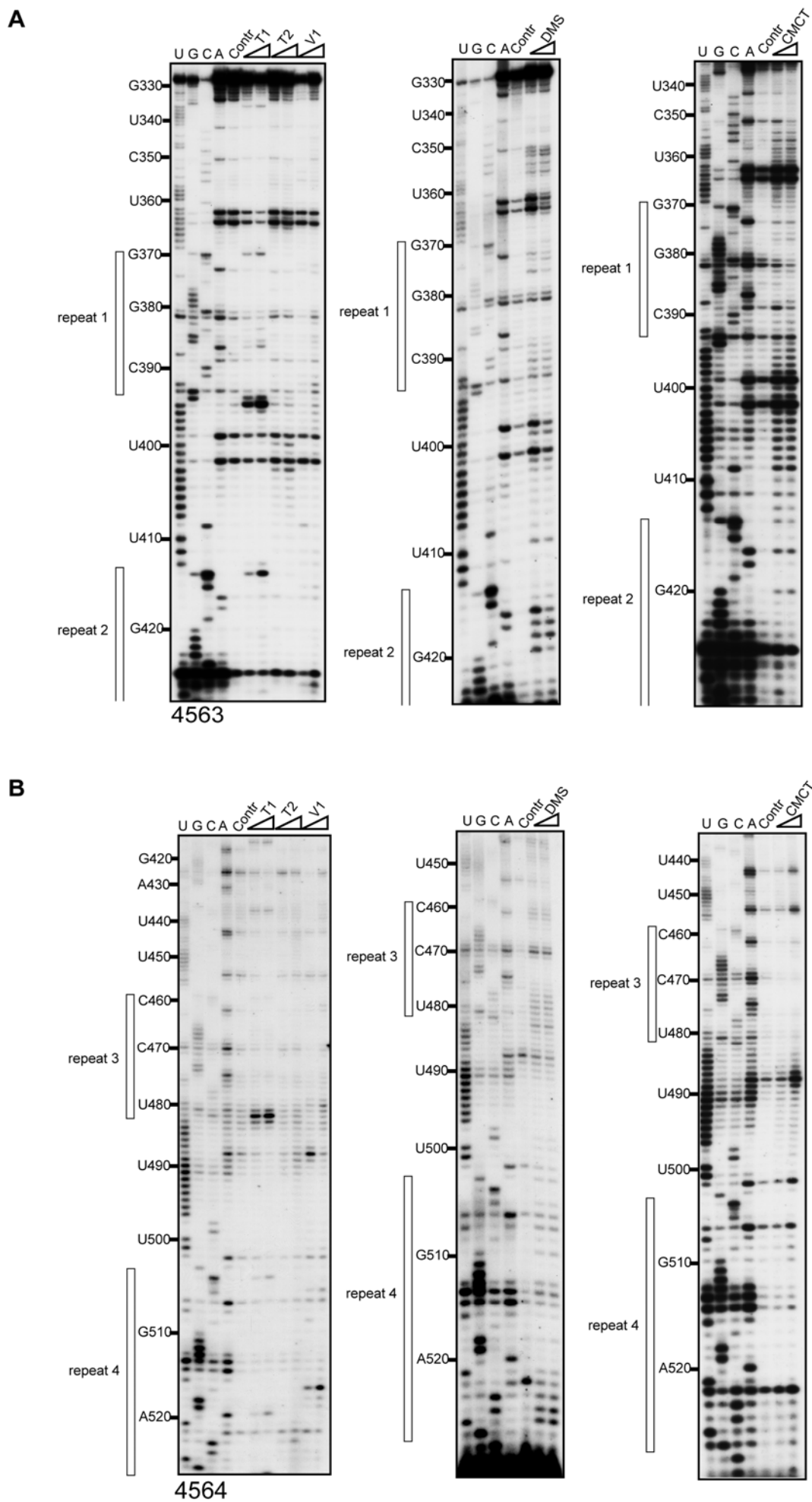


Figure S3



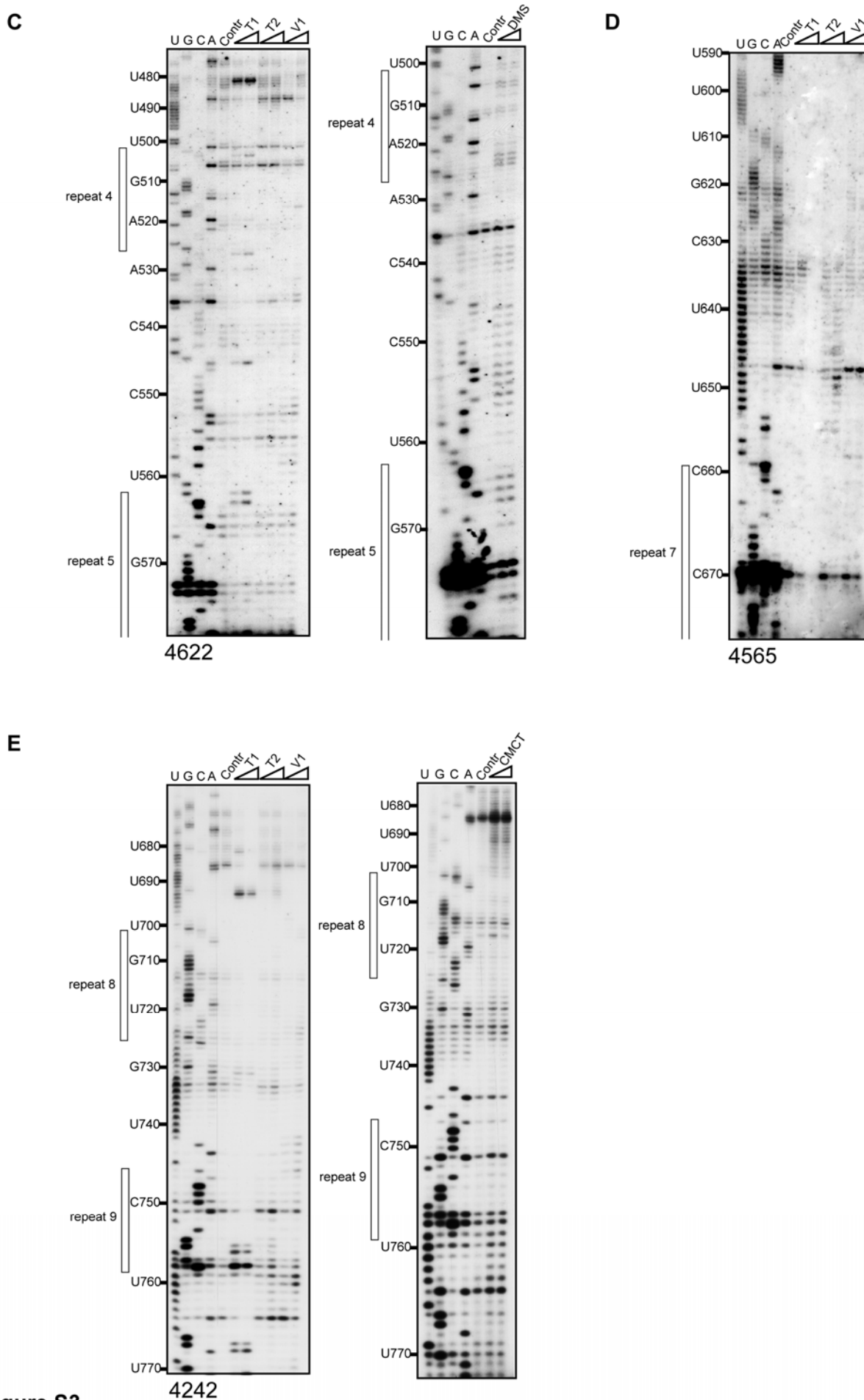


Figure S3

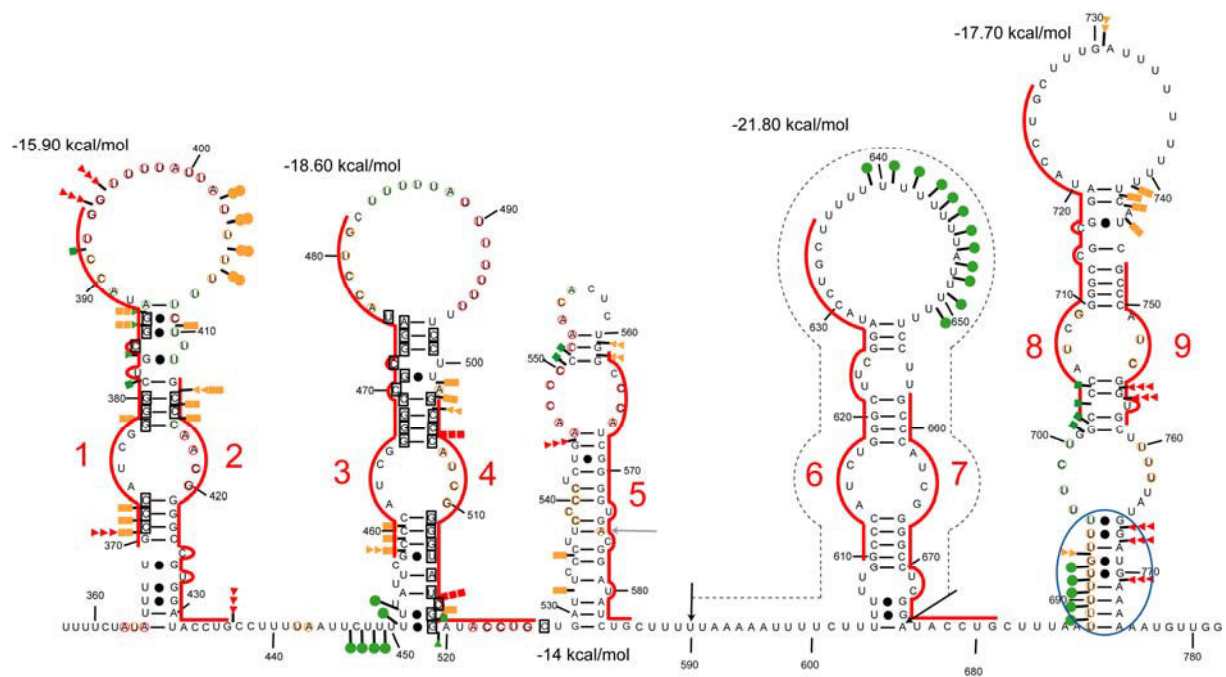


Figure S4

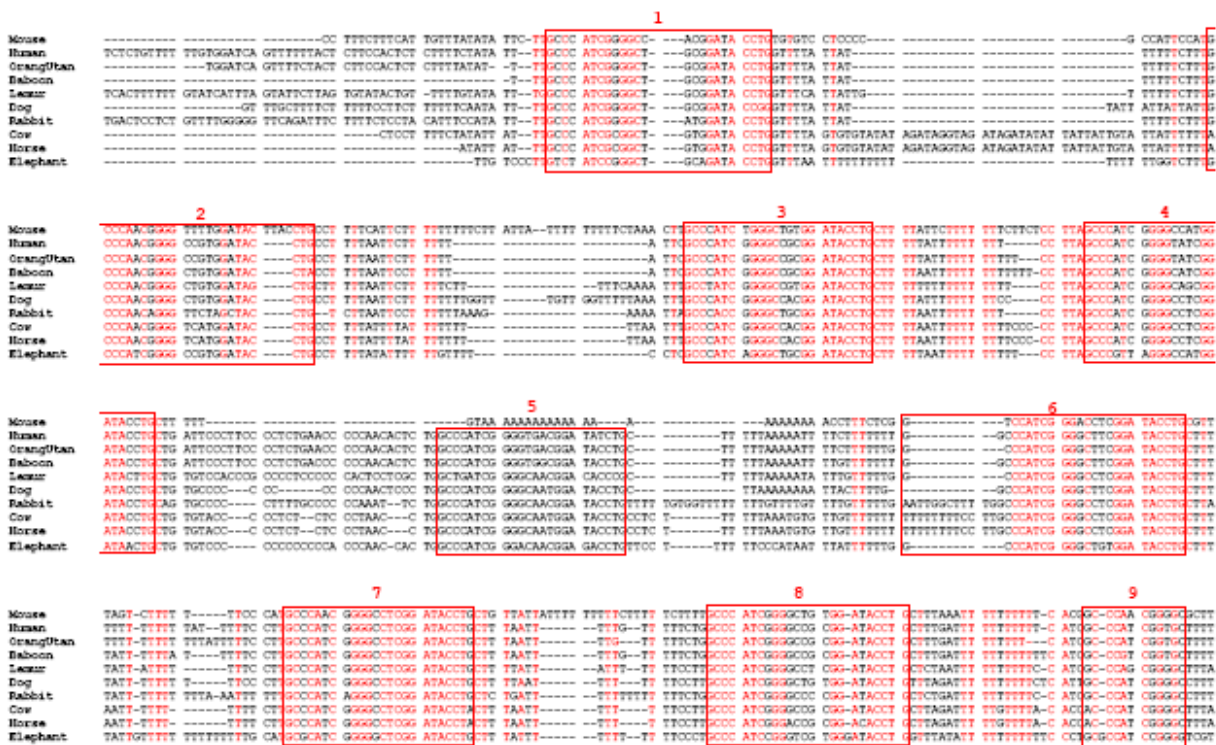


Figure S5

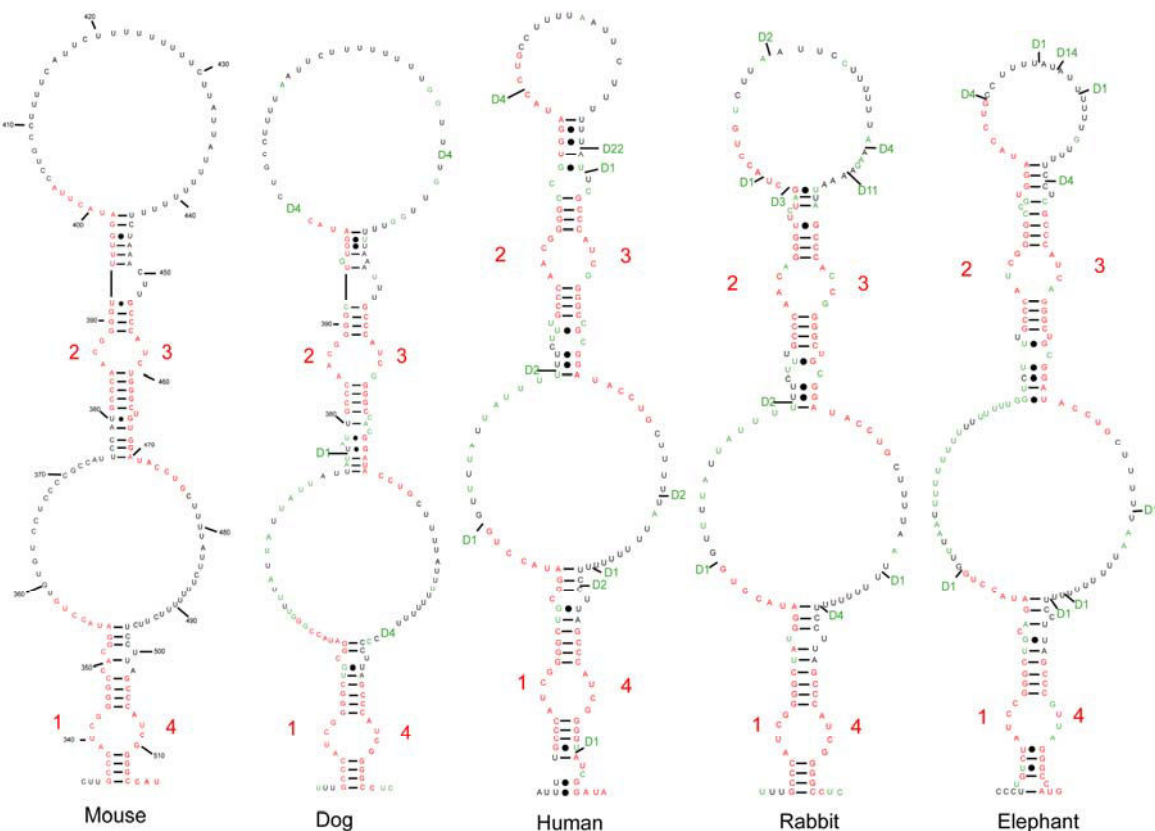
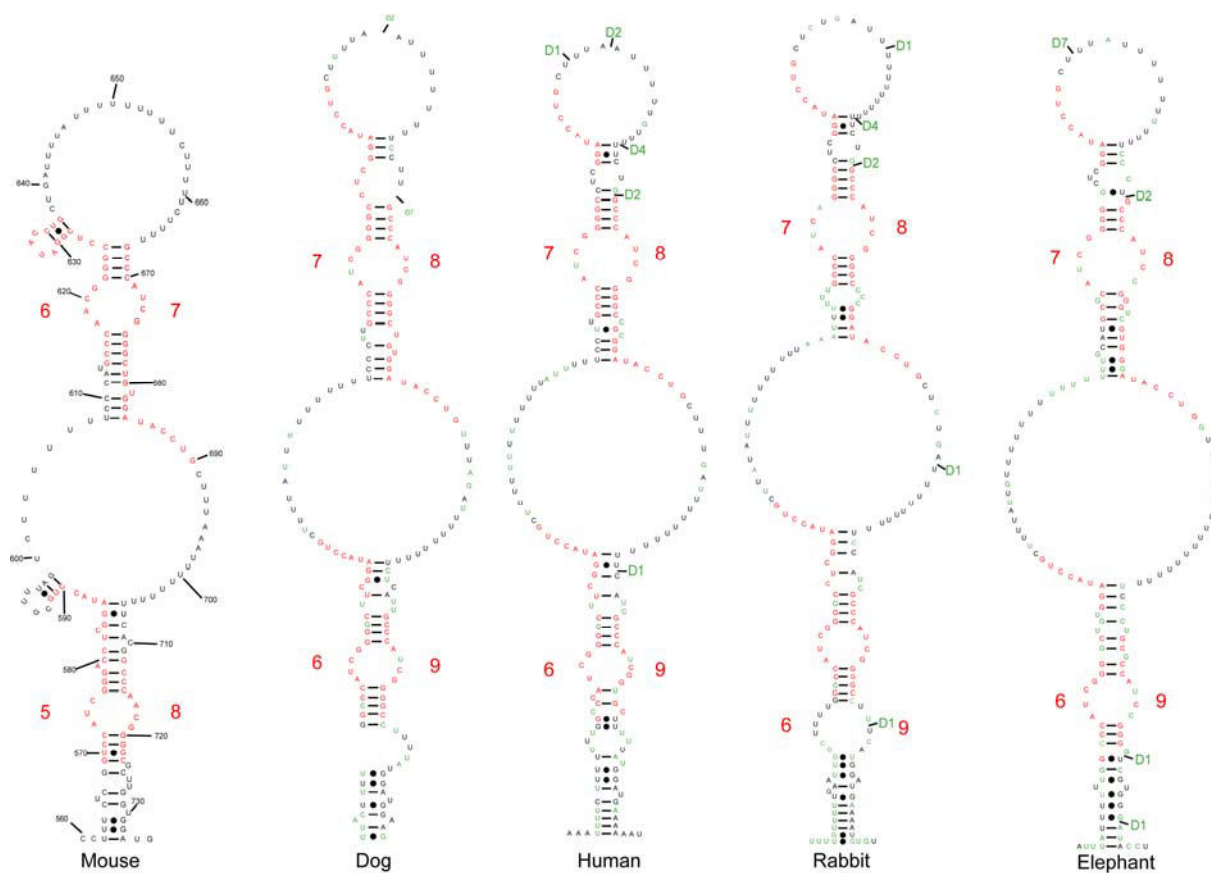


Figure S6



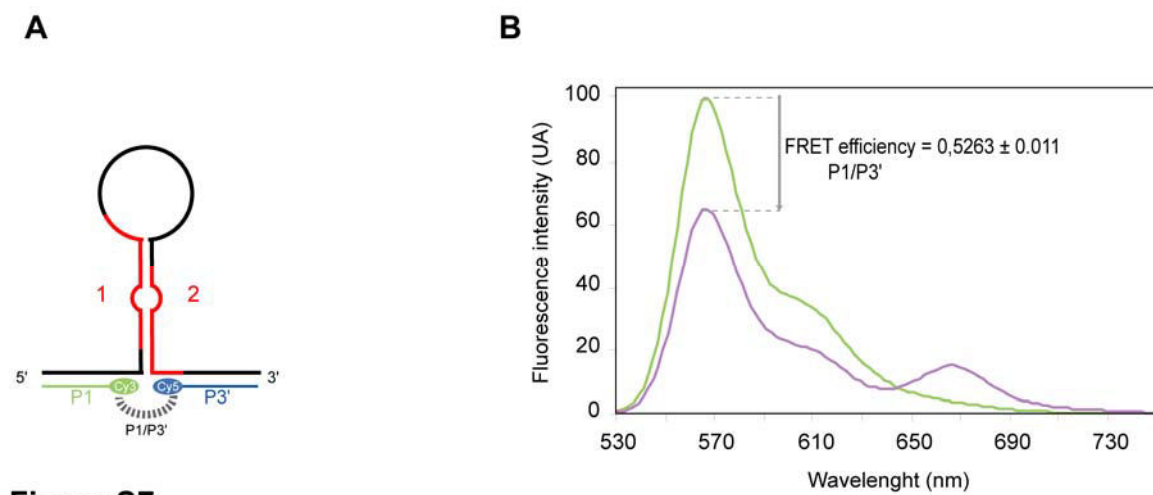
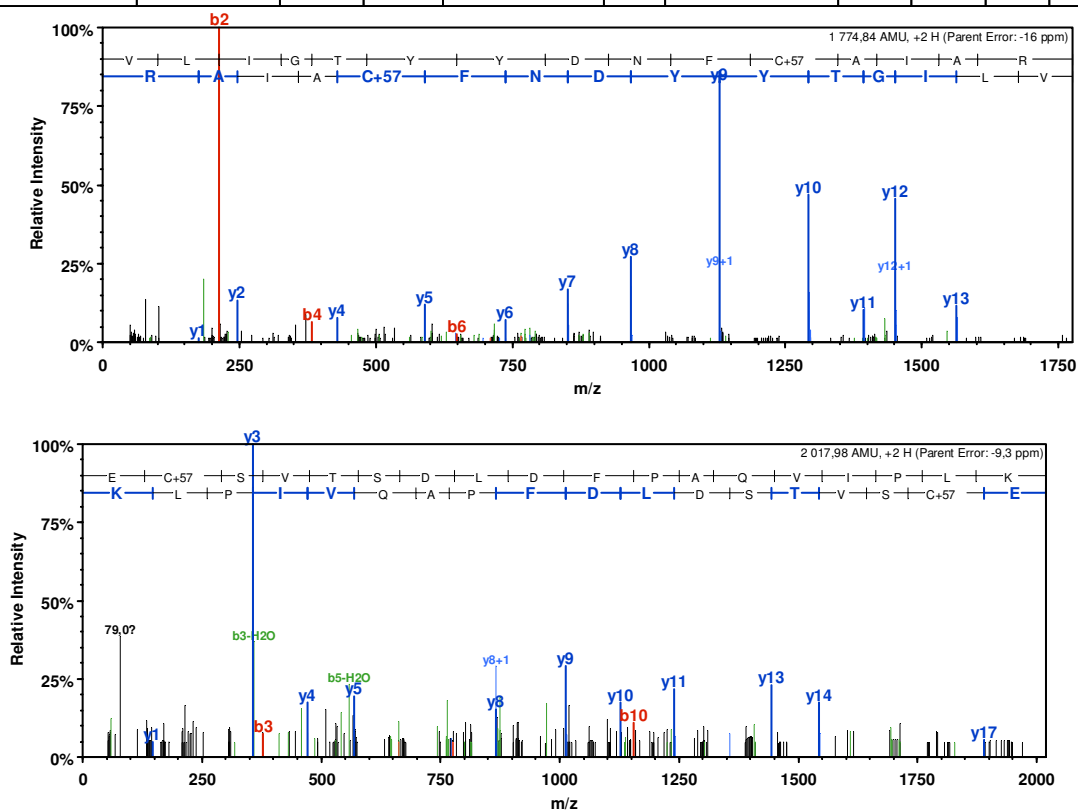


Figure S7

Figure S8

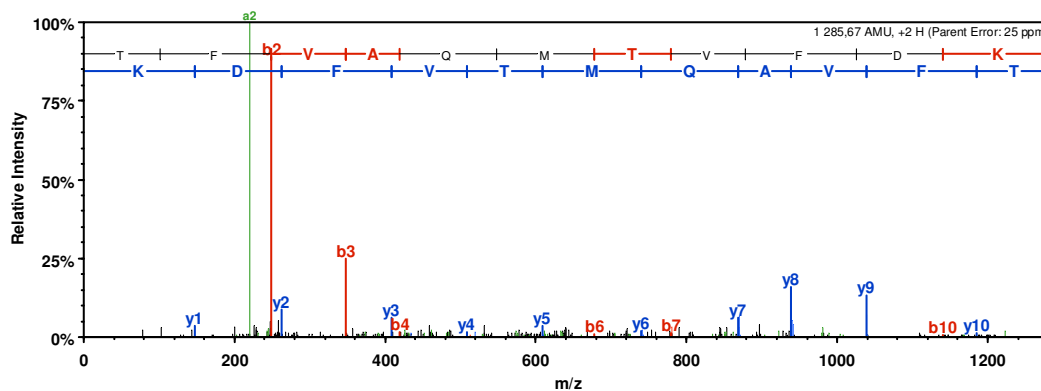
## A Identification of Enhancer of zeste homolog 2 (EZH2),

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
				ECSVTSDLDFPAQVIPLK	36.86	Carbamidomethyl	1010.00	2017.98	2	-9.3



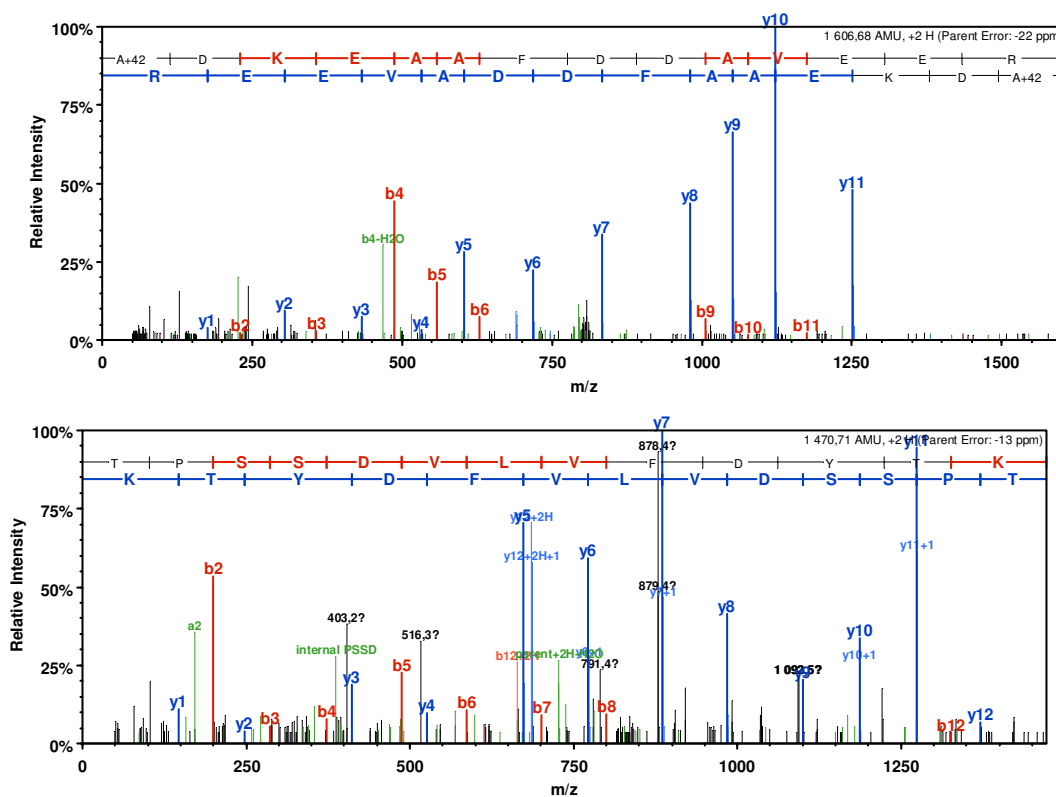
## B Identification of Polycomb protein suz12 (SUZ12),

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



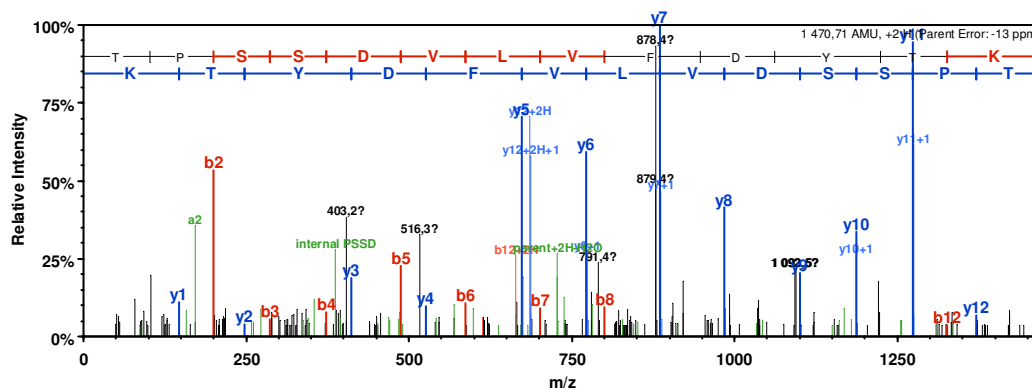
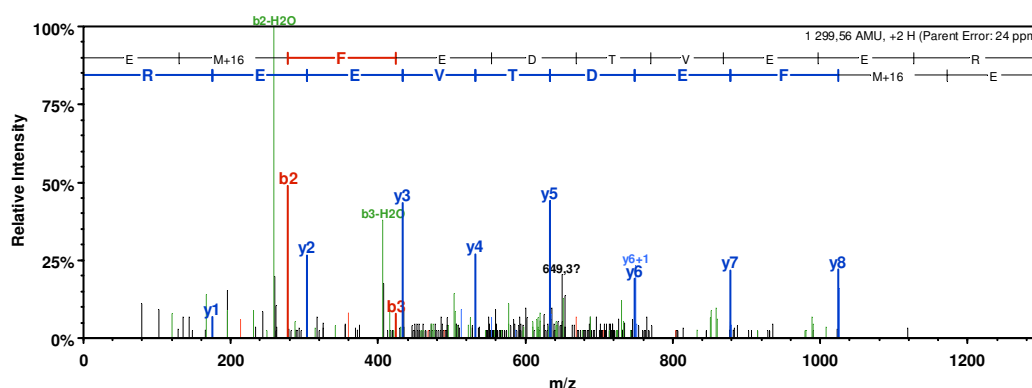
## C Identification of Retinoblastoma binding protein 4 (RbAp46),

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
				TPSSDLVFDYTK	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.

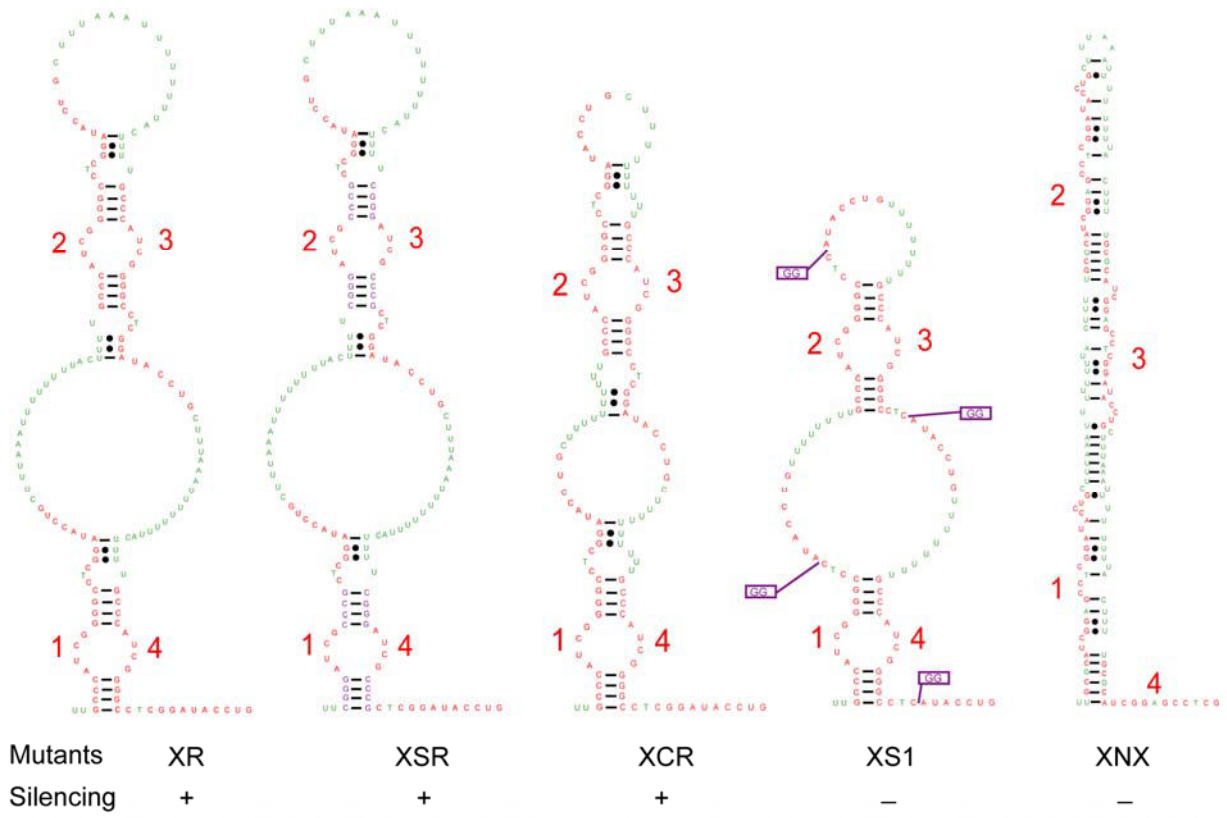


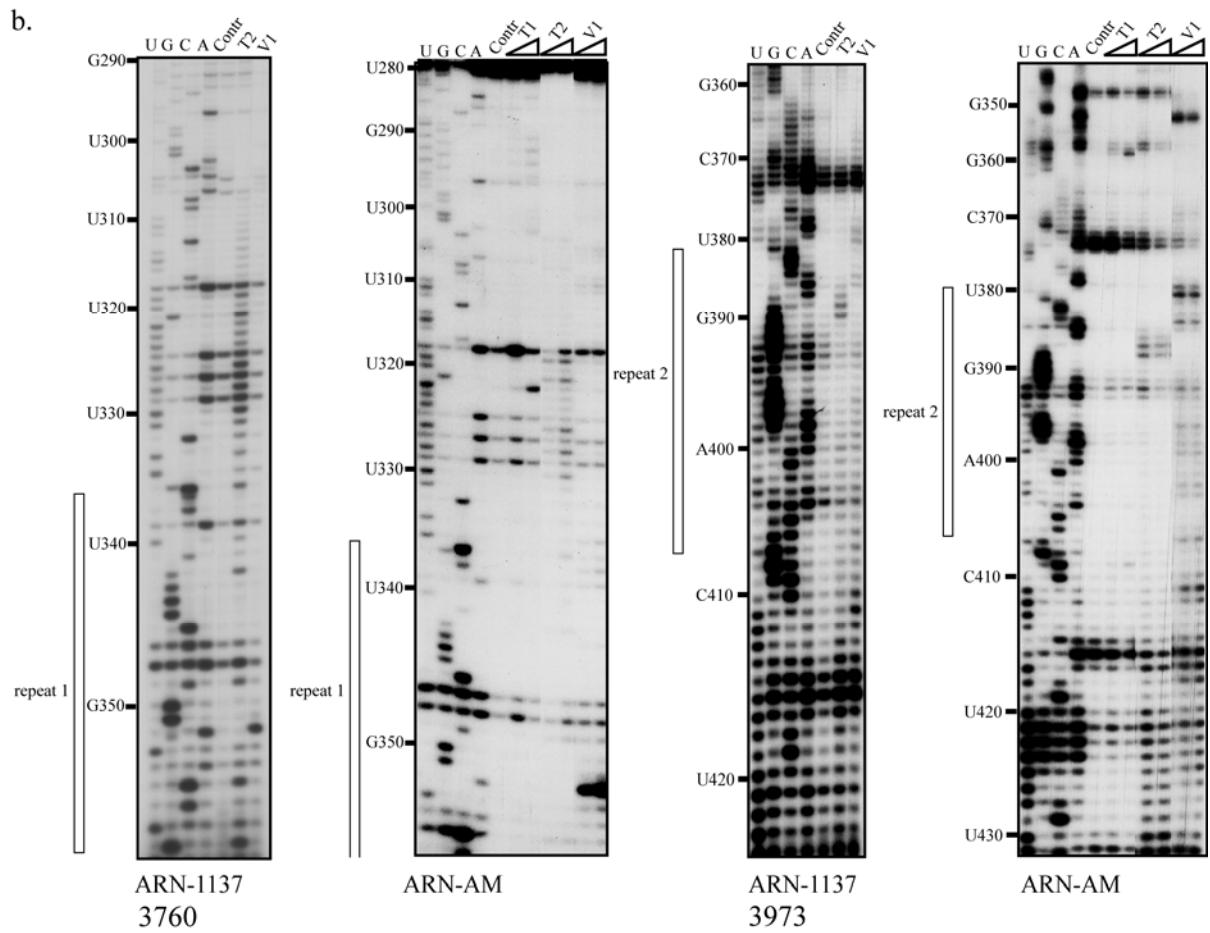
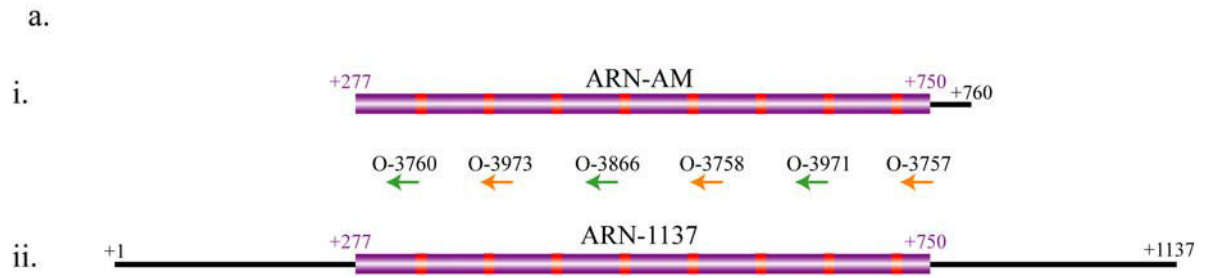
Figure S9

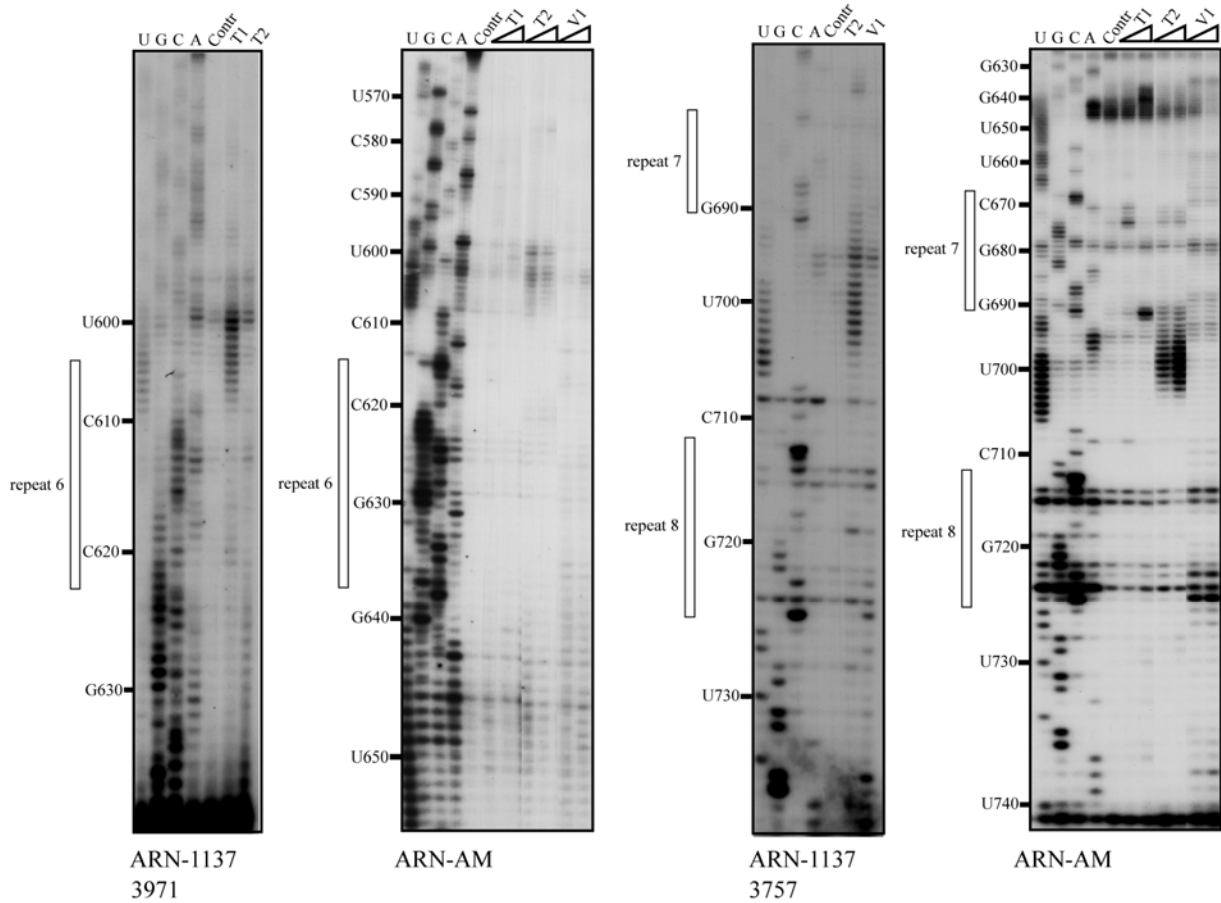
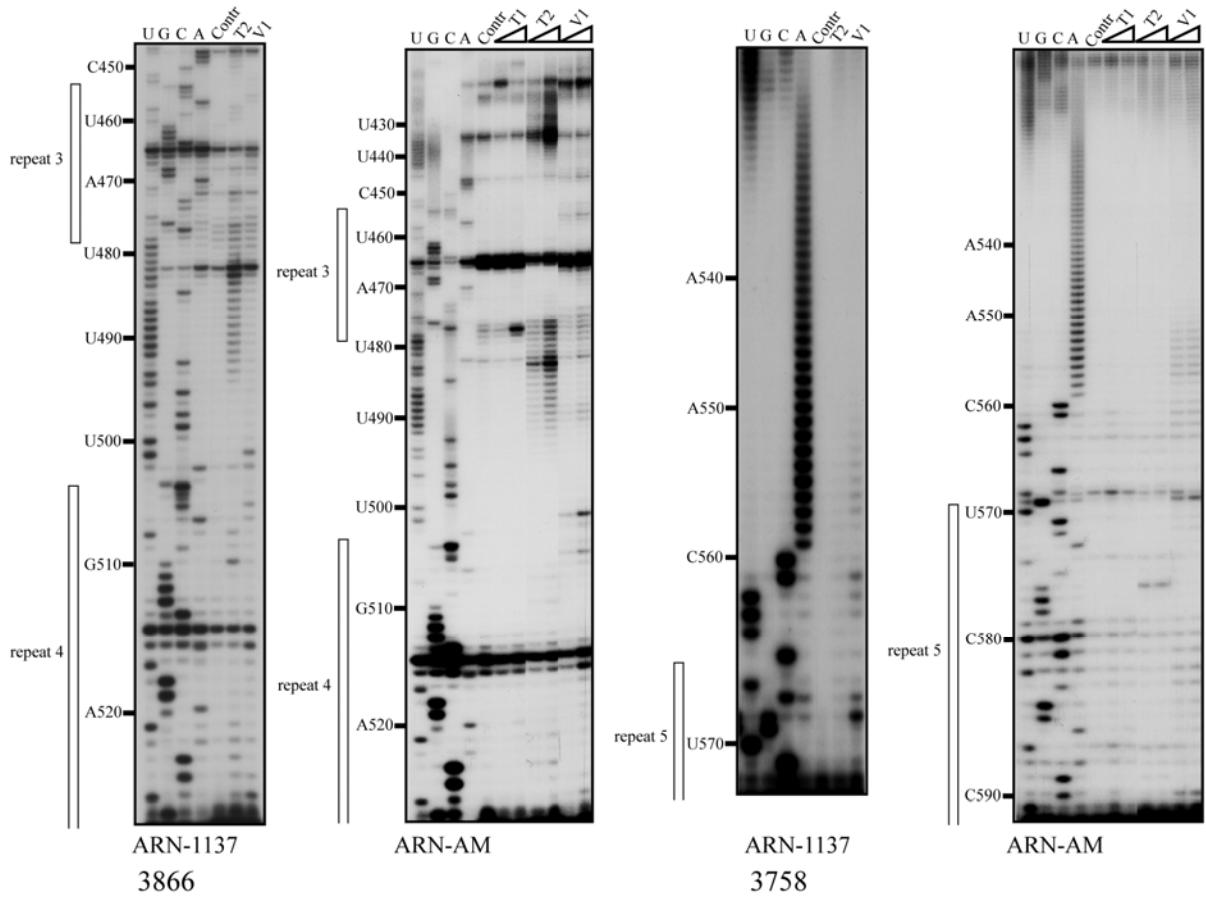
Table I

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	AAGCTTAAATTTAATACGACTCACTATAGGGATCAGTTTTTTACTCTTC	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	AGCCCCGATGGGCAAAAGA	661-679	
O-3973	GATGGGCAAGTTTAGA	444-460	
2D structure of Human A region			
O-4242	ACAACCCACAAAACCAACAT	776-796	
O-4563	AAAAAGAATTTAAAGGCAGG	432-452	
O-4564	GGGTTGAGAGGGAAGGGAA	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	GAAGAAAAAAGAATAAAG	473-496	Cy5
P5	ACAAAAGCAGGTATCCATG	515-534	Cy5
P6	GCAATAACTCACAAAACCAT	741-761	Cy3
P7	GGTTTTTTTTTTTTTTTTT	542-561	Cy5
P3'	AATGAAAAGGCAGGTAAGTA	399-418	Cy5

***Annexe des Chapitres A, B et C***

**Figures supplémentaires non contenues dans l'article  
concernant les analyses structurales**





**Figure 1 annexe : Comparaison des accessibilités aux enzymes des ARN-AM et ARN-1137 lors de l'analyse de la structure 2D de la région des A-repeats murine**

a. Schéma représentant les ARN utilisés pour l'analyse de la structure 2D de la région des A-repeats murine par emploi de sondes enzymatiques et chimiques

i. Région des A-repeats murine appelée ARN-AM

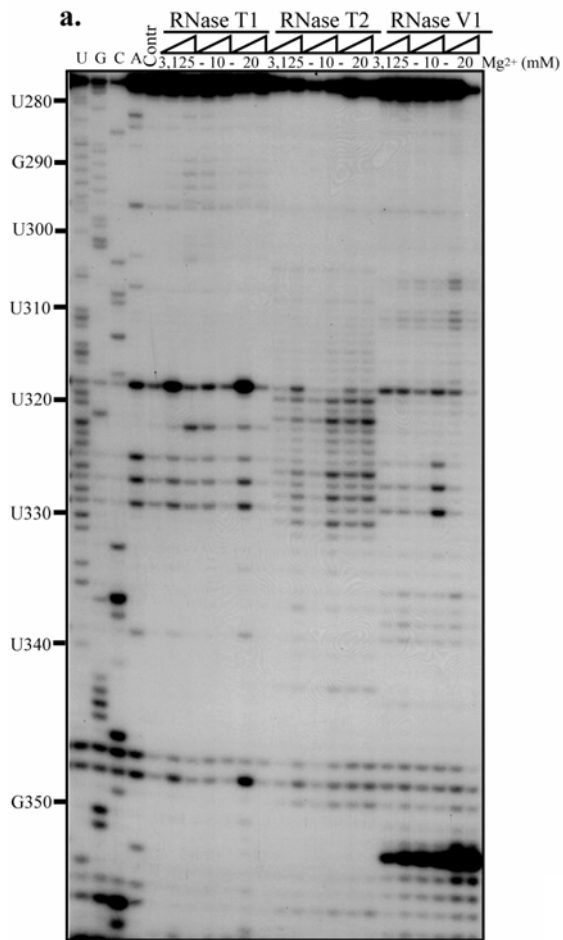
ii. ARN-1137

La région des A-repeats est symbolisée par un rectangle de couleur violette et les répétitions par des rectangles de couleur rouge (ce schéma n'est pas à l'échelle). Les oligonucléotides amorces utilisés pour les expériences d'extension d'amorce sont indiqués.

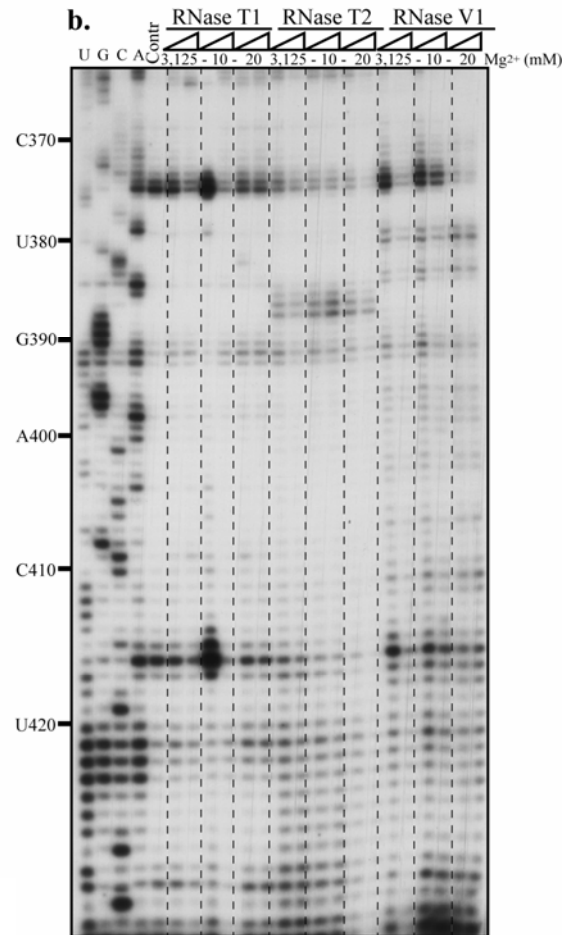
Les positions des nucléotides au sein de l'ARN Xist situés aux extrémités de chacun des fragments analysés sont indiquées.

b. Les ARN-AM et ARN-1137 ont été soumis à des coupures par les RNases T1, T2 et V1 et T1, T2, respectivement. Les conditions utilisées sont décrites dans la partie Matériel et Méthodes. La piste "contr" correspond à un contrôle sans enzyme. Les positions des coupures ont été identifiées par extension de chacun des oligonucléotides amorce radiomarquées en 5'. Les arrêts d'extension ont été positionnés dans la séquence, grâce à un séquençage du transcrit réalisé avec les mêmes oligonucléotides (pistes U, G, C, A). Les ADNc ont été fractionnés par électrophorèse en gel de polyacrylamide dénaturant à 7%. Les positions des nucléotides sont indiquées à gauche des autoradiographies. Les motifs répétés sont indiqués par des rectangles.

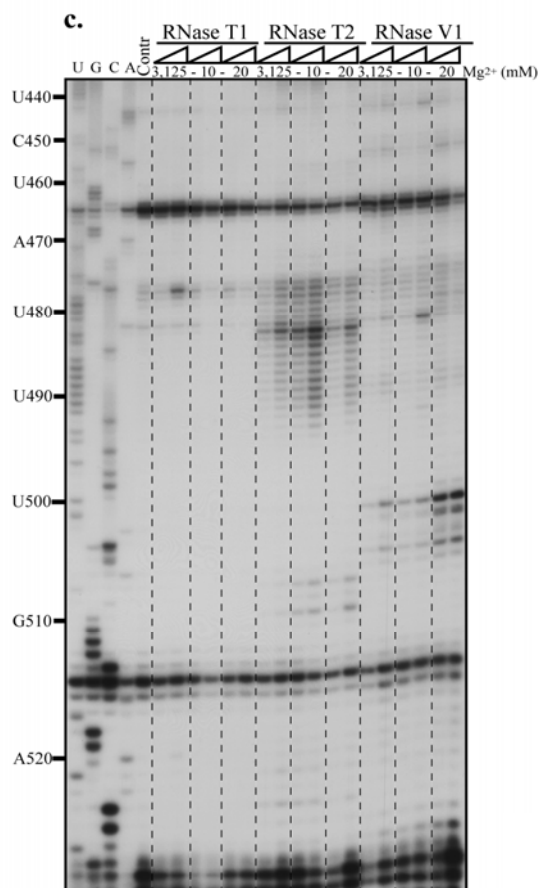
Les résultats obtenus pour une même amorce pour chacun des ARN-AM et ARN-1137 sont présentés en parallèle. Le nom de l'ARN et de l'amorce utilisée sont indiqués sous chaque panneau. Les régions analysables par chaque oligonucléotide sont indiquées dans la **Figure 40**.



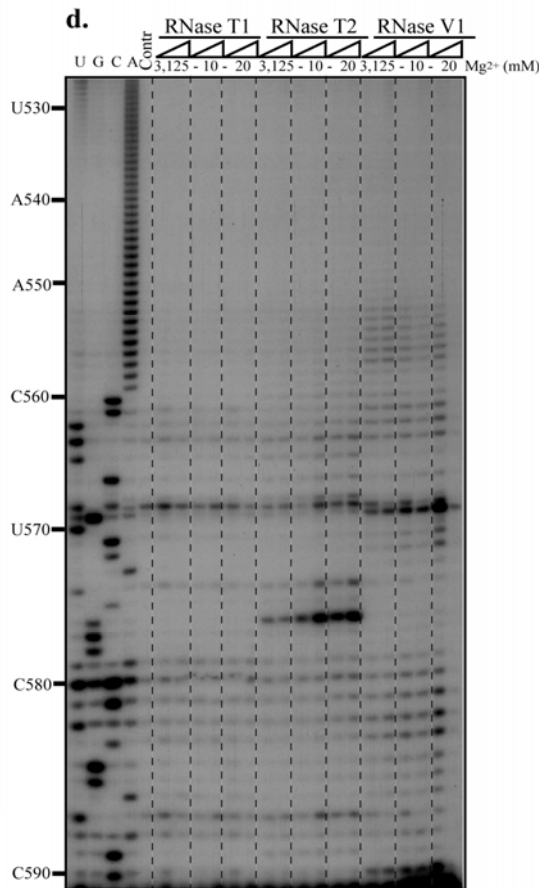
ARN-AM  
3760



ARN-AM  
3973

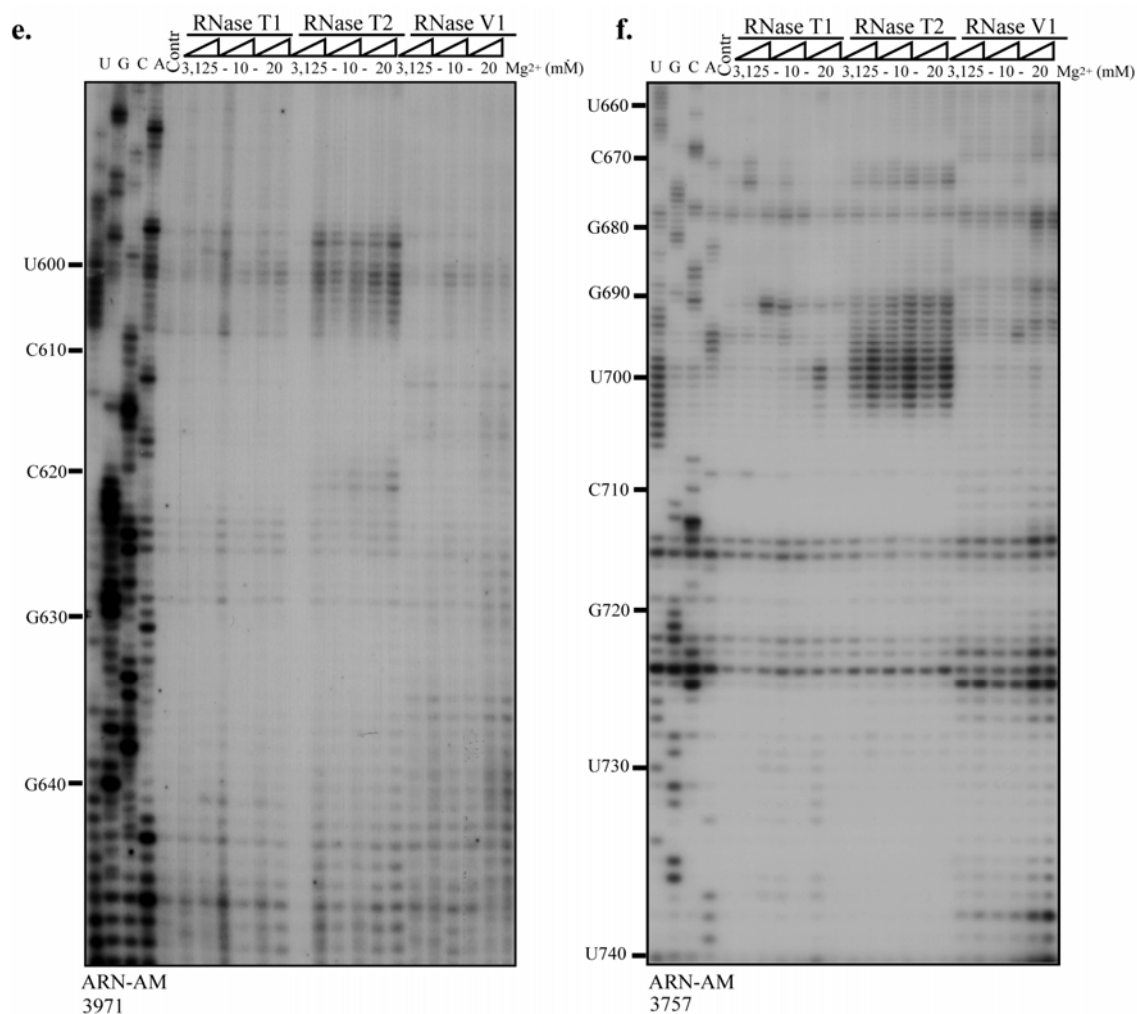


ARN-AM  
3866



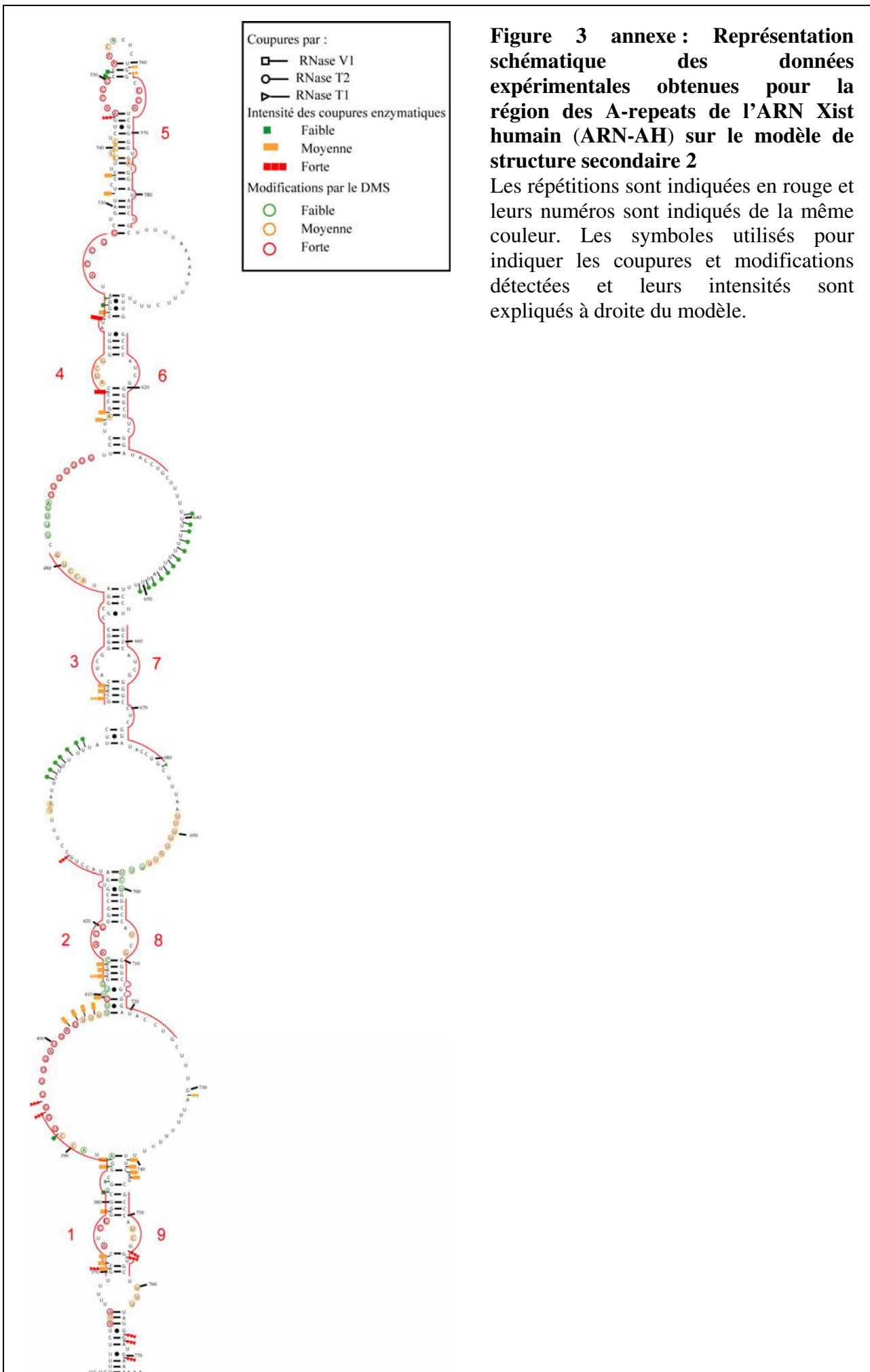
ARN-AM  
3757



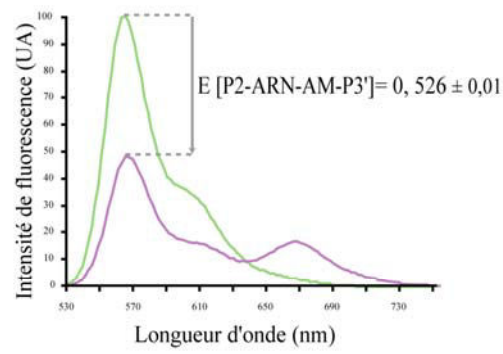
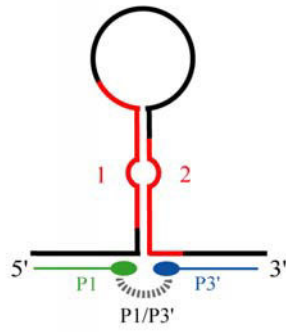


**Figure 2 annexe : Influence de la concentration en Mg<sup>2+</sup> sur l'accessibilité de l'ARN-AM aux sondes enzymatiques**

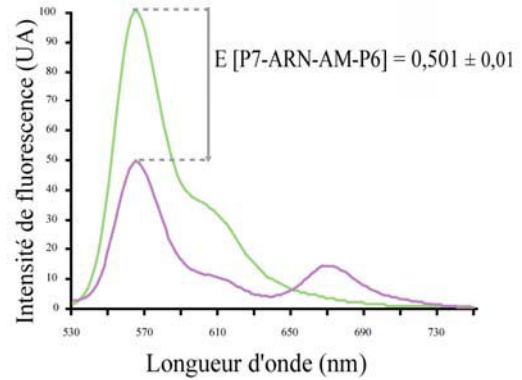
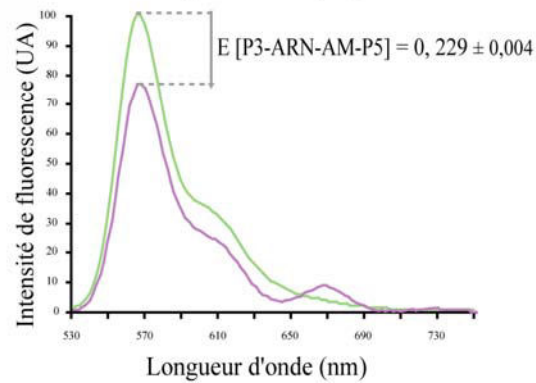
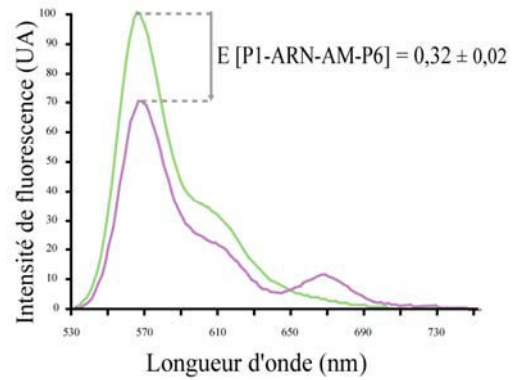
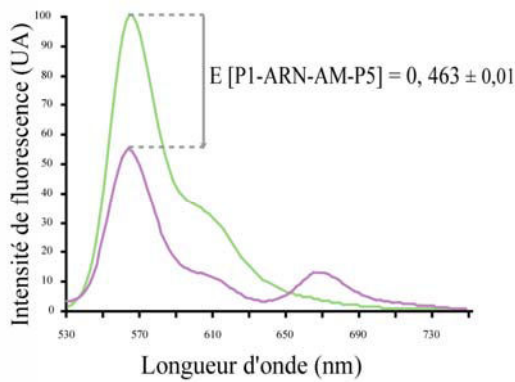
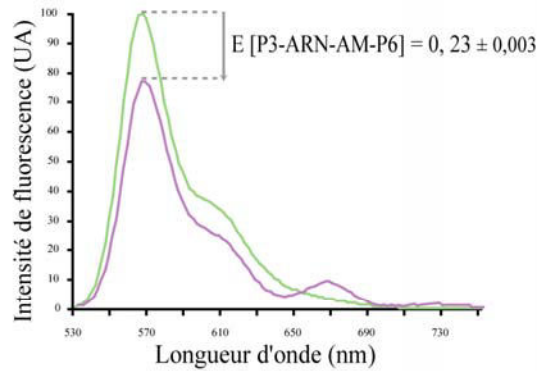
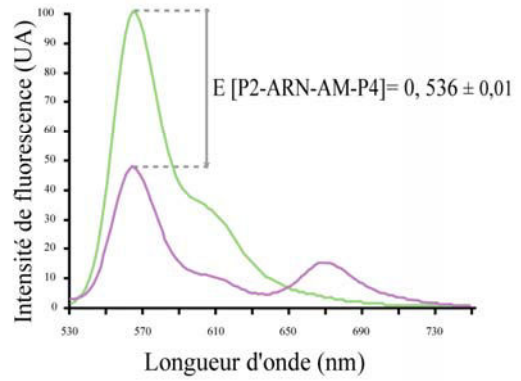
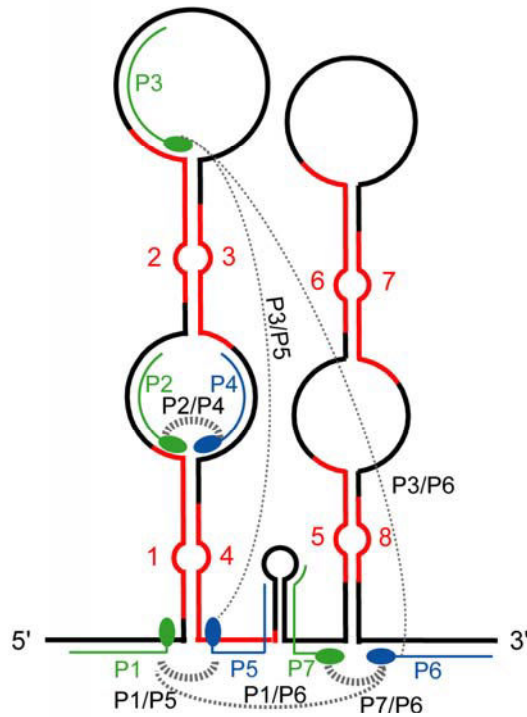
L'ARN-AM a été soumis à des coupures par les RNases T1, T2 et V1. Les conditions utilisées sont décrites dans la partie Matériel et Méthodes. Les analyses ont été effectuées sur l'ARN-AM et 3 concentrations de MgCl<sub>2</sub> (3,125 ; 10 et 20 mM) ont été testées. La piste "Contr" correspond à un contrôle sans enzyme. Les positions des coupures ont été identifiées par extension de chacune des oligonucléotides amorce radiomarqués en 5'. Les arrêts d'extension ont été positionnés dans la séquence, grâce à un séquençage du transcrit réalisé avec les mêmes oligonucléotides (pistes U, G, C, A). Les ADNc ont été fractionnés par électrophorèse en gel de polyacrylamide dénaturant à 7%. Les positions des nucléotides sont indiquées à gauche des autoradiographies. Les répétitions sont indiquées par des rectangles. Le nom de l'ARN et de l'amorce utilisée est indiqué sous chaque panneau. Les régions analysables par chaque oligonucléotide sont indiquées dans la Figure 40. Les panneaux a., b., c., d., e., f. correspondent aux résultats d'extension d'amorces obtenus avec les oligonucléotides 3760, 3973, 3866, 3757, 3971 et 3760, respectivement.



a.



b.

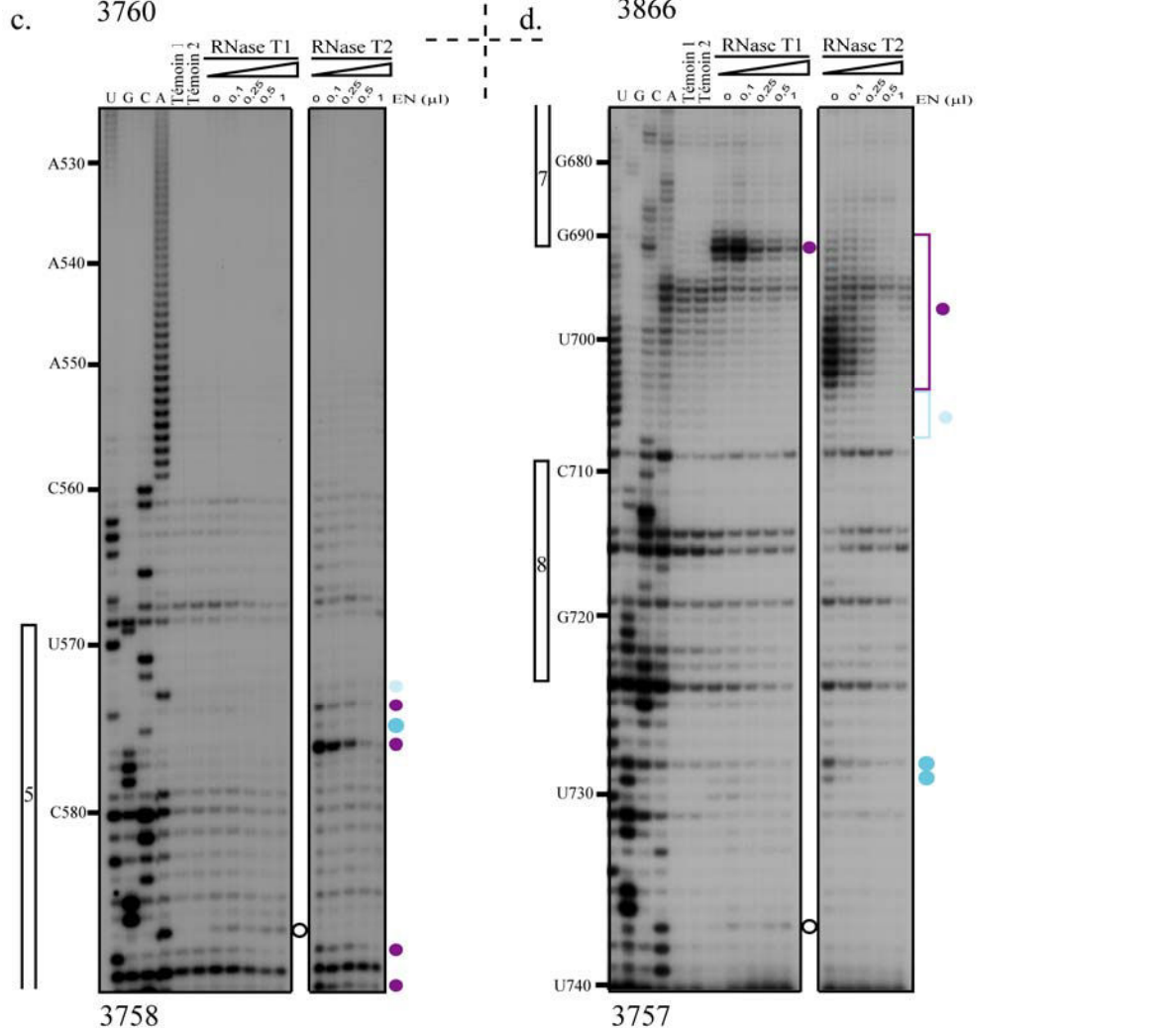
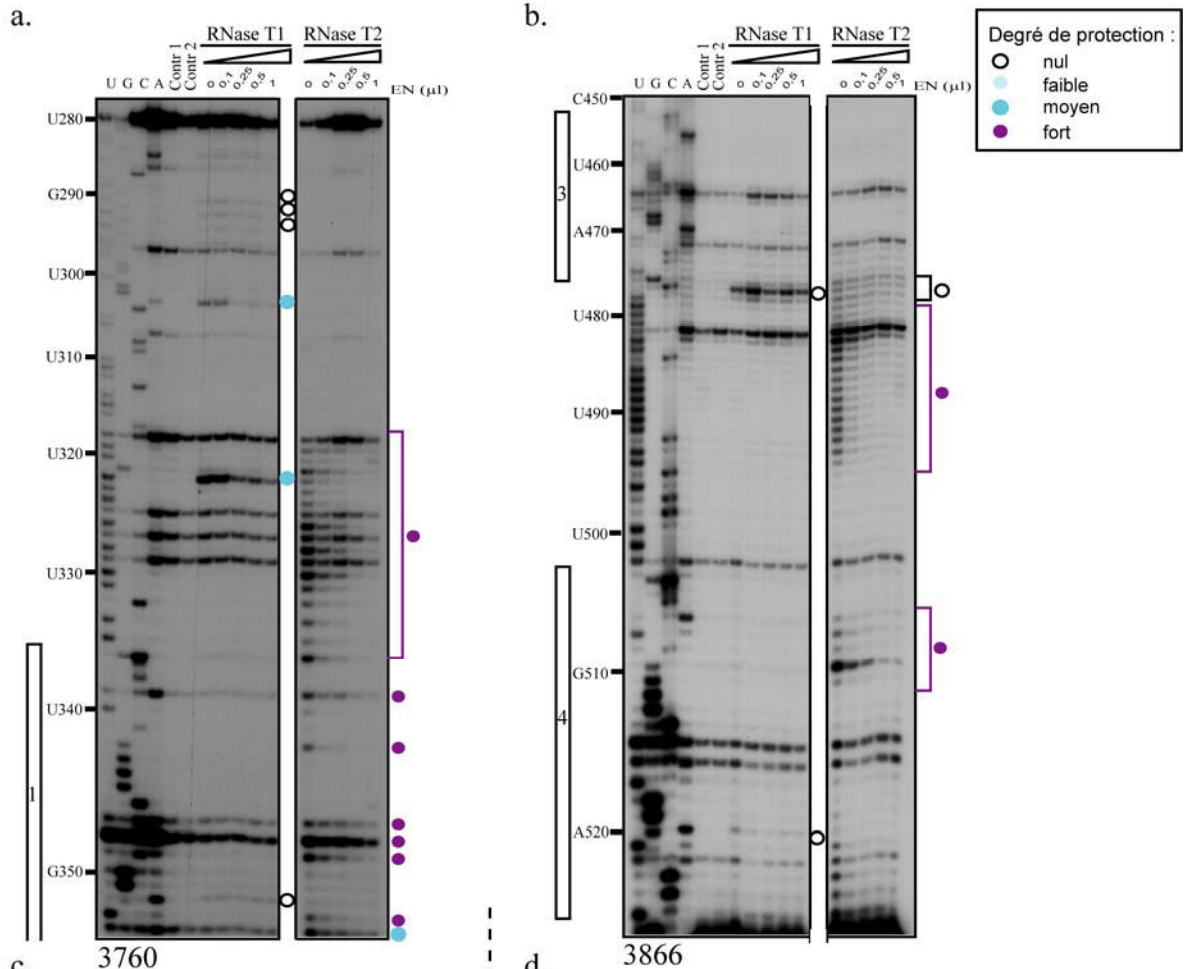


**Figure 4 annexe : Spectres d'émission de fluorescence enregistrés pour chacun des couples d'oligonucléotides utilisés lors des expériences de FRET qui ont permis de valider le modèle 3 de structure 2D de la région des A-repeats murine**

a. Transfert d'énergie obtenu pour le couple d'oligonucléotides (P1-P3') appariés à des séquences simple brin bordant une hélice formée par les répétitions 1 et 2 dans l'ARN-2R.

La structure 2D unique possible pour l'ARN-2R est représentée schématiquement ainsi que l'appariement obtenu avec la paire d'oligonucléotides (P1-P3'). Le spectre d'émission de fluorescence enregistré en présence de l'ARN-2R et de l'oligonucléotide P1 seul est représenté en vert. Celui obtenu en présence des 2 oligonucléotides P1 et P3' est représenté en violet. La diminution d'émission de fluorescence à 564 nm du donneur est indiquée.

b. Transferts d'énergie obtenus avec tous les couples d'oligonucléotides utilisés sur l'ARN de la région des A-repeats entière (ARN-AM). Le code couleur est le même que celui utilisé dans le panneau a.

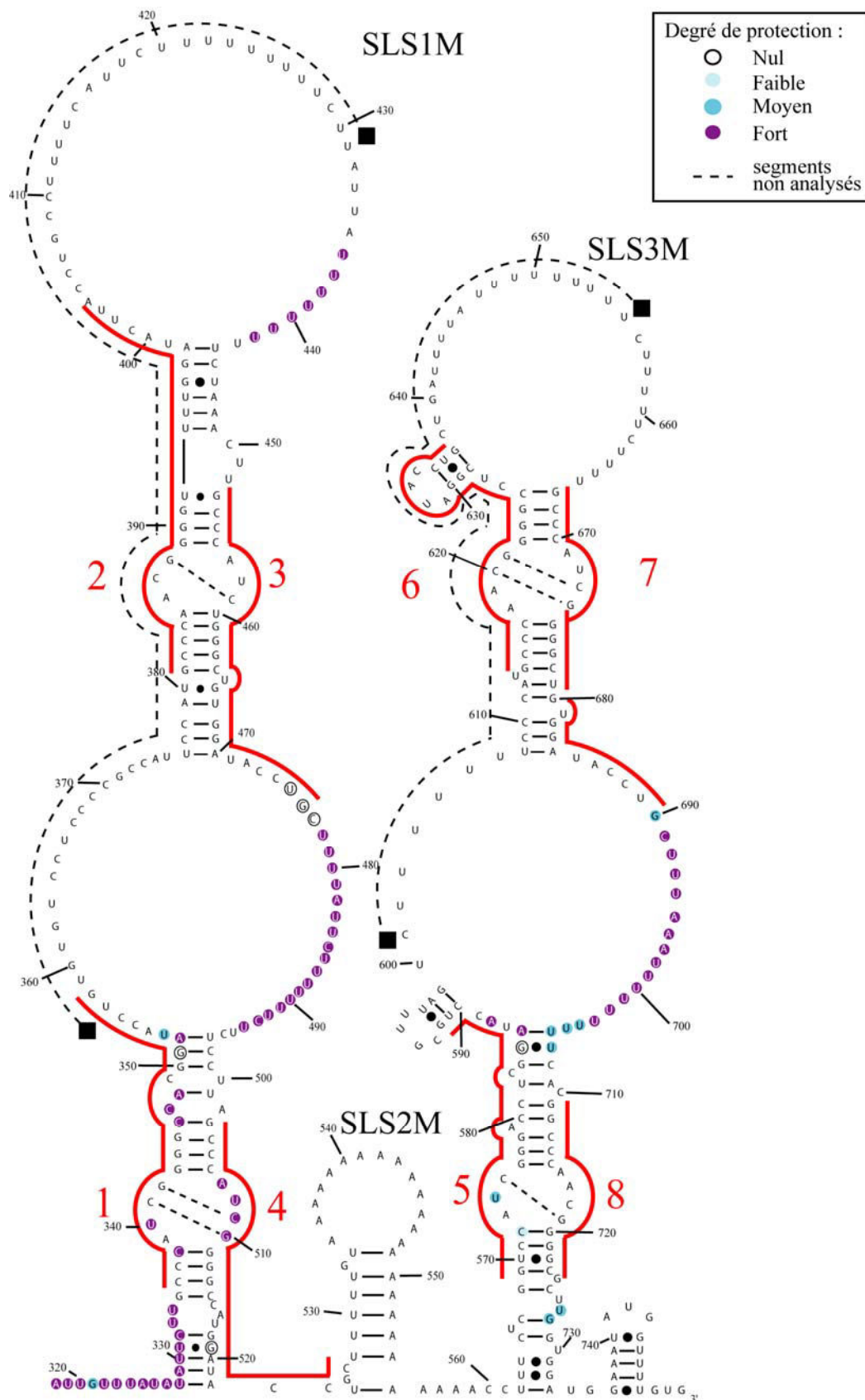


**Figure 5 annexe : Analyse de la protection contre l'action des RNases générée sur l'ARN-AM par fixation des protéines d'un extrait nucléaire de cellules HeLa**

Le transcrit a été soumis à une hydrolyse ménagée par les RNases T1 et T2 (0,5 U/1,2 pmol d'ARN) après incubation en présence ou en absence d'extrait nucléaire de cellule HeLa. Les conditions sont décrites dans la partie Matériel et Méthodes. Les positions des coupures ont été identifiées par extension des oligonucléotides 3760 (a.), 3866 (b.), 3758 (c.) et 3757 (d.) marqués en 5'. Les arrêts d'extension ont été positionnés dans la séquence grâce à des séquençages du transcrit réalisés avec les mêmes oligonucléotides (pistes U, G, C, A). Les ADNc ont été fractionnés par électrophorèse en gel de polyacrylamide dénaturant à 7%. Les positions des nucléotides sont indiquées à gauche de l'autoradiographie. La formation de complexes et les hydrolyses ont été réalisées en présence de  $MgCl_2$  à une concentration de 3,125 mM.

Le contrôle 1 (piste Contr 1) a été réalisé en absence de RNase et en absence d'extrait nucléaire (EN). Le contrôle 2 (Contr 2) a été réalisé en absence de RNase et en présence de 1  $\mu$ l d'extrait nucléaire. Les zones protégées sont indiquées à droite des autoradiographies par des cercles dont l'intensité est proportionnelle au degré de protection.

Les répétitions présentes dans la région des A-repeats sont indiquées par des rectangles à droite des autoradiographies.



**Figure 6 annexe : Représentation des protections contre l'hydrolyse par les RNases T1 et T2 générées par fixation des protéines d'un extrait nucléaire de cellules HeLa sur le modèle de structure 3 de la région des A-repeats**

Les résultats obtenus présentés dans la **Figure 5 annexe**, ont été représentés schématiquement sur le modèle 3 de structure 2D proposé pour l'ARN-AM. Les nucléotides protégés sont indiqués par des cercles et le code couleur est indiqué dans l'encart en haut à droite. Le numéro des répétitions est indiqué en rouge.





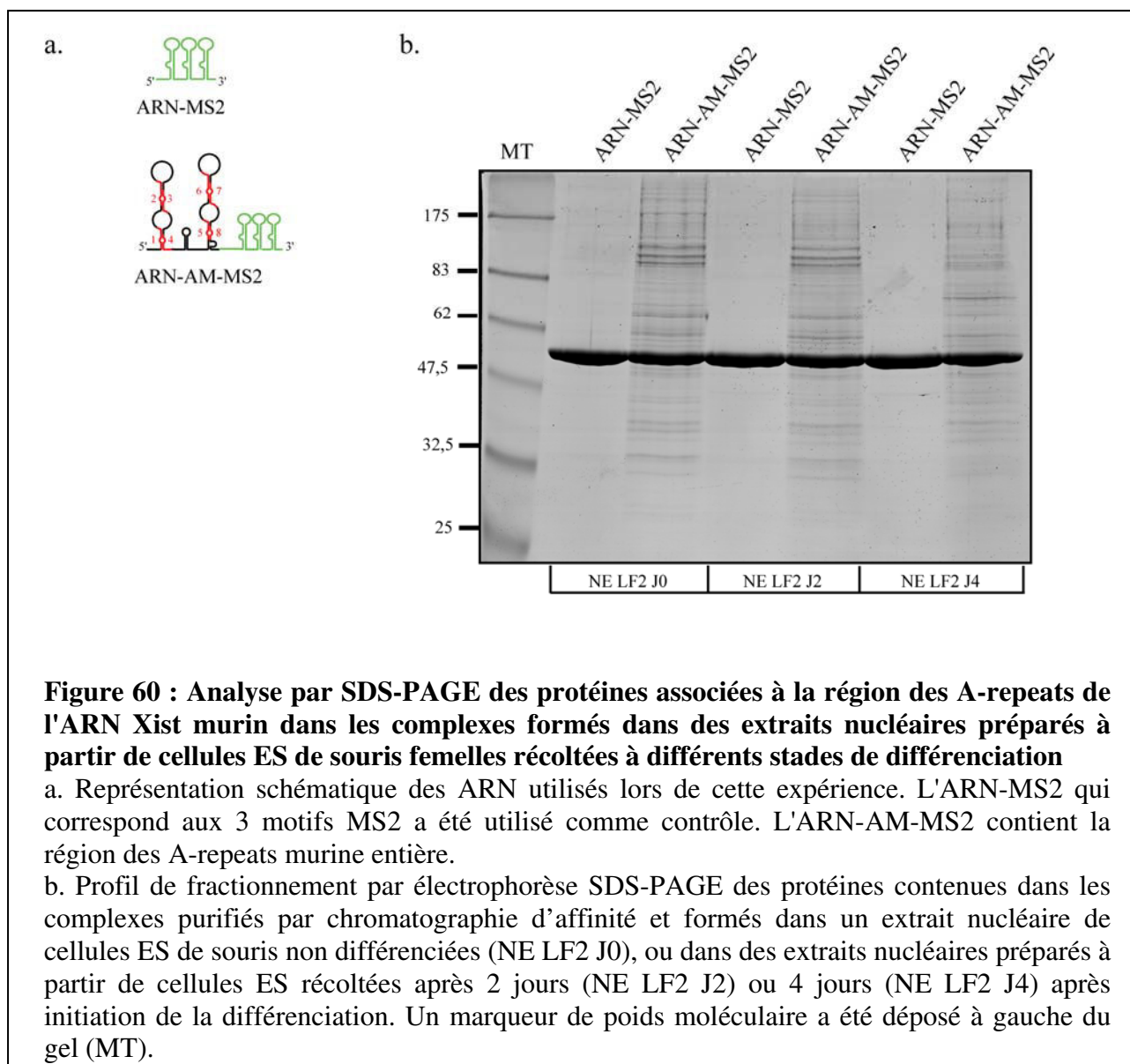


## **D. Bilan des résultats obtenus par purification des RNP formées avec la région des A-repeats et ses sous-fragments**

Dans l'article, nous avons décrit seulement une partie des analyses des RNP formées par incubation de la région des A-repeats en présence d'un extrait nucléaire de cellules ES non différenciées de souris femelles, les analyses complètes sont décrites dans ce chapitre. Par ailleurs, il était intéressant de comparer la composition en protéines des RNP formées avec la région des A-repeats en présence d'extraits nucléaires de cellules ES de souris femelles récoltées à différents temps après le début de la différenciation. En effet, durant cette période ont lieu les différentes étapes conduisant l'inactivation du chromosome X.

### **1. Purification des RNP formées avec la région des A-repeats entière dans des extraits nucléaires de cellules ES de souris femelles prélevées à différents stades de la différenciation**

J'ai préparé dans le laboratoire de P Avner des extraits nucléaires à partir de cellules ES de souris femelles, récoltées 2 jours (J2) et 4 jours (J4) après initiation de leur différenciation, c'est-à-dire après suppression dans le milieu de culture de l'inhibiteur de différenciation appelé LIF. Comme nous l'avons mentionné dans l'introduction, l'initiation de l'inactivation se déroule pendant cette fenêtre de temps. J'ai utilisé le même protocole que celui employé pour les cellules récoltées avant d'initier la différenciation (J0). Pour s'assurer de la reproductibilité des données, les expériences ont été réalisées 2 fois sur des lots d'extraits nucléaires différents et avec des lots d'ARN différents et en parallèle de chacune de ces expériences, les complexes formés avec l'extrait de cellules non différenciées étaient analysés. De manière reproductible, nous avons constaté que les RNP formées avec l'ARN-AM-MS2 (région des A-repeats entière, **Fig. 60a**) contenaient un grand nombre de protéines, ceci quel que soit l'extrait nucléaire utilisé. Néanmoins, d'après le profil de fractionnement, le contenu en protéines des RNP variait selon le stade de différenciation auquel les cellules avaient été récoltées. En particulier, un changement marquant du profil de fractionnement était observé pour l'expérience réalisée avec l'extrait nucléaire J4. Les profils de fractionnement des protéines obtenus avec les extraits nucléaires J0 et J2 étaient par contre très similaires (**Fig. 60b**).



Pour les 2 séries d'expériences réalisées, les pistes des gels SDS-PAGE ont été découpées de façon systématique tous les 2 mm et les protéines contenues dans les bandes ont été analysées par spectrométrie de masse. Ceci a permis d'identifier approximativement 126 protéines distinctes pour les expériences réalisées avec l'extrait nucléaire J0, 56 protéines pour les expériences réalisées avec l'extrait nucléaire J2 et 84 protéines pour les expériences réalisées avec l'extrait nucléaire J4. Ces nombres étaient légèrement variables pour les 2 expériences. Les protéines présentes dans l'expérience contrôle réalisée avec l'ARN-MS2 ne sont pas comptabilisées dans ces chiffres. La grande différence de profil de fractionnement des protéines observée entre les différents extraits nucléaires ne se retrouvait pas au niveau des résultats de spectrométrie de masse. La grande majorité des protéines identifiées avec l'extrait nucléaire J4 était retrouvée avec l'extrait nucléaire J0. Ceci peut peut-être s'expliquer par l'absence de données quantitatives de l'approche de spectrométrie de masse utilisée. En effet, les différences de profils électrophorétiques observées peuvent être dues à des différences de quantités relatives des protéines.

Face au très grand nombre de protéines identifiées dans chacune des analyses systématiques que nous avons réalisées par spectrométrie de masse et face au problème d'absence d'information sur les quantités relatives des protéines identifiées, nous avons décidé de rechercher s'il pouvait exister un lien entre les protéines identifiées et le processus d'inactivation, ceci en nous basant sur les données de la littérature. Comme nous l'avons déjà mentionné, les protéines présentes au sein des listes établies par S Sanglier et L Fouillen lors des expériences de spectrométrie de masse pouvaient soit être associées de manière aspécifique par des interactions ARN-protéines ou protéines-protéines, soit être impliquées dans le processus d'inactivation en :

1. renforçant la stabilité de la structure 2D de la région des A-repeats
2. jouant un rôle de plateforme et en recrutant des protéines fonctionnelles ayant un rôle dans l'inactivation
3. étant impliquées dans l'inactivation, soit en se liant directement à l'ARN, soit en se liant par l'intermédiaire de protéines plateformes

## **2. Stratégie utilisée pour essayer d'identifier les protéines des RNP pouvant avoir une importance fonctionnelle pour l'inactivation du chromosome X**

Nous avons au cours de ce travail de thèse, produit, purifié et analysé de nombreux complexes formés par la région des A-repeats des ARN Xist murin et humain en présence de différents extraits nucléaires. Les expériences réalisées avec l'extrait nucléaire de cellules HeLa devaient nous donner des informations sur les protéines se liant à la région des A-repeats dans des conditions où l'inactivation du chromosome X est établie. Les expériences réalisées avec les extraits nucléaires de cellules ES récoltées à différents stades de différenciation, devaient elles, nous apporter des informations sur les partenaires pouvant se lier à la région des A-repeats dans des conditions où l'inactivation du chromosome X est mise en place. Ainsi, avec les extraits de cellules HeLa, on pouvait espérer avoir parmi les protéines identifiées des protéines impliquées dans la maintenance de l'inactivation et avec les extraits nucléaires de cellules ES, on pouvait tirer des informations sur les protéines impliquées dans sa mise en place.

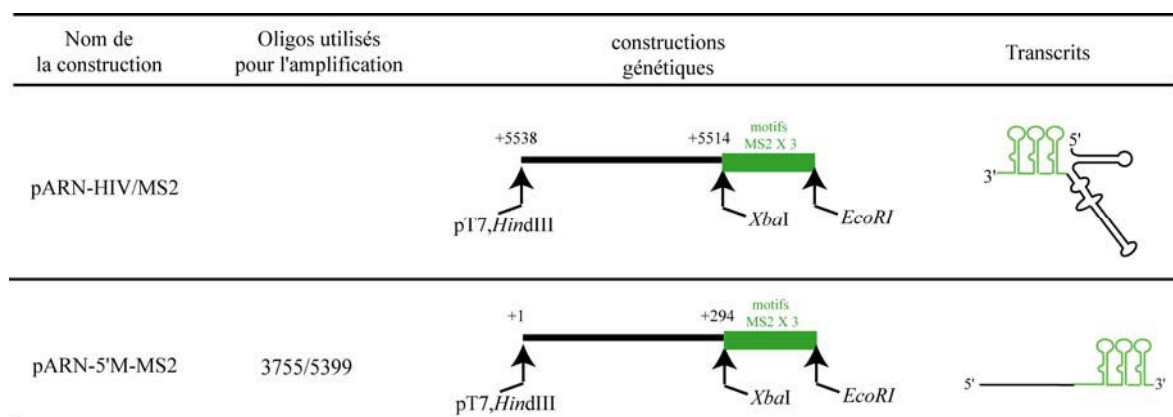
Pour essayer de prédire quelles protéines parmi toutes celles identifiées pouvaient être impliquées dans l'inactivation du chromosome X, il fallait réaliser une étude approfondie de la littérature. Pour cela, les fonctions connues des protéines que nous avons identifiées ont été recherchées dans la littérature. De plus, un défaut d'inactivation du chromosome X aussi bien au niveau de l'inactivation soumise à empreinte parentale (E3,5) qu'au niveau de l'inactivation aléatoire (E6,5) induit des phénotypes létaux (Hoki *et al.*, 2009). L'implication forte d'une protéine dans l'inactivation du chromosome X doit donc conduire à un défaut de développement embryonnaire dans la mesure où ce processus d'inactivation est essentiel au

développement. Aussi, nous avons recherché les données disponibles sur les phénotypes de cellules dans lesquelles les gènes codants pour certaines des protéines identifiées étaient éteints.

Dans l'interprétation des données, nous pouvions aussi comparer les données obtenues pour les différents sous-fragments de la région des A-repeats. En effet, la connaissance des segments de la région des A-repeats requis pour l'association des différentes protéines (**Tab. D**) pouvait apporter des informations sur le mode d'association des protéines et pouvait être prise en compte dans la recherche d'une fonction possible des protéines.

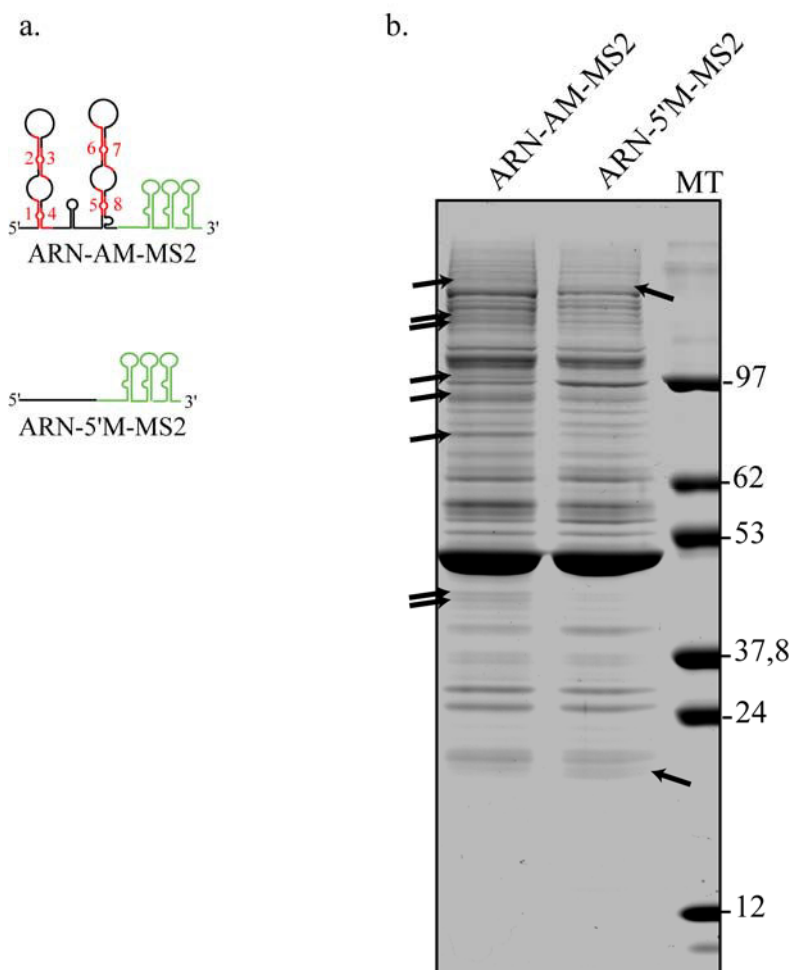
Dans l'interprétation des données, nous avons aussi tenu compte de connaissances acquises au laboratoire du fait que des purifications de RNP, suivies de l'analyse de leur contenu en protéines par spectrométrie de masse par l'équipe d'A Van Dorsselaer ont été réalisées sur un très grand nombre d'ARN différents. Ceci nous permettait de savoir que certaines des protéines que nous avons identifiées sont en fait retrouvées systématiquement quelque soit l'ARN étudié. Il s'agit soit de protéines n'ayant aucune spécificité au niveau de leur interaction avec l'ARN, soit de protéines reconnaissant des motifs généraux trouvés dans tout ARN, tels que des hélices ou des structures tige-boucle. Nous avons donc plus particulièrement focalisé nos études bibliographiques sur des protéines qui n'avaient pas été trouvées par les autres chercheurs du laboratoire, bien que nous ne puissions pas exclure l'implication dans le mécanisme d'inactivation de protéines ayant une faible spécificité de reconnaissance de l'ARN.

Pour avoir une meilleure idée de la spécificité vis-à-vis de la région des A-repeats des protéines que nous avons identifiées, nous avons réalisé en parallèle avec les mêmes extraits, des expériences de purification de RNP formées soit avec la région de l'ARN du VIH (BRU) (positions 5338 à 5514, ARN-HIV/MS2), soit avec la région de l'ARN Xist murin située en amont de la région des A-repeats (ARN-5'M-MS2). Les résultats obtenus avec l'ARN du VIH sont explicités dans l'article. Pour la région 5' de l'ARN Xist, nous avons cloné le fragment d'ADN correspondant à la région 5' de l'ARN Xist de souris compris entre les positions +1 à +294 (ARN-5'M-MS2) dans le vecteur pUC-T7-MS2, selon le protocole décrit dans la partie Matériel et Méthodes. Cette région de Xist a donc été produite sous forme fusionnée avec 3 motifs MS2 (**Fig. 61**).



**Figure 61 : Constructions réalisées pour les expériences de purification par chromatographie d'affinité MS2/MBP des RNP formées avec une région de l'ARN du VIH (positions 5538 à 5514) et la région 5' terminale de l'ARN Xist murin et après incubation en extrait nucléaire de cellules ES non différenciées**

Pour l'ARN-5'M-MS2, les expériences ont été réalisées avec un extrait nucléaire de cellules ES non différenciées et le contenu en protéines a été analysé par spectrométrie de masse. La **Figure 62** présente le fractionnement par SDS-PAGE du contenu en protéines des RNP après purification par chromatographie d'affinité MS2/MBP. Les expériences ont également été réalisées en parallèle avec l'ARN-AM-MS2. Un point marquant était la grande similitude des profils obtenus pour ces 2 régions de l'ARN Xist à l'exception de quelques bandes supplémentaires ou renforcées dans le cas de la région des A-repeats et de 2 bandes renforcées dans le cas de la région 5'.



**Figure 62 : Fractionnement par SDS-PAGE des protéines contenues dans les RNP formées avec la région des A-repeats murine entière et avec la région 5' terminale de l'ARN Xist murin dans des extraits nucléaires de cellules ES non différenciées de souris femelles**

a. Représentation schématique des ARN utilisés lors de cette expérience. L'ARN-AM-MS2 correspond à la région des A-repeats murine entière et l'ARN-5'M-MS2 correspond à la région des positions +1 à +294 de l'ARN Xist.

b. Profil de fractionnement par électrophorèse en SDS-PAGE des protéines contenues dans les complexes après une purification par chromatographie d'affinité MS2/MBP. Un marqueur de poids moléculaire a été déposé à droite du gel (MT).

Les différences observées entre les profils de fractionnement sont indiquées par des flèches.

2.1 Les protéines des RNP qui pourraient être impliquées dans l'activité de la région des A-repeats sur la base des données bibliographiques

2.1.1 Les protéines qui pourraient être intéressantes et qui sont retrouvées avec tous les extraits que nous avons utilisés ou seulement avec l'extrait nucléaire de cellules HeLa

Les protéines identifiées ont été classées par catégories : les protéines hnRNP, les protéines impliquées dans le métabolisme de l'ADN, les protéines impliquées dans le métabolisme de l'ARN, les protéines impliquées dans la transcription et la traduction et enfin une catégorie dans laquelle sont classées les protéines ayant d'autres rôles (**Tab. II, III et**

**IV).** Le **Tableau II** présente la comparaison des résultats d'identification par spectrométrie de masse des protéines contenues dans les RNP formées avec la région des A-repeats humaine (2 expériences avec l'ARN-AM-MS2, **Exp. 1 et Exp. 2, Tab. I annexe**) et les sous-fragments de cette région (**Exp. 3, Tab. II annexe**). Le **Tableau III** présente la comparaison des résultats d'identification par spectrométrie de masse du contenu en protéines des RNP formées avec la région des A-repeats murine (2 expériences, **Exp. 4 et Exp. 5, Tab. III annexe**) avec des extraits nucléaires de cellules ES récoltées à différents stades de différenciation. Le **Tableau IV** présente la comparaison des résultats d'identification par spectrométrie de masse du contenu en protéines des RNP formées avec la région des A-repeats murine, différents sous-fragments de cette région et la région 5' terminale de l'ARN Xist murin avec l'extrait nucléaire de cellules ES non différenciées (**Exp. 6, Tab. IV annexe**).

Avant d'explicitier plus en détail les résultats de spectrométrie de masse, il est important de faire un certain nombre de remarques sur les variations qui apparaissent entre données de plusieurs types d'expériences :

1.) Tout d'abord, le fait que le nombre de protéines détectées varie selon les expériences décrites est dans certains cas lié au fait que l'identification des protéines contenues dans les RNP formées avec la région des A-repeats par spectrométrie de masse a débuté il y a 3 ans en collaboration avec l'équipe d'A Van Dorsselaer. Les méthodes d'analyses ont évolué au cours de ces 3 ans. Lors des premières expériences réalisées avec l'extrait nucléaire de cellules HeLa et la région des A-repeats humaine complète, seules les bandes visualisées après coloration au bleu de Coomassie colloïdale étaient découpées pour les analyses de spectrométrie de masse (**Exp. 1, Tab. I annexe**). Par la suite, les pistes ont été découpées de façon systématique afin de détecter toutes les protéines, même celles présentes en faible quantité (**Exp. 2, 3, 4, 5 et 6, Tab. II, III annexe**). Ceci explique l'augmentation du nombre de protéines détectées. De plus, une méthode d'identification de spectrométrie de masse plus sensible a parfois été utilisée, ce qui explique également l'augmentation du nombre de protéines détectées pour certaines expériences (**Exp. 3, 4, 5 et 6**).

2.) Ensuite, le fait que le nombre des peptides identifiés pour les extraits nucléaires de cellules ES récoltées à différents stades de la différenciation était plus petit que pour un extrait nucléaire de cellules HeLa. Ceci peut être expliqué par une quantité plus faible de matériel analysé. Néanmoins, au sein des 2 expériences réalisées avec un même type d'extrait nucléaire, les nombres relatifs de peptides observés pour les différentes protéines étaient similaires. Pour faciliter la lecture du chapitre, les protéines que nous considérons comme pouvant avoir un rôle dans la fonction de la région des A-repeats sont soulignées dans le texte et surlignées dans les tableaux.



**Tableau II : Comparaison des résultats d'identification par spectrométrie de masse des protéines contenues dans les RNP formées avec la région des A-repeats humaine**

Numéros d'accension	Noms des Protéines	Exp. 1	Exp. 2	Exp. 3			
		ARN-AH-MS2	ARN-AH-MS2	ARN-MS2	ARN-1R-MS2	ARN-2R-MS2	ARN-SLS1H-MS2
<b>hnRNP</b>							
Q13151 ROA0_HUMAN	Heterogeneous nuclear ribonucleoprotein A0 ( <b>hnRNP A0</b> )	3	3			6	6
P09651 ROA1_HUMAN	<b>Heterogeneous nuclear ribonucleoprotein A1</b> (Helix-destabilizing protein)	4	6	7	6	22	24
Q99729 ROAA_HUMAN	Heterogeneous nuclear ribonucleoprotein A/B ( <b>hnRNP A/B</b> )	10	1				13
P22626 ROA2_HUMAN	Heterogeneous nuclear ribonucleoproteins A2/B1 ( <b>hnRNP A2 / hnRNP B1</b> )					6	20
P51991 ROA3_HUMAN	Heterogeneous nuclear ribonucleoprotein A3 ( <b>hnRNP A3</b> )						6
P07910 HNRPC_HUMAN	Heterogeneous nuclear ribonucleoproteins C1/C2 ( <b>hnRNP C1 / hnRNP C2</b> )						5
Q14103 HNRPD_HUMAN	Heterogeneous nuclear ribonucleoprotein D0 ( <b>hnRNP D0</b> )	29	18		2	13	31
Q15365 PCBP1_HUMAN	Poly(rC)-binding protein 1 (Alpha-CP1) ( <b>hnRNP-E1</b> )	4					13
Q15366 PCBP2_HUMAN	Poly(rC)-binding protein 2 (Alpha-CP2) ( <b>hnRNP-E2</b> )					9	22
P38159 HNRPG_HUMAN	Heterogeneous nuclear ribonucleoprotein G ( <b>hnRNP G</b> )						4
P61978 HNRPK_HUMAN	Heterogeneous nuclear ribonucleoprotein K ( <b>hnRNP K</b> )					15	41
P14866 HNRPL_HUMAN	Heterogeneous nuclear ribonucleoprotein L ( <b>hnRNP L</b> )						6
Q8WVV9 HNRLH_HUMAN	<b>Heterogeneous nuclear ribonucleoprotein L-like</b> (Stromal RNA-regulating factor)						10
O43390 HNRPR_HUMAN	Heterogeneous nuclear ribonucleoprotein R ( <b>hnRNP R</b> )	8	6	2		9	23
O60506 HNRPQ_HUMAN	Heterogeneous nuclear ribonucleoprotein Q ( <b>hnRNP Q</b> )	17	22			14	26
Q00839 HNRPU_HUMAN	Heterogeneous nuclear ribonucleoprotein U ( <b>hnRNP U</b> )	26	9	9	6	31	53
P26599 PTBP1_HUMAN	<b>Polypyrimidine tract-binding protein 1</b> (PTB)	61	51			25	51
Q61990 PCBP2_MOUSE	<b>Poly(rC)-binding protein 2</b> (Alpha-CP2)						7
Q99020 ROAA_MOUSE	Heterogeneous nuclear ribonucleoprotein A/B (hnRNP A/B) (CARG-binding factor-A)					3	14
Q96AE4 FUBP1_HUMAN	Far upstream element-binding protein 1 ( <b>FUSE-binding protein 1</b> ) (FBP)	32	11				31
Q92945 FUBP2_HUMAN	Far upstream element-binding protein 2 ( <b>FUSE-binding protein 2</b> )	45	19				54
<b>Métabolisme de l'ADN</b>							
Q92922 SMRC1_HUMAN	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfam1						7
O60264 SMCA5_HUMAN	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfam1	5					19

Q8TAQ2 SMRC2_HUMAN	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamI						7
P12956 KU70_HUMAN	<b>ATP-dependent DNA helicase 2 subunit 1</b>	6		11			17
P13010 KU86_HUMAN	<b>ATP-dependent DNA helicase 2 subunit 2</b>	4		10			21
P11387 TOP1_HUMAN	<b>DNA topoisomerase 1</b> (EC 5.99.1.2) (DNA topoisomerase I)		5				17
P11388 TOP2A_HUMAN	<b>DNA topoisomerase 2-alpha</b> (EC 5.99.1.3) (DNA topoisomerase II, alpha isozyme)			3			9
Q9Y5B9 SPT16_HUMAN	<b>FACT complex subunit SPT16</b>						4
Q08945 SSRP1_HUMAN	<b>FACT complex subunit SSRP1</b>						4
O00139 KIF2A_HUMAN	<b>Kinesin-like protein KIF2A</b> (Kinesin-2) (HK2)	5					5
Q99661 KIF2C_HUMAN	<b>Kinesin-like protein KIF2C</b> (Mitotic centromere-associated kinesin) (MCAK)	5					7
P07355 ANXA2_HUMAN	<b>Annexin A2</b> (Annexin II) (Lipocortin II) (Calpactin I heavy chain)						6
P26358 DNMT1_HUMAN	<b>DNA (cytosine-5)-methyltransferase 1</b> (EC 2.1.1.37) (Dnmt1)						4
P20585 MSH3_HUMAN	<b>DNA mismatch repair protein Msh3</b> (Divergent upstream protein) (DUP)						4
P78527 PRKDC_HUMAN	<b>DNA-dependent protein kinase catalytic subunit</b> (EC 2.7.11.1)	57	3	26			32
P29372 3MG_HUMAN	<b>DNA-3-methyladenine glycosylase</b> (EC 3.2.2.21) (3-methyladenine DNA glycosidase)						5
P49916 DNL3_HUMAN	<b>DNA ligase 3</b> (EC 6.5.1.1) (DNA ligase III)		8				8
Q13112 CAF1B_HUMAN	<b>Chromatin assembly factor 1 subunit B</b> (CAF-1 subunit B)						2
Q14683 SMC1A_HUMAN	<b>Structural maintenance of chromosomes protein 1A</b> (SMC1alpha protein) (Sb1.8)	4					14
Q14839 CHD4_HUMAN	<b>Chromodomain helicase-DNA-binding protein 4</b> (EC 3.6.1.-)	4					22
Q16531 DDB1_HUMAN	<b>DNA damage-binding protein 1</b> (Damage-specific DNA-binding protein 1)	6					5
Q86U86 PB1_HUMAN	<b>Protein polybromo-1</b> (hPB1) (Polybromo-1D) (BRG1-associated factor 180) (BAF180)						13
Q9NRL2 BAZ1A_HUMAN	<b>Bromodomain adjacent to zinc finger domain protein 1A</b>						11
Q9UQE7 SMC3_HUMAN	<b>Structural maintenance of chromosomes protein 3</b>						15
Q13263 TIF1B_HUMAN	<b>Transcription intermediary factor 1-beta</b> (TIF1-beta)	3					3

Métabolisme de l'ARN							
P17844 DDX5_HUMAN	<b>Probable ATP-dependent RNA helicase DDX5</b> (EC 3.6.1.-) (DEAD box protein 5)	1	7			3	13
Q92841 DDX17_HUMAN	<b>Probable ATP-dependent RNA helicase DDX17</b> (EC 3.6.1.-) (DEAD box protein 17)						4
Q9H0S4 DDX47_HUMAN	<b>Probable ATP-dependent RNA helicase DDX47</b> (EC 3.6.1.-) (DEAD box protein 47)						4
Q92499 DDX1_HUMAN	<b>ATP-dependent RNA helicase DDX1</b> (EC 3.6.1.-) (DEAD box protein 1)	13	8				30
O00571 DDX3X_HUMAN	<b>ATP-dependent RNA helicase DDX3X</b> (EC 3.6.1.-)						12
Q08211 DHX9_HUMAN	<b>ATP-dependent RNA helicase A</b> (EC 3.6.1.-) (Nuclear DNA helicase II) (NDH II)	7	22				8
Q13868 EXOS2_HUMAN	<b>Exosome complex exonuclease RRP4</b> (EC 3.1.13.-)						3
Q9NQ75 EXOS3_HUMAN	<b>Exosome complex exonuclease RRP40</b> (EC 3.1.13.-)						3
Q9NPD3 EXOS4_HUMAN	<b>Exosome complex exonuclease RRP41</b> (EC 3.1.13.-)						5
Q9Y2L1 RRP44_HUMAN	<b>Exosome complex exonuclease RRP44</b> (EC 3.1.13.-)						3
Q9NQ74 EXOS5_HUMAN	<b>Exosome complex exonuclease RRP46</b> (EC 3.1.13.-)						2
Q01780 EXOSX_HUMAN	<b>Exosome component 10</b> (Polymyositis/scleroderma autoantigen 2)	6	5				9
P09012 SNRPA_HUMAN	<b>U1 small nuclear ribonucleoprotein A (U1 snRNP protein A)</b>		5			6	3
P09661 RU2A_HUMAN	<b>U2 small nuclear ribonucleoprotein A' (U2 snRNP-A')</b>		3			5	9
Q43395 PRPF3_HUMAN	<b>U4/U6 small nuclear ribonucleoprotein Prp3</b> (Pre-mRNA-splicing factor 3)						7
O75643 U520_HUMAN	<b>U5 small nuclear ribonucleoprotein 200 kDa helicase</b> (EC 3.6.1.-)						3
P08621 RU17_HUMAN U1	<b>small nuclear ribonucleoprotein 70 kDa (U1 snRNP 70 kDa)</b> (snRNP70)						5
P14678 RSMB_HUMAN ou P63162 RSMN_HUMAN	<b>Small nuclear ribonucleoprotein-associated proteins B and B' (snRNP-B)</b> ou <b>Small nuclear ribonucleoprotein-associated protein N (snRNP-N)</b>		2			3	8
O75533 SF3B1_HUMAN	<b>Splicing factor 3B subunit 1</b> (Spliceosome-associated protein 155) (SAP 155)		5				13
Q13435 SF3B2_HUMAN	<b>Splicing factor 3B subunit 2</b> (Spliceosome-associated protein 145) (SAP 145)						11
Q15393 SF3B3_HUMAN	<b>Splicing factor 3B subunit 3</b> (Spliceosome-associated protein 130) (SAP 130)					3	6
Q07955 SFRS1_HUMAN	<b>Splicing factor, arginine/serine-rich 1</b>					4	6
P84103 SFRS3_HUMAN	<b>Splicing factor, arginine/serine-rich 3</b> (Pre-mRNA-splicing factor SRP20)						8

Q01081 U2AF1_HUMAN	<b>Splicing factor U2AF 35 kDa subunit</b> (U2 auxiliary factor 35 kDa subunit)	14	9			13	15
P26368 U2AF2_HUMAN	<b>Splicing factor U2AF 65 kDa subunit</b> (U2 auxiliary factor 65 kDa subunit)	12	20			13	22
Q15637 SF01_HUMAN	<b>Splicing factor 1</b> (Zinc finger protein 162) (Transcription factor ZFM1)	16	12			8	12
P23246 SFPQ_HUMAN	<b>Splicing factor, proline- and glutamine-rich</b>	5	8			6	12
<b>Transcription/Traduction</b>							
P68104 EF1A1_HUMAN	<b>Elongation factor 1-alpha 1</b> (EF-1-alpha-1) (Elongation factor 1 A-1) (eEF1A-1)	35	12		13	11	12
P24534 EF1B_HUMAN	<b>Elongation factor 1-beta</b> (EF-1-beta)	4				2	
P29692 EF1D_HUMAN	<b>Elongation factor 1-delta</b> (EF-1-delta) (Antigen NY-CO-4)	19					6
P26641 EF1G_HUMAN	<b>Elongation factor 1-gamma</b> (EF-1-gamma) (eEF-1B gamma)	28		3		1	16
P57772 SELB_HUMAN	<b>Selenocysteine-specific elongation factor</b> (Elongation factor sec)						4
P16017 EF1A_CANAL	Elongation factor 1-alpha (EF-1-alpha) - Candida albicans (Yeast)						3
P13549 EF1A0_XENLA	Elongation factor 1-alpha, somatic form (EF-1-alpha-S)						7
P05198 IF2A_HUMAN	<b>Eukaryotic translation initiation factor 2 subunit 1</b>	8					11
P20042 IF2B_HUMAN	<b>Eukaryotic translation initiation factor 2 subunit 2</b>						4
P41091 IF2G_HUMAN	<b>Eukaryotic translation initiation factor 2 subunit 3</b>						4
Q14152 IF3A_HUMAN	<b>Eukaryotic translation initiation factor 3 subunit 10</b> (eIF-3 theta)						4
O60841 IF2P_HUMAN	<b>Eukaryotic translation initiation factor 5B</b> (eIF-5B)						37
<b>Autres</b>							
P46782 RS5_HUMAN	<b>40S ribosomal protein S5</b>						2
Q12905 ILF2_HUMAN	<b>Interleukin enhancer-binding factor 2</b>	9	17	5	5	13	20
Q12906 ILF3_HUMAN	<b>Interleukin enhancer-binding factor 3</b>	15	12	24	8	16	31
O43809 CPSF5_HUMAN	<b>Cleavage and polyadenylation specificity factor 5</b>	9				6	15
Q16630 CPSF6_HUMAN	<b>Cleavage and polyadenylation specificity factor 6</b>	6	4				4
Q8N684 CPSF7_HUMAN	<b>Cleavage and polyadenylation specificity factor 7</b>	4	5				8
Q8IX12 CCAR1_HUMAN	<b>Cell division cycle and apoptosis regulator protein 1</b>						13
Q9NYY4 CD2L7_HUMAN	<b>Cell division cycle 2-related protein kinase 7</b> (EC 2.7.11.22)						5

Q9UJX2 CDC23_HUMAN	<b>Cell division cycle protein 23 homolog</b> (Anaphase-promoting complex subunit 8)					9
O43143 DHX15_HUMAN	<b>Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15</b> (EC 3.6.1.-)		7			5
O60231 DHX16_HUMAN	<b>Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX16</b> (EC 3.6.1.-)					7
Q9NSD9 SYFB_HUMAN	<b>Phenylalanyl-tRNA synthetase beta chain</b> (EC 6.1.1.20)	4		2		
P35249 RFC4_HUMAN	<b>Replication factor C subunit 4</b> (Replication factor C 37 kDa subunit)					4
P78347 GTF2I_HUMAN	<b>General transcription factor II-I</b> (GTFII-I) (TFII-I)					7
Q12789 TF3C1_HUMAN	<b>General transcription factor 3C polypeptide 1</b>					15
O75909 CCNK_HUMAN	<b>Cyclin-K</b>					5
P12004 PCNA_HUMAN	<b>Proliferating cell nuclear antigen</b> (PCNA) (Cyclin)					3
O43670 ZN207_HUMAN	<b>Zinc finger protein 207</b>					3
O43684 BUB3_HUMAN	<b>Mitotic checkpoint protein BUB3</b>	5				11
O43913 ORC5_HUMAN	<b>Origin recognition complex subunit 5</b>					5
O75475 PSIP1_HUMAN	<b>PC4 and SFRS1-interacting protein</b> (Lens epithelium-derived growth factor)					6
O75607 NPM3_HUMAN	<b>Nucleoplasmin-3</b>					4
O94776 MTA2_HUMAN	<b>Metastasis-associated protein MTA2</b> (Metastasis-associated 1-like 1)	5				11
O95232 CROP_HUMAN	<b>Cisplatin resistance-associated overexpressed protein</b>					3
O95602 RPA1_HUMAN	<b>DNA-directed RNA polymerase I largest subunit</b> (EC 2.7.7.6)					5
P05455 LA_HUMAN	<b>Lupus La protein</b> (Sjogren syndrome type B antigen) (SS-B) (La ribonucleoprotein)		2			5
P06748 NPM_HUMAN	<b>Nucleophosmin</b> (NPM) (Nucleolar phosphoprotein B23) (Numatrin)	11	25		9	18
P09874 PARP1_HUMAN	<b>Parp 1</b>	70	32	31	14	89
P11021 GRP78_HUMAN	<b>78 kDa glucose-regulated protein precursor</b> (GRP 78)	9				13
P11142 HSP7C_HUMAN	<b>Heat shock cognate 71 kDa protein</b> (Heat shock 70 kDa protein 8)	16	6	4		10
P11940 PABP1_HUMAN	<b>Polyadenylate-binding protein 1</b> (Poly(A)-binding protein 1) (PABP 1)	16	6			5
P17480 UBF1_HUMAN	<b>Nucleolar transcription factor 1</b> (Upstream-binding factor 1) (UBF-1)		4			4
P19338 NUCL_HUMAN	<b>Nucleolin</b> (Protein C23)	14	23		15	9
P19525 E2AK2_HUMAN	<b>Interferon-induced, double-stranded RNA-activated protein kinase</b>	2	5	7		8

P23919 DTYMK_HUMAN	<b>Thymidylate kinase</b> (EC 2.7.4.9) (dTMP kinase)					3	5
P26640 SYV_HUMAN	<b>Valyl-tRNA synthetase</b> (EC 6.1.1.9) (Valine-tRNA ligase) (ValRS) (Protein G7a)	20					4
P35269 T2FA_HUMAN	<b>Transcription initiation factor IIF subunit alpha</b> (EC 2.7.11.1) (TFIIF-alpha)						3
P35637 FUS_HUMAN	<b>RNA-binding protein FUS</b> (Oncogene FUS) (Oncogene TLS)						6
P35659 DEK_HUMAN	<b>Protein DEK</b>						11
P40108 ALDH_CLAHE	<b>Aldehyde dehydrogenase</b> (EC 1.2.1.3) (ALDDH) (Allergen Cla h 10) (Cla h 3)					10	
P42285 SK2L2_HUMAN	<b>Superkiller viralicidic activity 2-like 2</b> (EC 3.6.1.-)	3	2				17
P42704 LPPRC_HUMAN	<b>130 kDa leucine-rich protein</b> (LRP 130) (GP130)	7					21
P43243 MATR3_HUMAN	<b>Matrin-3</b>	2	7				15
P48382 RFX5_HUMAN	<b>DNA-binding protein RFX5</b> (Regulatory factor X subunit 5)						3
P51532 SMCA4_HUMAN	<b>Probable global transcription activator SNF2L4</b> (EC 3.6.1.-)						20
P62258 1433E_HUMAN	<b>14-3-3 protein epsilon</b> (14-3-3E)						3
P62826 RAN_HUMAN	<b>GTP-binding nuclear protein Ran</b> (GTPase Ran) (Ras-like protein TC4)				6	7	6
P67809 YBOX1_HUMAN	<b>Nuclease sensitive element-binding protein 1</b> (Y-box-binding protein 1)						3
P68400 CSK21_HUMAN	<b>Casein kinase II subunit alpha</b> (EC 2.7.11.1) (CK II)						3
P78346 RPP30_HUMAN	<b>Ribonuclease P protein subunit p30</b> (EC 3.1.26.5) (RNaseP protein p30)						3
Q00341 VIGLN_HUMAN	<b>Vigilin</b> (High density lipoprotein-binding protein) (HDL-binding protein)	5					26
Q01085 TIAR_HUMAN	<b>Nucleolysin TIAR</b> (TIA-1-related protein)	4	3				14
Q07666 SAM68_HUMAN	<b>KH domain-containing, RNA-binding, signal transduction-associated protein 1</b> (p21)						5
Q13112 CAF1B_HUMAN	<b>Chromatin assembly factor 1 subunit B</b> (CAF-1 subunit B)						2
Q13330 MTA1_HUMAN	<b>Metastasis-associated protein MTA1</b>						3
Q13416 ORC2_HUMAN	<b>Origin recognition complex subunit 2</b>						3
Q13428 TCOF_HUMAN	<b>Treacle protein</b> (Treacher Collins syndrome protein)						17
Q14258 TRI25_HUMAN	<b>Tripartite motif-containing protein 25</b> (Zinc finger protein 147)	18	20				27
Q14444 GPIA1_HUMAN	<b>GPI-anchored membrane protein 1</b> (GPI-anchored protein p137) (p137GPI)						3
Q14498 RBM39_HUMAN	<b>RNA-binding protein 39</b> (RNA-binding motif protein 39)	3	6				9
Q14676 MDC1_HUMAN	<b>Mediator of DNA damage checkpoint protein 1</b> (Nuclear factor with BRCT domains 1)						22

Q14683 SMC1A_HUMAN	<b>Structural maintenance of chromosomes protein 1A</b> (SMC1alpha protein) (Sb1.8)	4					14
Q14978 NOLC1_HUMAN	<b>Nucleolar phosphoprotein p130</b> (Nucleolar 130 kDa protein)						4
Q15233 NONO_HUMAN	<b>Non-POU domain-containing octamer-binding protein</b> (NonO protein)					13	21
Q15717 ELAV1_HUMAN	<b>ELAV-like protein 1</b> (Hu-antigen R) (HuR)					6	3
Q16531 DDB1_HUMAN	<b>DNA damage-binding protein 1</b> (Damage-specific DNA-binding protein 1)	6					5
Q16666 IFI16_HUMAN	<b>Gamma-interferon-inducible protein Irfi-16</b>	7					9
Q52LJ0 FAM98B_HUMAN	<b>Protein FAM98B</b>						5
Q8IWS0 PHF6_HUMAN	<b>PHD finger protein 6</b> (PHD-like zinc finger protein)						2
Q8NE71 ABCF1_HUMAN	<b>ATP-binding cassette sub-family F member 1</b> (ATP-binding cassette 50)	23	15	4		11	26
Q8TCS8 PNPT1_HUMAN	<b>Polyribonucleotide nucleotidyltransferase 1, mitochondrial precursor</b>						6
Q92616 GCN1L_HUMAN	<b>GCN1-like protein 1</b> (HsGCN1)	55					3
Q92797 SYMPK_HUMAN	<b>Symplekin</b>						3
Q92979 NEP1_HUMAN	<b>Probable ribosome biogenesis protein NEP1</b> (Nucleolar protein EMG1 homolog)					5	11
Q96EP5 DAZP1_HUMAN	<b>DAZ-associated protein 1</b> (Deleted in azoospermia-associated protein 1)						14
Q96T60 PNKP_HUMAN	<b>Bifunctional polynucleotide phosphatase/kinase</b>						6
Q9BQ13 KCD14_HUMAN	<b>BTB/POZ domain-containing protein KCTD14</b>						3
Q9BQ75 CC026_HUMAN	<b>Uncharacterized protein C3orf26</b>						4
Q9BQG0 MBB1A_HUMAN	<b>Myb-binding protein 1A</b>						5
Q9BSD7 U334_HUMAN	<b>Probable UPF0334 kinase-like protein C1orf57</b>					4	8
Q9BVC5 ASHWN_HUMAN	<b>Ashwin</b>						11
Q9H0G5 CCD55_HUMAN	<b>Coiled-coil domain-containing protein 55</b>						3
Q9P258 RCC2_HUMAN	<b>Protein RCC2</b> (Telophase disk protein of 60 kDa) (RCC1-like protein TD-60)						3
Q9UBD5 ORC3_HUMAN	<b>Origin recognition complex subunit 3</b> (Origin recognition complex subunit Latheo)						9
Q9UKD2 MRT4_HUMAN	<b>mRNA turnover protein 4 homolog</b>						2
Q9UQE7 SMC3_HUMAN	<b>Structural maintenance of chromosomes protein 3</b>						15
Q9Y224 CN166_HUMAN	<b>Protein C14orf166</b>					4	11

Q9Y230 RUVB2_HUMAN	<b>RuvB-like 2</b> (EC 3.6.1.-) (48 kDa TATA box-binding protein-interacting protein)					7
Q9Y266 NUDC_HUMAN	<b>Nuclear migration protein nudC</b> (Nuclear distribution protein C homolog)					3
Q9Y3B2 EXOS1_HUMAN	<b>3'-5' exoribonuclease CSL4 homolog</b> (EC 3.1.13.-) (Exosome component 1)					3
Q9Y3I0 CV028_HUMAN	<b>UPF0027 protein C22orf28</b>					15
Q9Y6K5 OAS3_HUMAN	<b>2'-5'-oligoadenylate synthetase 3</b> (EC 2.7.7.-) ((2-5')oligo(A) synthetase 3)	2	9			7
Q9NR30 DDX21_HUMAN	<b>Nucleolar RNA helicase 2</b> (EC 3.6.1.-) (Nucleolar RNA helicase II)					17
P07437 TBB5_HUMAN	<b>Tubulin beta chain</b> (Tubulin beta-5 chain)					8
P68363 TBAK_HUMAN	<b>Tubulin alpha-ubiquitous chain</b> (Alpha-tubulin ubiquitous) (Tubulin K-alpha-1)					2
P60709 ACTB_HUMAN	<b>Actin, cytoplasmic 1</b> (Beta-actin)					12
P27816 MAP4_HUMAN	<b>Microtubule-associated protein 4</b> (MAP 4)					9
P35579 MYH9_HUMAN	<b>Myosin-9</b> (Myosin heavy chain, nonmuscle IIa) (Nonmuscle myosin heavy chain IIa)					2
Q96019 ACL6A_HUMAN	<b>Actin-like protein 6A</b> (53 kDa BRG1-associated factor A)					5
Q9UJX4 ANC5_HUMAN	<b>Anaphase-promoting complex subunit 5</b> (APC5) (Cyclosome subunit 5)					4
Q9UJX3 APC7_HUMAN	<b>Anaphase-promoting complex subunit 7</b> (APC7) (Cyclosome subunit 7)					6
Q3SWX9 RAD21_BOVIN	<b>Double-strand-break repair protein rad21 homolog</b> - Bos taurus (Bovine)					3



**Tableau III : Comparaison des résultats d'identification par spectrométrie de masse du contenu en protéines des RNP formées avec la région des A-repeats murine**

Numéro d'accèsion	Noms des protéines	Exp. 4						Exp. 5					
		J0 ARN-MS2	J0 ARN-AM-MS2	J2 ARN-MS2	J2 ARN-AM-MS2	J4 ARN-MS2	j4 ARN-AM-MS2	J0 ARN-MS2	J0 ARN-AM-MS2	J2 ARN-MS2	J2 ARN-AM-MS2	J4 ARN-MS2	j4 ARN-AM-MS2
<b>hnRNPs</b>													
Q9CX86 Q9CX86_MOUSE	<b>Heterogeneous nuclear ribonucleoprotein A0</b> (hnRNP A0) homolog - Mus musculus	2	3	1	5		2	2	5		2	1	5
Q3U8W9 Q3U8W9_MOUSE	<b>Heterogeneous nuclear ribonucleoprotein R</b> full insert sequence - Mus musculus		3				5	2	6		4		10
Q3U6P5 Q3U6P5_MOUSE	<b>Heterogeneous nuclear ribonucleoprotein C</b> full insert sequence - Mus musculus		7		5		4		3		2		
Q3THB3 Q3THB3_MOUSE	<b>Heterogeneous nuclear ribonucleoprotein M</b> full insert sequence - Mus musculus		4				1		2				3
Q3TIK8 Q3TIK8_MOUSE	Heterogeneous nuclear ribonucleoprotein A1 - Rattus norvegicus	7	8	2	11		5		6		4	3	6
O88569 ROA2_MOUSE	<b>Heterogeneous nuclear ribonucleoprotein B1</b> - Rattus norvegicus		5		4				3		4		2
Q60668 HNRPD_MOUSE	<b>Heterogeneous nuclear ribonucleoprotein D0</b> - Mus musculus		5		5		5		6		4		8
Q9Z130 HNRDL_MOUSE	<b>Heterogeneous nuclear ribonucleoprotein D-like</b> - Mus musculus		7		3				5		1		6
Q91VM5 Q91VM5_MOUSE	<b>Heterogeneous nuclear ribonucleoprotein G</b> - Rattus norvegicus		3		1				2		2		4

P61979 HNRPK_MOUSE	<b>Heterogeneous nuclear ribonucleoprotein K</b> - Mus musculus		5		2		4		8		8		8
Q7TMK9 HNRPQ_MOUSE	<b>Heterogeneous nuclear ribonucleoprotein Q</b> - Mus musculus		7		3		5		10		8		11
Q00PI9 HNRRL2_MOUSE	<b>Heterogeneous nuclear ribonucleoprotein U-like protein 2</b> - Mus musculus								3				
Q3T984 Q3T984_MOUSE	<b>Polypyrimidine tract binding protein 1</b> full insert sequence - Mus musculus		8				10		20		9		17
Q3U8W9 Q3U8W9_MOUSE	<b>Heterogeneous nuclear ribonucleoprotein R</b> full insert sequence - Mus musculus		3				5	2	6		4		10
Q9Z204 HNRPC_MOUSE	<b>Heterogeneous nuclear ribonucleoprotein C</b> full insert sequence - Mus musculus		7		5		4		3		2		
Q3THB3 Q3THB3_MOUSE	<b>Heterogeneous nuclear ribonucleoprotein M</b> full insert sequence - Mus musculus		4				1		2				3
Q91WJ8 FUBP1_MOUSE	<b>Far upstream element (FUSE) binding protein 1</b> - FBP1 Mus musculus		7				1						2
A2AJ72 A2AJ72_MOUSE	<b>Far upstream element (FUSE) binding protein 3</b> - Mus musculus		5		2				17		8		9
Q99PF5 FUBP2_RAT	<b>Far upstream element-binding protein 2</b> - KSRP Rattus norvegicus		10				9		6		10		10

A2AU61 A2AU61_MOUSE	<b>HnRNP-associated with lethal yellow</b> (Fragment) - Mus musculus		2										
Q3TVV6 Q3TVV6_MOUSE	<b>Heterogeneous nuclear ribonucleoprotein U hnRNP/SAF-A</b> full insert sequence - Mus musculus	2	11	1	3	1	8	1	14		11		12
B2M1R7 B2M1R7_MOUSE	<b>Poly(RC) binding protein 2</b> - Mus musculus		1										
P60335 PCBP1_MOUSE	<b>Poly(rC)-binding protein 1</b> - Mus musculus		2		2								
Q3UEB3 PUF60_MOUSE Q9WV25 PUF60_RAT	<b>Poly(U)-binding-splicing factor PUF60</b> - Mus musculus		6		4				7		6		7
Q66H20 PTBP2_RAT Q91Z31 PTBP2_MOUSE	<b>Polypyrimidine tract-binding protein 2</b> - Rattus norvegicus								2		1		3
<b>Métabolisme de l'ADN</b>													
Q3V0E7 Q3V0E7_MOUSE	<b>TAR DNA binding protein</b> full insert sequence - Mus musculus	2	3						7		3		
P13864 DNMT1_MOUSE	<b>DNA (cytosine-5)-methyltransferase 1</b> - Mus musculus		3										
Q04750 TOP1_MOUSE	<b>DNA topoisomerase 1</b> - Mus musculus		1										
Q6PET9 Q6PET9_MOUSE	<b>DNA topoisomerase 2</b> (Fragment) - Mus musculus	1	3				2	4	1				
Q01320 TOP2A_MOUSE	<b>DNA topoisomerase 2-alpha</b> - Mus musculus	10	27	4	3		13	21	19		12		8
P41516 TOP2A_RAT	<b>DNA topoisomerase 2-alpha</b> - Rattus norvegicus	1	8					8	5		1		



Q08943 SSRP1_MOUSE	<b>Structure specific recognition protein 1</b> (Fragment) - Mus musculus				1		1						
Q80U70 SUZ12_MOUSE	<b>Suppressor of zeste 12 homolog</b> - Mus musculus		2										
Q62318 TIF1B_MOUSE	<b>Tripartite motif protein 28</b> - Mus musculus	11	13		3				6		5		
<b>Métabolisme de l'ARN</b>													
Q91VR5 DDX1_MOUSE	<b>DEAD (Asp-Glu-Ala-Asp) box polypeptide 1</b> full insert sequence - Mus musculus		5										
Q64213 SF01_MOUSE	<b>Zinc finger protein 162 full insert sequence</b> - Mus musculus		3				3				6		3
Q92119 EXOS4_MOUSE	<b>RRP41</b> full insert sequence - Mus musculus		1										
Q70133 DHX9_MOUSE	<b>ATP-dependent RNA helicase A</b> - Mus musculus	12	3	2			5	3	6		5		
Q3UAI4 Q3UAI4_MOUSE	<b>Splicing factor 3b subunit 2</b> full insert sequence - Mus musculus		2										
Q9JII5 DAZP1_MOUSE	<b>DAZ associated protein 1</b> full insert sequence - Mus musculus		4										
Q501J6 DDX17_MOUSE	<b>Probable RNA-dependent helicase p72</b> - Mus musculus		5						8				
Q9D287 SPF27_MOUSE	<b>Breast carcinoma amplified sequence 2</b> full insert sequence - Mus musculus		1				2						

A1L333 A1L333_MOUSE	<b>DEAD (Asp-Glu-Ala-Asp) box polypeptide 5</b> - Mus musculus	4	14		4		8	2	16		11		8
P70318 TIAR_MOUSE	<b>Nucleolysin TIAR</b> - Mus musculus		3		2						4		4
Q8K4Z5 SF3A1_MOUSE	<b>Splicing factor 3a</b> subunit 1 full insert sequence - Mus musculus		2										
Q3TVJ3 Q3TVJ3_MOUSE	<b>DEAD (Asp-Glu-Ala-Asp) box polypeptide 21</b> full insert sequence - Mus musculus	2	7		5		11		7		6		13
Q9D824 FIP1_MOUSE	Pre-mRNA 3'-end-processing factor FIP1 - Rattus norvegicus								1				
Q99KP6 PRP19_MOUSE	<b>Prp19 beta protein</b> - Mus musculus		2										
A5D6Q4 A5D6Q4_MOUSE	<b>Prpf8 protein</b> - Mus musculus		12				2				4		
O35286 DHX15_MOUSE	<b>Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15</b> - Mus musculus		3				1						
O88796 RPP30_MOUSE	<b>Ribonuclease P protein subunit p30</b> - Mus musculus		2				2						
A6QRH4 A6QRH4_MOUSE	<b>RNA binding motif protein 39</b> (Fragment) - Mus musculus		3						5				
Q3UFZ6 Q3UFZ6_MOUSE	<b>RNA granule protein 105</b> - Rattus norvegicus		2										
P56959 FUS_MOUSE	<b>RNA-binding protein FUS</b> - Mus musculus		4		3				3		2		
A1L3C2 A1L3C2_MOUSE	<b>Rpo1-1 protein</b> - Mus musculus		2		1								
P35550 FBRL_MOUSE	<b>rRNA 2'-O-methyltransferase fibrillarin</b> - Rattus norvegicus		2		2								

Q497V5 SRBD1_MOUSE	<b>S1 RNA-binding domain-containing protein 1</b> - Mus musculus								1				
Q9D0D4 DIMIT1_MOUSE	<b>DIMIT1</b> - Rattus norvegicus						2						
P27048 RSMB_MOUSE	<b>Small nuclear ribonucleoprotein B</b> - Mus musculus		4		2		3		2		2		2
P62315 SMD1_MOUSE	<b>Small nuclear ribonucleoprotein D1</b> - Rattus norvegicus										2		2
P62305 RUXE_MOUSE	<b>Small nuclear ribonucleoprotein E</b> - Mus musculus								2				
P62317 SMD2_MOUSE	<b>Small nuclear ribonucleoprotein Sm D2</b> - Mus musculus										2		3
P62320 SMD3_MOUSE	<b>Small nuclear ribonucleoprotein Sm D3</b> - Mus musculus	1											
Q9D554 SF3A3_MOUSE	<b>Splicing factor 3a subunit 3</b> - Rattus norvegicus		1										
Q99NB9 SF3B1_MOUSE	<b>Splicing factor 3B subunit 1</b> - Mus musculus		2						2		4		
Q8QZY9 SF3B4_MOUSE	<b>Splicing factor 3B subunit 4</b> - Rattus norvegicus										2		
Q921M3 SF3B3_MOUSE	<b>Splicing factor 3b subunit 3</b> - Mus musculus		2						4				
Q8VIJ6 SFPQ_MOUSE	<b>Splicing factor proline/glutamine rich</b> - Mus musculus	8	10		12		12		13		10		11
P26369 U2AF2_MOUSE	<b>Splicing factor U2AF 65 kDa subunit</b> - Mus musculus		7		2		6		6		3		7

Q6PDM2 SFRS1_MOUSE	<b>Splicing factor arginine/serine-rich 1</b> - Mus musculus		6		4		2		8		7		7
Q62093 SFRS2_MOUSE	<b>Splicing factor arginine/serine-rich 2</b> - Mus musculus								2				2
P84104 SFRS3_MOUSE	<b>Splicing factor arginine/serine-rich 3</b> - Mus musculus		3		2		3		1		1		2
O35326 SFRS5_MOUSE	<b>Splicing factor arginine/serine-rich 5</b> - Mus musculus		2		1						2		1
Q62189 SNRPA_MOUSE	<b>U1 small nuclear ribonucleoprotein A</b> - Mus musculus		4		3				3		3		
A2RS68 A2RS68_MOUSE B2RZ74 B2RZ74_RAT Q62376 RU17_MOUSE	<b>U1 small nuclear ribonucleoprotein polypeptide A</b> - Mus musculus		3		2		2						
P57784 RU2A_MOUSE	<b>U2 small nuclear ribonucleoprotein A'</b> - Mus musculus		4		6		2		5		5		3
Q14C24 Q14C24_MOUSE Q3KR55 Q3KR55_RAT Q9D883 U2AF1_MOUSE	<b>U2 small nuclear ribonucleoprotein auxiliary factor (U2AF) 1</b> - Mus musculus		4		7		5		3		7		4
<b>Transcription/Traduction</b>													
P62908 RS3_MOUSE	<b>40S ribosomal protein S3</b> - Mus musculus		4		3		3						
P62702 RS4X_MOUSE	<b>40S ribosomal protein S4 X isoform</b> - Mus musculus						2						
P62754 RS6_MOUSE	<b>40S ribosomal protein S6</b> - Mus musculus						1						
P62082 RS7_MOUSE	<b>40S ribosomal protein S7</b> - Mus musculus		2		3		4						2



P47963 RL13_M OUSE	<b>60S ribosomal protein L13</b> - Rattus norvegicus		1										
P19253 RL13A_M OUSE	<b>60S ribosomal protein L13a</b> - Mus musculus						2						
P35980 RL18_M OUSE	<b>60S ribosomal protein L18</b> - Mus musculus				2		1						
P62717 RL18A_M OUSE	<b>60S ribosomal protein L18a</b> - Mus musculus						2						
P67984 RL22_M OUSE	<b>60S ribosomal protein L22</b> - Mus musculus							2	2		2	2	2
P62900 RL31_M OUSE	<b>60S ribosomal protein L31</b> - Mus musculus									2			2
P47911 RL6_M OUSE	<b>60S ribosomal protein L6</b> - Mus musculus				2		2						
P14148 RL7_M OUSE	<b>60S ribosomal protein L7</b> - Mus musculus				1								
P51410 RL9_M OUSE	<b>60S ribosomal protein L9</b> - Rattus norvegicus						2						
O35643 AP1B1_M OUSE	<b>AP-1 complex subunit beta-1</b> - Mus musculus						2						
P10126 EF1A1_M OUSE	<b>Elongation factor 1-alpha 1</b> - Mus musculus	7	5	3	5	4	4	7		6	2	5	3
A2AH85 A2AH85_M OUSE	<b>Elongation factor Tu GTP binding domain containing 2</b> - Mus musculus		8				2						
P10630 IF4A2_M OUSE	<b>Eukaryotic initiation factor 4A-II</b> - Mus musculus		1										
P68101 IF2A_RA T	<b>Eukaryotic translation initiation factor 2 subunit 1</b> - Rattus norvegicus		6		2								

Q63505 TF3C1_RAT	<b>General transcription factor 3C polypeptide 1</b> - Rattus norvegicus		2										
Q148Y1 Q148Y1_MOUSE	General transcription factor IIC polypeptide 4 - Mus musculus		2										
Q9Z0N1 IF2G_MOUSE	<b>Eukaryotic translation initiation factor 2 subunit 3</b> structural gene X-linked full insert sequence - Mus musculus		5				2		3		2		5
P25444 RS2_MOUSE	<b>Ribosomal protein S2</b> - Rattus norvegicus						1						
P62242 RS8_MOUSE	<b>Ribosomal protein S8</b> - Rattus norvegicus				2		4						
Q91WK2 EIF3H_MOUSE	<b>Eukaryotic translation initiation factor 3 subunit 3</b> (gamma) full insert sequence - Mus musculus						1						
P25976 UBF1_MOUSE	<b>Upstream binding transcription factor RNA polymerase I</b> - Mus musculus		6										
<b>Autres</b>													
Q3TJ01 Q3TJ01_MOUSE	<b>HSPC117</b> homolog - Mus musculus		3										
Q99K48 NONO_MOUSE	<b>non-POU-domain-containing octamer binding protein</b> full insert sequence (Fragment) - Mus musculus	8	9		8		5		9		7		6
B2RRX1 B2RRX1_MOUSE	<b>Actin beta cytoplasmic</b> - Mus musculus	6	4	5	8	8	10						1
P63260 ACTG_MOUSE	<b>Actin cytoplasmic 2</b> - Rattus norvegicus					1	2						

Q9CPP0 NPM3_MOUSE	<b>Nucleoplasmin 3</b> full insert sequence - Mus musculus	1	1		8				1				
Q3U0Y9 Q3U0Y9_MOUSE	<b>Phenylalanine-tRNA synthetase-like alpha subunit</b> full insert sequence - Mus musculus	2					2						
Q6P4T2 Q6P4T2_MOUSE	<b>Activating signal cointegrator 1 complex</b> subunit 3-like 1 - Mus musculus		19					7				3	
Q3UV17 Q3UV17_MOUSE	similar to <b>Keratin 2p</b> - Mus musculus				1							1	1
Q20BD0 Q20BD0_MOUSE	<b>AIF-C1</b> - Rattus norvegicus	3	9		7	2	9		10		10		9
P00762 TRY1_RAT	<b>Anionic trypsin-1 precursor</b> - Rattus norvegicus				1								
Q3TK27 Q3TK27_MOUSE	<b>Nucleostemin</b> full insert sequence - Mus musculus							1					
Q9WVA3 BUB3_MOUSE	<b>Budding uninhibited</b> by benzimidazoles 3 homolog - Rattus norvegicus		1										
P08779 K1C16_HUMAN	cDNA FLJ75925 highly similar to Homo sapiens <b>keratin 16</b> - Homo sapiens				6		3						
Q3TF62 Q3TF62_MOUSE	53 kDa <b>BRG1-associated factor A</b> (Actin-related protein Baf53a) full insert sequence - Mus musculus	1	3										
Q3UCF2 Q3UCF2_MOUSE	<b>Cell division cycle 5-like</b> - Rattus norvegicus		2										
Q08288 LYAR_MOUSE Q6AYK5 LYAR_RAT	<b>Cell growth-regulating nucleolar protein</b> - Mus musculus				1			2					

Q5SUF2 CROP_MOUSE	<b>Cisplatin resistance-associated overexpressed protein</b> - Mus musculus		2										
Q9EPU4 CPSF1_MOUSE	<b>Cleavage and polyadenylation specificity factor subunit 1</b> - Mus musculus		7						2		7		1
Q4KM65 CPSF5_RAT Q9CQF3 CPSF5_MOUSE	<b>Cleavage and polyadenylation specificity factor subunit 5</b> - Rattus norvegicus		4		4		7				2		3
Q6NVF9 CPSF6_MOUSE	<b>Cleavage and polyadenylation specificity factor subunit 6</b> - Mus musculus						3						
Q5XI29 CPSF7_RAT Q8BTV2 CPSF7_MOUSE	<b>Cleavage and polyadenylation specificity factor subunit 7</b> - Rattus norvegicus		2										
A2AEK0 A2AEK0_MOUSE Q8BIQ5 CSTF2_MOUSE	<b>Cleavage stimulation factor 3' pre-RNA subunit 2</b> - Mus musculus								1		6		
A2AP97 A2AP97_MOUSE	<b>Cleavage stimulation factor 3' pre-RNA subunit 1</b> - Mus musculus		2						2				2
A2AQH9 A2AQH9_MOUSE Q99LI7 CSTF3_MOUSE	<b>Cleavage stimulation factor 3' pre-RNA subunit 3</b> - Mus musculus		4				1		8		7		5
P60824 CIRBP_MOUSE	<b>Cold-inducible RNA-binding protein</b> - Mus musculus								2				
P70372 ELAV1_MOUSE	<b>ELAV-like protein 1</b> - Mus musculus		4		4		5		5		3		4
P62827 RAN_MOUSE	<b>GTP-binding nuclear protein Ran</b> - Mus musculus	3	4		4	4	2						

P68040 GBLP_M MOUSE	<b>Guanine nucleotide-binding protein subunit beta-2-like 1</b> - Rattus norvegicus		2		3		2						
P63017 HSP7C_ MOUSE	<b>Heat shock cognate 71 kDa protein</b> - Mus musculus	12	11	3	2	2	2		7		2		10
Q8VDJ3 VIGLN_ MOUSE Q9Z1A6 VIGLN_ RAT	High density lipoprotein binding protein - Rattus norvegicus <b>Vigilin</b>		5		1		4						
P63158 HMGB1_ MOUSE	<b>High mobility group protein B1</b> - Mus musculus						2						
Q00547 HMMR_ MOUSE	<b>Hyaluronan mediated motility receptor</b> - Mus musculus		3										
P52296 IMB1_RA T	<b>Importin subunit beta-1</b> - Rattus norvegicus	1	2										
Q88477 IF2B1_M MOUSE	<b>Insulin-like growth factor 2 mRNA-binding protein 1</b> - Mus musculus	3	5	2			3		4		3		7
B2CY77 B2CY77_ MOUSE	<b>Laminin receptor (Fragment)</b> - Mus musculus	2	4		3		3						
P19137 LAMA1_ MOUSE	<b>Laminin subunit alpha-1 precursor</b> - Mus musculus						17						
P02469 LAMB1_ MOUSE	<b>Laminin subunit beta-1 precursor</b> - Mus musculus						4						
P02468 LAMC1_ MOUSE	<b>Laminin subunit gamma-1 precursor</b> - Mus musculus						2						
Q6PB66 LPPRC_ MOUSE	<b>Leucine-rich PPR motif-containing protein</b> mitochondrial precursor - Mus musculus		5				2				7		
Q8K3Y3 LN28A_ MOUSE	<b>Lin-28 homolog A</b> - Mus musculus				4								

P32067 LA_MOUSE	<b>Lupus La protein</b> homolog - Mus musculus	3	3	2			2	1	3		3		4
Q8K310 MATR3_MOUSE	<b>Matrin-3</b> - Rattus norvegicus		5						10		6		9
P27546 MAP4_MOUSE	<b>Microtubule-associated protein 4</b> - Mus musculus		2										
Q8C854 MYEF2_MOUSE	<b>Myelin basic protein expression factor 2 repressor</b> - Mus musculus		2						3				
Q7TPV4 MBB1A_MOUSE	<b>MYB binding protein (P160) 1a</b> full insert sequence - Mus musculus	3	4	2	2	5	14	8	17		11		8
Q8CIM0 Q8CIM0_MOUSE	<b>Cleavage and polyadenylation specificity factor 3</b> full insert sequence - Mus musculus		2						3		3		3
Q3U2B5 Q3U2B5_MOUSE	<b>Symplekin</b> homolog - Mus musculus		4						7		6		
A2ATZ4 A2ATZ4_MOUSE	<b>Novel protein</b> - Mus musculus		1										
Q45VK5 Q45VK5_MOUSE	<b>Nuclear factor 90</b> - Mus musculus	9	7	9	2	9	7	15	19	11	8	15	15
P62960 YBOX1_MOUSE	<b>Nuclease-sensitive element-binding protein 1</b> - Mus musculus	3	5		2		1		3		2		4
A2APD6 A2APD6_MOUSE	<b>Nucleolar protein 5A</b> - Mus musculus						1						
A2AA61 A2AA61_MOUSE Q9CRB2 NOLA2_MOUSE	<b>Nucleolar protein family A member 2</b> - Mus musculus						1						
P09405 NUCL_MOUSE	<b>Nucleolin</b> - Mus musculus	7	16	1	13	4	17		10		12		13
P13084 NPM_RAT	<b>Nucleophosmin</b> - Rattus norvegicus	3	4		2			4	4		5	3	5
Q9WTM5 RUVB2_MOUSE	<b>RuvB-like protein 2</b> full insert sequence - Mus musculus	5	9		2		3						

Q99JF8 PSIP1_MOUSE	<b>PC4 and SFRS1</b> interacting protein 1 - Mus musculus		2										
Q8CCS6 PABP2_MOUSE	<b>Polyadenylate-binding protein 2</b> - Mus musculus							3		2			2
P50580 PA2G4_MOUSE	<b>Proliferation-associated protein 2G4</b> - Mus musculus		2										
Q7TNV0 DEK_MOUSE	Protein <b>DEK</b> - Mus musculus		5										
P97376 FRG1_MOUSE	Protein FRG1 - Mus musculus		1										
B2RXM2 B2RXM2_MOUSE	<b>Putative uncharacterized protein</b> - Mus musculus	7	8		5		4	1	7		4		5
Q9CXY6 ILF2_MOUSE	<b>Putative uncharacterized protein</b> - Rattus norvegicus	10	6	8		5	2	13	14	11	7	10	13
Q4ZF3 DDX49_MOUSE	<b>Putative uncharacterized protein</b> - Rattus norvegicus		2		7								
Q3TIX9 SNUT2_MOUSE	<b>Putative uncharacterized protein</b> - Rattus norvegicus		3										
Q91V55 Q91V55_MOUSE	<b>Putative uncharacterized protein</b> - Rattus norvegicus				2		2						
P12970 RL7A_MOUSE	<b>Putative uncharacterized protein</b> - Rattus norvegicus				1		6						
P97855 G3BP1_MOUSE	<b>Ras GTPase-activating protein-binding protein 1</b> - Mus musculus		2										
Q4VA29 Q4VA29_MOUSE	RIKEN cDNA <b>2700060E02</b> gene - Mus musculus		4				2						
Q8BHD7 ROD1_MOUSE	<b>ROD1 regulator of differentiation 1</b> - Mus musculus						2						
P07724 ALBU_MOUSE	<b>Serum albumin precursor</b> - Mus musculus		1										

Q91YR7 PRP6_MOUSE	Similar to RIKEN cDNA 1190003A07 - Rattus norvegicus		2										
P80316 TCPE_MOUSE	<b>T-complex protein 1</b> subunit epsilon - Mus musculus		2										
O08583 THOC4_MOUSE Q0VBL5 Q0VBL5_MOUSE	<b>THO complex subunit 4</b> - Mus musculus		3		3					2		1	
Q497C1 Q497C1_RAT	<b>Thyroid hormone receptor interactor 12</b> - Rattus norvegicus									5			
O08784 TCOF_MOUSE	<b>Treacle protein</b> - Mus musculus		1										
B2KGS5 B2KGS5_MOUSE	<b>Tripartite motif protein 25</b> - Mus musculus		3				3						
A2RSY6 TRM1L_MOUSE	<b>TRM1-like protein</b> - Mus musculus	3	2										
Q8BHA0 CR037_MOUSE	<b>Uncharacterized protein C18orf37</b> homolog - Mus musculus		2										
P61965 WDR5_MOUSE	<b>WD repeat domain 5</b> - Mus musculus		1										
B1AQG7 B1AQG7_MOUSE	<b>Zinc finger protein 207</b> (Fragment) - Mus musculus		1										
B2RUG7 B2RUG7_MOUSE O88532 ZFR_MOUSE	<b>Zinc finger RNA binding protein</b> - Mus musculus								4				



**Tableau IV : Comparaison des résultats d'identification par spectrométrie de masse du contenu en protéines des RNP formées avec la région des A-repeats murine, différents sous-fragments de cette région et la région 5' terminale de l'ARN Xist murin avec l'extrait nucléaire de cellules ES non différenciées**

		Exp. 6					
Noms des protéines	Noms d'accension	ARN-MS2	ARN-1R-MS2	ARN-2RM-MS2	ARN-SLS1M-MS2	ARN-AM-MS2	ARN-5'M-MS2
<b>hnRNPs</b>							
<b>Far upstream element-binding protein 1, FBP1</b>	FUBP1_MOUSE				4	4	4
<b>Far upstream element-binding protein 2, KSRP</b>	FUBP2_RAT				9	7	7
Far upstream element-binding protein 3	FUBP3_HUMAN				3	1	
Heterogeneous nuclear ribonucleoprotein A/B	ROAA_MOUSE	2	2	2	7	9	12
Heterogeneous nuclear ribonucleoprotein A0	ROA0_HUMAN	2		1	3	2	2
Heterogeneous nuclear ribonucleoprotein A1	ROA1_MOUSE	1		1	3	4	6
Heterogeneous nuclear ribonucleoprotein A3	ROA3_MOUSE		1	1		2	2
Heterogeneous nuclear ribonucleoprotein D0	HNRPD_MOUSE			2	6	7	8
Heterogeneous nuclear ribonucleoprotein D-like	HNRDL_MOUSE			2		2	2
Heterogeneous nuclear ribonucleoprotein G	HNRPG_RAT				2	1	
Heterogeneous nuclear ribonucleoprotein K	HNRPK_MOUSE			1	3	8	
Heterogeneous nuclear ribonucleoprotein L	HNRPL_MOUSE				1	4	
Heterogeneous nuclear ribonucleoprotein Q	HNRPQ_MOUSE			2	8	4	7
Heterogeneous nuclear ribonucleoprotein R	HNRPR_HUMAN			3	7	2	
<b>Heterogeneous nuclear ribonucleoprotein U, SAF/A</b>	HNRPU_HUMAN	1	3	4	7	12	8
Heterogeneous nuclear ribonucleoprotein U-like protein 2	HNRL2_MOUSE				1		1
Heterogeneous nuclear ribonucleoproteins A2/B1	ROA2_MOUSE				1	3	2
Heterogeneous nuclear ribonucleoproteins C1/C2	HNRPC_MOUSE				3	2	
Poly(rC)-binding protein 1	PCBP1_MOUSE			1	1		
Poly(U)-binding-splicing factor PUF60	PUF60_MOUSE			1	3	3	5
<b>Polypyrimidine tract-binding protein 1</b>	PTBP1_MOUSE			4	16	7	11
<b>Métabolisme de l'ADN</b>							
<b>Breast cancer metastasis-suppressor 1-like protein, BRM1L</b>	BRM1L_MOUSE				1		
<b>Bromodomain adjacent to zinc finger domain protein 1A, BAZ1A</b>	BAZ1A_MOUSE				1		
<b>DNA methyl transferase, DNMT1</b>	DNMT1_MOUSE					2	
DNA topoisomerase 2-alpha	TOP2A_MOUSE				43	36	33
DNA topoisomerase 2-alpha	TOP2A_RAT						
DNA topoisomerase 2-beta	TOP2B_MOUSE				1	1	
DNA-directed RNA polymerase I subunit RPA1	RPA1_MOUSE				3	2	
DNA-directed RNA polymerase I subunit RPA2	RPA2_MOUSE				3	1	

DNA-directed RNA polymerases I and III subunit RPAC1	RPAC1_MOUSE				1		
DNA-directed RNA polymerases I, II, and III subunit RPABC3	RPAB3_MOUSE					1	
DnaJ homolog subfamily C member 21	DJC21_HUMAN				1		
<b>Enhancer of zeste homolog 2, EZH2</b>	EZH2_MOUSE					1	
<b>Histone H2B type 1-F/J/L</b>	H2B_MOUSE				1	1	
<b>Histone H4</b>	H4_MOUSE					1	
<b>Histone-binding protein RBBP4, RbAp46-48</b>	RBBP4_MOUSE					2	
<b>Paired amphipathic helix protein Sin3a</b>	SIN3A_MOUSE				2	1	
Putative methyltransferase METT10D	METT10_MOUSE				1		
Replication factor C subunit 4	RFC4_MOUSE				1		
<b>Sister chromatid cohesion protein PDS5 homolog A, PDS5A</b>	PDS5A_MOUSE					2	
<b>SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5</b>	SMCA5_MOUSE					5	1
<b>Telomere-associated protein, RIF1</b>	RIF1_MOUSE				3	5	
<b>Transcription intermediary factor 1-beta, TIF1b</b>	TIF1B_MOUSE				3	2	3
<b>Métabolisme de l'ARN</b>							
116 kDa U5 small nuclear ribonucleoprotein component	U5S1_MOUSE						1
<b>ATP-dependent RNA helicase A, RHA</b>	DHX9_MOUSE			1			5
ATP-dependent RNA helicase DDX3X	DDX3X_MOUSE				1	1	
DAZ-associated protein 1	DAZP1_HUMAN				2	1	3
Nucleolar RNA helicase 2	DDX21_MOUSE			1	1	9	5
<b>Nucleolysin TIAR</b>	TIAR_MOUSE				1		
Pre-mRNA-processing factor 6	PRP6_MOUSE					1	
Probable ATP-dependent RNA helicase DDX17	DDX17_MOUSE				1	3	
Probable ATP-dependent RNA helicase DDX47	DDX47_MOUSE					1	
Probable ATP-dependent RNA helicase DDX49	DDX49_MOUSE				1		
Probable ATP-dependent RNA helicase DDX5	DDX5_HUMAN				7	8	8
Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	DHX15_MOUSE				1		
RNA-binding protein 39	RBM39_HUMAN,RBM39_PONPY				1		
RNA-binding protein FUS	FUS_MOUSE					2	4
S1 RNA-binding domain-containing protein 1	SRBD1_MOUSE				1		
Small nuclear ribonucleoprotein E	RUXE_MOUSE					2	
Small nuclear ribonucleoprotein F	RUXF_MOUSE					1	
Small nuclear ribonucleoprotein G	RUXG_MOUSE					1	1
Small nuclear ribonucleoprotein Sm D1	SMD1_MOUSE					1	4
Small nuclear ribonucleoprotein Sm D2	SMD2_MOUSE				1	2	2
Small nuclear ribonucleoprotein Sm D3	SMD3_MOUSE						1
Small nuclear ribonucleoprotein-associated protein N	RSMB_MOUSE,RSMN_MOUSE						1
Small subunit processome component 20 homolog	UTP20_MOUSE					1	

Splicing factor 1	SF01_MOUSE			3	8		
Splicing factor 3B subunit 1	SF3B1_MOUSE						1
Splicing factor 3B subunit 3	SF3B3_MOUSE					1	4
Splicing factor U2AF 35 kDa subunit	U2AF1_MOUSE			1	1	5	5
Splicing factor U2AF 65 kDa subunit	U2AF2_MOUSE			5	11	4	5
Splicing factor, arginine/serine-rich 1-like protein 1	SFRS1_MOUSE			1		4	
Splicing factor, arginine/serine-rich 3	SFRS3_MOUSE					3	
Splicing factor, proline- and glutamine-rich	SFPQ_MOUSE				5	4	6
TAR DNA-binding protein 43	TADBP_MOUSE				2		7
U1 small nuclear ribonucleoprotein A	SNRPA_MOUSE						4
U2 small nuclear ribonucleoprotein A'	RU2A_MOUSE						1
U5 small nuclear ribonucleoprotein 200 kDa helicase	U520_HUMAN				2		
Zinc finger RNA-binding protein	ZFR_MOUSE				1		
<b>Transcription/Traduction</b>							
40S ribosomal protein S23	RS23_MOUSE					1	
40S ribosomal protein S3	RS3_MOUSE					1	
40S ribosomal protein S3a	RS3A_MOUSE				1		
40S ribosomal protein S7	RS7_MOUSE				1		1
40S ribosomal protein SA	RSSA_MOUSE				2	1	
60S ribosomal protein L22	RL22_MOUSE					2	
60S ribosomal protein L27	RL27_MOUSE					1	1
AP-1 complex subunit beta-1	AP1B1_MOUSE					3	1
Elongation factor 1-alpha 1	EF1A1_MOUSE					1	
Elongation factor 1-delta	EF1D_MOUSE					1	
Elongation factor 1-gamma	EF1G_MOUSE					1	
Eukaryotic translation initiation factor 2 subunit 1	IF2A_MOUSE					2	
FACT complex subunit SSRP1	SSRP1_MOUSE				1		
Nucleolar transcription factor 1	UBF1_MOUSE					1	
TFIIH basal transcription factor complex p62 subunit	TF2H1_MOUSE				1		
<b>Autres</b>							
78 kDa glucose-regulated protein precursor	GRP78_MOUSE				1		
Actin-like protein 6A	ACL6A_MOUSE				1		
Alpha-S1-casein precursor	CASA1_CAPHI					1	
Casein kinase II subunit alpha	CSK21_MOUSE				1		
Cleavage and polyadenylation specificity factor subunit 1	CPSF1_MOUSE				2	1	
Cleavage and polyadenylation specificity factor subunit 2	CPSF2_MOUSE				2		
Cleavage and polyadenylation specificity factor subunit 3	CPSF3_HUMAN				1	1	
Cleavage and polyadenylation specificity factor subunit 7	CPSF7_MOUSE				1		
Cleavage stimulation factor 50 kDa subunit	CSTF1_MOUSE				1		
Cleavage stimulation factor 77 kDa subunit	CSTF3_MOUSE			1	1	1	1
Cold-inducible RNA-binding protein	CIRBP_MOUSE				1		
CUG-BP- and ETR-3-like factor 1	CELF1_MOUSE						1
ELAV-like protein 1	ELAV1_MOUSE			2	2	1	
Heat shock cognate 71 kDa protein	HSP7C_MOUSE				2	2	2
Importin subunit beta-1	IMB1_MOUSE			1			
Insulin-like growth factor 2 mRNA-binding protein 1	IF2B1_MOUSE			4	1	1	1

Interleukin enhancer-binding factor 2	ILF2_MOUSE	7	1	6	7	5	9
Interleukin enhancer-binding factor 3	ILF3_MOUSE	9	1	5	6	6	5
Leucine-rich PPR motif-containing protein, mitochondrial precursor	LPPRC_MOUSE				7	10	1
<b>Lin-28 homolog A, Lin28</b>	<b>LN28A_MOUSE</b>						<b>1</b>
Lupus La protein homolog	LA_MOUSE			1	1	2	2
Lysozyme C type P precursor	LYSCP_MOUSE	1					
Matrin-3	MATR3_MOUSE			5	5	5	8
Myb-binding protein 1A	MBB1A_MOUSE			3	13	11	15
Myelin expression factor 2	MYEF2_MOUSE						1
Non-POU domain-containing octamer-binding protein	NONO_MOUSE			1	2		
Nucleolin	NUCL_MOUSE		16	1	22	33	15
Nucleophosmin	NPM_MOUSE			1	4	5	4
Nucleoplasmin-3	NPM3_MOUSE				1		2
Paraspeckle component 1	PSPC1_MOUSE						2
PC4 and SFRS1-interacting protein	PSIP1_MOUSE				2	4	
<b>Poly [ADP-ribose] polymerase 1, PARP1</b>	<b>PARP1_RAT</b>				5	7	
Probable E3 ubiquitin-protein ligase TRIP12	TRIPC_HUMAN				4	2	
Probable global transcription activator SNF2L4	SMCA4_HUMAN				3	1	
Protein FAM60A	FA60A_MOUSE				1		1
Retrovirus-related Pol polyprotein from transposon 17.6	POL3_DROME					1	
Ribonuclease P protein subunit p30	RPP30_MOUSE				1		
RuvB-like 1	RUVB1_MOUSE					1	
Serine/threonine-protein phosphatase PP1-beta catalytic subunit	PP1B_MOUSE				2	1	
Symplekin	SYMPK_MOUSE				3	3	3
THO complex subunit 4	THOC4_MOUSE				1	1	
Treacle protein	TCOF_MOUSE				1		
Tripartite motif-containing protein 25	TRI25_MOUSE				3		1
TRM1-like protein	TRM1L_MOUSE			1	4	3	
Ubiquitin carboxyl-terminal hydrolase 7	UBP7_MOUSE				1		
<b>Vigilin</b>	<b>VIGLN_MOUSE</b>				1	2	
Y-box-binding protein 2	DBPA_MOUSE, YBOX2_MOUSE				1	1	
Zinc finger CCCH domain-containing protein 7B	Z3H7B_HUMAN			1			

## Les protéines hnRNP :

Comme nous l'avons mentionné dans l'introduction, l'ARN Xist est transcrit par l'ARN polymérase II. Comme tout autre transcrit de ce type, il est susceptible de lier des protéines hnRNP. Comme nous nous y attendions, les protéines hnRNP A1, A2/B, D, E2, R, Q, U (SAF-A), PTB et KSRP (FUSE Binding protein 2) étaient identifiées, ceci avec un nombre important de peptides par rapport aux autres protéines identifiées (jusqu'à 61 peptides pour la PTB dans le cas de la région des A-repeats humaine entière) (**Tab II, III et IV**). On peut noter que les protéines hnRNP U/SAF-A, PTB et KSRP étaient majoritaires parmi toutes ces protéines hnRNP. Bien qu'il soit possible que ce caractère majoritaire soit le reflet de la quantité respective des protéines dans les extraits nucléaires, l'identification de ces protéines validait notre hypothèse d'un rôle possible de PTB et KSRP sur la structuration de la région des A-repeats.

La protéine hnRNP U/SAF-A (Scaffold Attachment Factor A), que nous avons présentée dans le chapitre introduction, a de multiples fonctions chez l'homme. Il a été montré qu'elle est significativement enrichie sur le Xi dans la lignée cellulaire HEK293 et que cette localisation dépend strictement de son site de fixation aux ARN (Helbig et Fackelmayer, 2003). Elle peut être extraite du Xi par un traitement à la RNase (Fackelmayer, 2005). Il est donc envisageable que SAF-A soit liée à ce chromosome par l'intermédiaire de l'ARN Xist. Le fait que nous détectons la protéine SAF-A (hnRNP U) avec plus de 50 peptides pour les RNP formées avec l'ARN-SLS1H-MS2 (**Tab. II**) est en accord avec ces données. L'étude sur les complexes formés avec les différents fragments de la région des A-repeats montre qu'un élément répété est suffisant pour sa fixation. Nous notons néanmoins que cette protéine semble avoir une faible spécificité de reconnaissance de l'ARN. En effet, elle a été retrouvée au laboratoire dans beaucoup de RNP formées avec un grand nombre d'ARN différents. Nous l'avons aussi détectée avec l'ARN-5'M-MS2 (8 peptides, **Tab. IV**) et parfois avec l'ARN-MS2 (**Tab. IV**).

La protéine PTB qui reconnaît des séquences polypyrimidine riches en résidus U pourrait, comme nous l'avons dit, expliquer la protection des grandes boucles internes et terminales dans nos expériences de test de la protection de la région des A-repeats contre l'action des nucléases en extrait nucléaire. Sa détection (61 peptides pour l'ARN-AH-MS2 par exemple, **Tab. II**) est en accord avec cette hypothèse (**Tab. II, III et IV**). D'après les analyses que nous avons réalisées sur des fragments de la région des A-repeats, la présence de 2 éléments répétés avec leurs éléments espaceurs est suffisante pour fixer PTB (**Tab. II et IV**). Elle a aussi été retrouvée au laboratoire dans plusieurs RNP reconstituées et purifiées par la méthode MS2, mais en général pas au même taux que celui estimé pour la région des A-repeats sur la base du nombre de peptides. Nous l'avons aussi détectée dans les RNP formées avec la région 5' terminale de l'ARN Xist (ARN-5'M-MS2) (11 peptides, **Tab. IV**). Mais, les



protéine peut aussi expliquer nos données de protection contre l'action des RNases en extrait nucléaire. De manière intéressante, nous ne détectons une fixation de KSRP que pour un ARN comportant au moins 4 éléments répétés (**Tab II et IV**). Comme PTB, elle se lie aussi à l'ARN-5'M-MS2 (7 peptides, **Tab. IV**). Elle a été retrouvée au laboratoire dans un certain nombre de RNP reconstituées et purifiées par la méthode MS2, mais comme pour PTB, le nombre de peptides détectés dans ces cas était inférieur à celui observé pour la région des A-repeats. Comme pour PTB, nous pensons que cette protéine peut intervenir dans la structure de la région des A-repeats mais peut aussi être impliquée dans le recrutement d'autres protéines sur la région des A-repeats. En effet, en plus de ses sites de fixation aux ARN, elle comporte des domaines pouvant servir d'ancrage pour d'autres protéines.

Notons qu'une protéine appartenant à la même famille que KSRP appelée FBP1, est aussi retrouvée avec un nombre de peptides important dans les RNP formées avec la région des A-repeats (32 pour la région des A-repeats humaine, **Tab. II**). Comme pour KSRP, sa fixation nécessite la présence d'au moins 4 éléments répétés et leurs séquences espaceurs (**Tab. IV**). Notons qu'elle se lie aussi à l'ARN-5'M-MS2 (4 peptides, **Tab. IV**).

Comme nous l'avons mentionné précédemment, nous avons émis l'hypothèse que la protéine hnRNP C1/C2 pourrait se lier à la région des A-repeats du fait de son affinité connue pour les séquences poly(U) et AGUA(U)<sub>5</sub>GUGGA. Toutefois, un faible nombre de peptides de cette protéine était détecté dans nos analyses (seulement 5 peptides pour SLS1H, **Tab. II**). Il faut noter que toutes les protéines ne sont pas détectées avec la même efficacité par spectrométrie de masse. Cela dépend de leur degré d'ionisation.

### **Les protéines impliquées dans le métabolisme de l'ADN :**

Nous avons également identifié un nombre important de protéines impliquées dans le métabolisme de l'ADN : des protéines hélicases, des protéines de réparation de l'ADN et des topo-isomérases. Il faut noter que ces différentes classes de protéines sont aussi souvent retrouvées au laboratoire dans les autres RNP purifiées par la méthode MS2. Bien que ces protéines n'aient pas été prioritaires dans la suite de l'étude, nous ne pouvons pas exclure qu'elles ne participent pas au rôle de la région des A-repeats.

Les protéines de la classe SWI/SNF que nous avons mentionnées dans l'introduction ont été identifiées seulement pour la région des A-repeats entière et avec un nombre de peptides allant de 1 à 19 en fonction des échantillons (**Tab. II, III et IV**). Comme les complexes SWI/SNF sont impliqués dans l'activation transcriptionnelle (pour revue, Reisman et al., 2009), les protéines de ces complexes pourraient se lier à la région des A-repeats complète et réguler positivement l'expression de Xist. Notons qu'elle se lie aussi à l'ARN-5'M-MS2 (1 peptide, **Tab. IV**).

Nous avons également identifié des protéines contenant des bromodomaines et des chromodomaines. Elles sont identifiées avec un nombre de peptides important (jusqu'à 22 peptides pour la protéine CHD4 « Chromodomain helicase-DNA-binding protein 4 », **Tab. II**). Elles nécessitent la formation d'une des deux grandes structures tige-boucle de la région des A-repeats pour se lier (**Tab. II**). Comme nous l'avons mentionné dans l'introduction, ces protéines contenant des bromodomaines et chromodomaines servent à détecter les modifications des histones et participent aux mécanismes mis en place après leur détection. Comme la région des A-repeats permet la mise en place des modifications requises pour l'inactivation du chromosome X, il est intéressant de les avoir détectées. Elles ont parfois aussi été détectées dans certaines autres RNP purifiées au laboratoire mais pas de manière systématique ni au même niveau. Nous ne les avons pas détectées dans les RNP formées avec les extraits nucléaires de cellules ES de souris peut-être parce que nous avons moins de matériel et qu'elles sont moins abondantes dans ces extraits à l'exception de BAZ-1A. Les données nous amènent donc à penser que la protéine BAZ-1A, qui contient un bromodomaine (Jones *et al.*, 2000), pourrait être intéressante (**Tab. IV**).

Nous avons aussi détecté la protéine Dnmt1 dans les RNP formées avec la région des A-repeats entière murine en extrait nucléaire de cellules ES non différenciées (5 peptides, **Tab. III et IV**). Elle ne lie pas aux sous-fragments 1R et 2R et n'est pas retrouvée pour l'ARN-5'M-MS2 (**Tab. II et IV**). Il semble donc que cette protéine se lie spécifiquement à la région des A-repeats entière ou à sa moitié 5'. Cette observation est particulièrement importante car Dnmt1 est impliquée dans la méthylation de l'ADN, assurant la maintenance de l'inactivation, et l'invalidation de son gène chez la souris est létale. L'absence de Dnmt1 induit une déméthylation massive du génome, provoquant l'expression bi-allélique des gènes soumis à l'empreinte génomique parentale (Li *et al.*, 1992). La maintenance de l'inactivation du chromosome X nécessite la présence de Dnmt1. Le fait qu'elle soit recrutée par l'intermédiaire de l'ARN Xist et plus particulièrement de la région des A-repeats est donc une hypothèse très intéressante qu'il sera important de tester.

Nous avons détecté la protéine TIF1b dans les RNP formées avec la région des A-repeats entière et l'ARN-SLS1H (approximativement 5 peptides pour chaque identification, **Tab. II, III et IV**). Il a été montré que TIF1b régule l'expression du gène *MEST* (mesoderm specific transcript) par son interaction avec la protéine HP1. Elle est essentielle à l'établissement et à la maintenance de la structure hétérochromatinienne de ce gène, état caractérisé par la présence de modifications H3K9me3, H4K20me3 des histones et l'hyperméthylation de l'ADN (marques de la chromatine inactive du Xi). De plus, la disruption de l'interaction entre les protéines TIF1b et HP1 provoque le relargage de TIF1b du promoteur de *MEST* et induit la réactivation de la chromatine (Riclet *et al.*, 2008). Cette protéine impliquée dans l'hétérochromatinisation du gène *MEST* pourrait donc être impliquée



dans l'inactivation du chromosome X. Cette protéine est retrouvée dans les RNP formées avec l'ARN-5'M-MS2 (3 peptides) mais aussi dans beaucoup de RNP étudiées au laboratoire ce qui n'est toutefois pas incompatible avec une activité dans l'inactivation du chromosome X.

Nous avons aussi identifié un nombre de peptides important de la protéine PARP-1 (70 peptides pour les RNP formées avec l'ARN-AH-MS2) avec tous les extraits nucléaires utilisés (**Tab. II, III et IV**). Un élément répété ne suffit pas pour qu'elle se fixe mais elle se lie à tous les autres ARN et a été retrouvée très fréquemment à un taux élevé dans les autres RNP étudiées au laboratoire. Néanmoins, nous ne l'avons pas détectée dans les RNP formées avec la région 5' terminale de l'ARN Xist. Nous soulignerons des données suggérant son implication dans l'inactivation du chromosome X. Lorsque son gène (*PARP-1*) est éteint dans des cellules HEK 293, les gènes du Xi sont réactivés, ce qui suggère son implication possible dans le maintien de l'inactivation du chromosome X (Nusinow *et al.*, 2007). Elle intervient aussi dans d'autres processus liés au métabolisme de l'ADN. En association avec SWAP70, la nucléoline (identifiée avec beaucoup de peptides, 28 pour l'ARN-SLS1H et 23 pour la région des A-repeats entière, **Tab. II**), et la nucléophosmine (aussi identifiée avec beaucoup de peptides, 18 pour l'ARN-SLS1H par exemple, **Tab. II**), elle intervient dans un mécanisme de recombinaison de l'ADN (Borggreffe *et al.*, 1998). Nous ne pouvons pas exclure que PARP-1 ait été retenue dans les RNP par interaction avec la nucléoline ou la nucléophosmine qui ne semblent pas avoir une spécificité étroite de reconnaissance des ARN. Néanmoins, PARP-1 peut être une protéine intéressante à étudier ultérieurement.

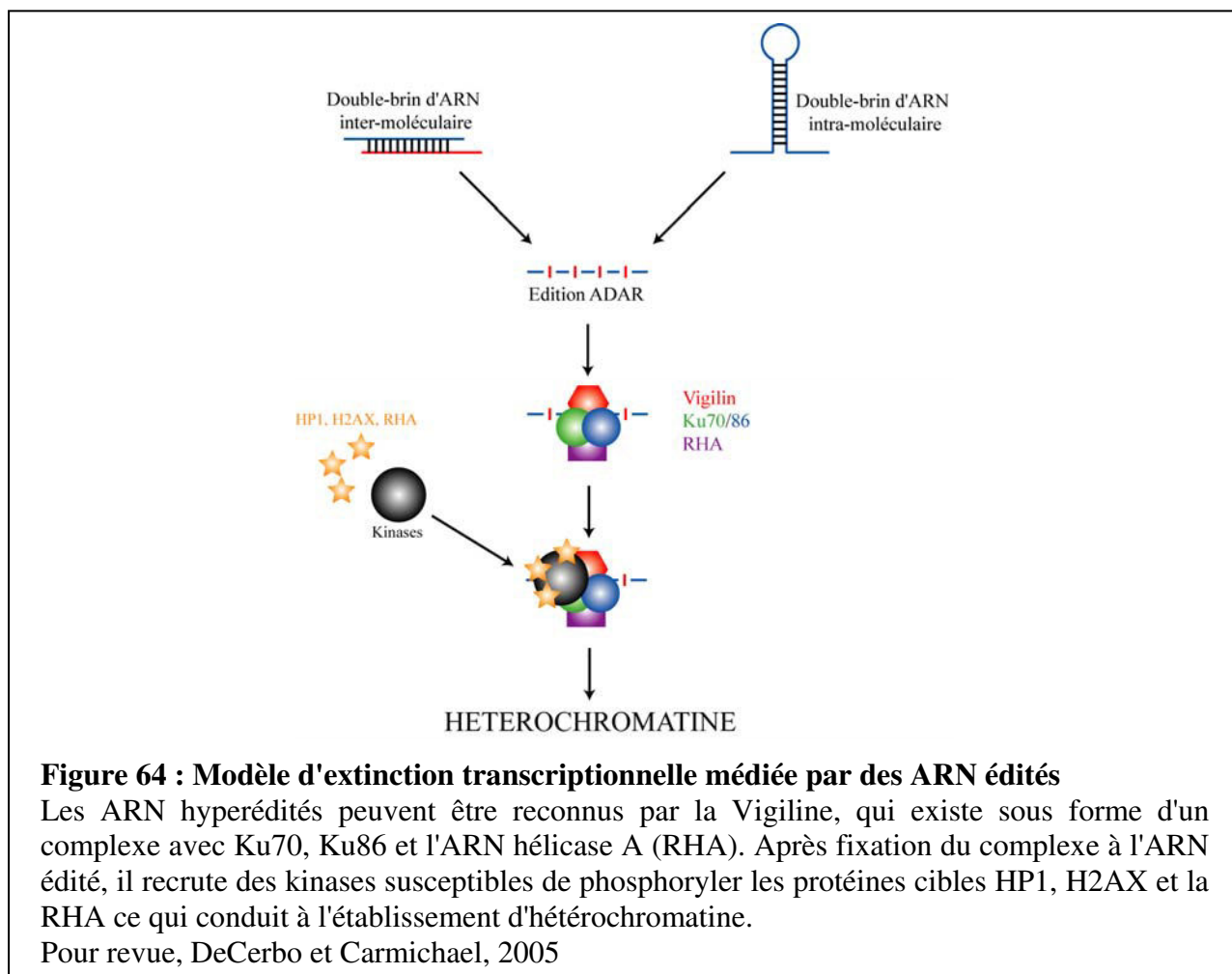
### **Les protéines impliquées dans le métabolisme de l'ARN :**

Nous avons identifié un nombre important de protéines impliquées dans l'épissage des ARN. Même si la région des A-repeats ne contient pas de site d'épissage, la détection de ces protéines n'était pas surprenante. En effet, la région des A-repeats contient des régions riches en résidus U qui peuvent être reconnues par U<sub>2</sub>AF35 (jusqu'à 15 peptides, **Tab. II et IV**). La fixation d'U<sub>2</sub>AF35 sur la séquence polypyrimidine des ARN pré-messagers permet la fixation d'autres facteurs d'épissage en particulier U<sub>2</sub>AF65, un partenaire d'U<sub>2</sub>AF35 (jusqu'à 22 peptides, **Tab. II**), ZFM1 (jusqu'à 16 peptides, **Tab. II**) et des autres facteurs d'épissage identifiés (**Tab. II, III et IV**). On ne peut pas exclure que U<sub>2</sub>AF35 joue un rôle dans d'autres mécanismes que l'épissage, mais aucune donnée n'a été obtenue dans ce sens jusqu'à présent. Vu la présence de grandes boucles riches en résidus U dans la région 5' terminale de l'ARN Xist, nous avons aussi détecté les protéines U<sub>2</sub>AF35 et U<sub>2</sub>AF65 dans les RNP formées avec cet ARN (**Tab. IV**).

Nous avons détecté la protéine Sam68 dans les RNP formées avec l'ARN-SLS1H-MS2 (5 peptides, **Tab. II**). Elle est intéressante car elle se lie à des séquences UAAA et UUUA (Tremblay et Richard, 2006), séquences qui sont présentes dans les régions liant les

motifs répétés. Elle est impliquée dans les mécanismes de régulation de l'épissage. Notons aussi qu'elle recrute également l'histone méthyl-transférase PRMT1 (pour revue, Lukong et Richard, 2003). Cette méthylase permet la mise en place des H4R3me3, marques de l'hétérochromatine du Xi. Cette protéine pourrait donc être intéressante même si nous ne l'avons pas détectée dans les RNP formées en extrait nucléaire de cellules ES. Elle a été retrouvée dans beaucoup de RNP étudiées au laboratoire mais il s'agissait souvent d'études sur l'épissage alternatif.

Nous avons identifié la Vigiline et l'ARN hélicase A (RHA) pour les RNP formées avec l'ARN-SLS1M-MS2 et la région des A-repeats entière en extrait nucléaire de cellules HeLa et ceci avec beaucoup de peptides : respectivement 26 et 22 (**Tab. II, III et IV**). La liaison n'est détectée qu'à partir d'ARN renfermant au moins 4 éléments répétés. On sait que la vigiline est présente au niveau de l'hétérochromatine dans le noyau. La vigiline et la RHA se retrouvent dans un même complexe protéique. En effet, le complexe formé par la protéine d'édition ADAR, les protéines Ku70, Ku86, la RHA et la vigiline, permet de réguler l'expression génique. Dans le noyau, l'adénosine déaminase ARN dépendante (ADAR) agit sur des doubles hélices d'ARN et convertit les résidus A en inosine, ce qui conduit à une fragmentation des ARN, fragments participant au recrutement de kinases qui phosphorylent des cibles telles que HP1, la RHA et l'histone H2AX. Ces phosphorylations participent à l'établissement du silence transcriptionnel de la chromatine (pour revue, DeCerbo et Carmichael, 2005 ; Wang *et al.*, 2005) (**Fig. 64**). On peut imaginer que du fait de la complexation de la vigiline ou la RHA avec Xist, la région des A-repeats puisse être éditée, leurs détections dans les RNP formées avec la région des A-repeats pourraient avoir un sens. Notons néanmoins que la vigiline a été retrouvée dans d'autres RNP étudiées au laboratoire.



D'autres protéines identifiées pourraient aussi avoir un lien avec l'inactivation du chromosome X : la protéine TIAR qui a été détectée avec un nombre important de peptides (jusqu'à 14 peptides avec l'ARN-SLS1H-MS2). Elle est impliquée dans plusieurs processus cellulaires. Elle contient 3 domaines de fixation aux ARN RRM et se lie à des séquences riches en résidus A et U. Elle ne se liait que lorsque l'ARN renfermait au moins 4 éléments répétés (**Tab. II et IV**) et n'était pas retrouvée avec l'ARN-5'M-MS2 (**Tab. IV**). Elle était néanmoins trouvée dans certaines autres RNP étudiées au laboratoire.

La protéine DEK a été détectée dans les RNP formées avec l'ARN-SLS1H-MS2 en extrait nucléaire de cellules HeLa (5 peptides, **Tab. II**) et avec les RNP formées en extrait nucléaire de cellule ES non différenciées (5 peptides, **Tab. III**), mais pas dans les RNP formées avec les sous-fragments de la région des A-repeats en extrait nucléaire de cellules ES non différenciées (**Tab. IV**). La protéine DEK est connue pour son implication dans le remodelage de la chromatine. C'est une sous-unité d'un complexe de répression transcriptionnel incluant HDAC2 (Hollenbach *et al.*, 2002). La possibilité que DEK s'associe à la région des A-repeats est donc intéressante à tester car elle pourrait recruter l'histone désacétylase HDAC permettant de désacétyler les histones, préalable à une extinction transcriptionnelle par méthylation.

### Les protéines impliquées dans la transcription et la traduction :

Nous avons détecté beaucoup de protéines impliquées dans la transcription, la traduction ou d'autres processus cellulaires mais, soit elles étaient détectées avec un faible nombre de peptides, soit lorsqu'un nombre plus élevé de peptides était détecté, ces protéines étaient aussi détectées dans la majorité des RNP étudiées au laboratoire comme c'est le cas pour la protéine eEFA1-1 (35 peptides pour la région des A-repeats). Nous considérerons donc que des études ultérieures sur ces protéines ne sont pas prioritaires.

En conclusion, sur la base de notre exploitation de la littérature, 14 protéines retrouvées dans les expériences, quelque soit l'origine de l'extrait (à l'exception de Sam68 seulement trouvée dans les extraits HeLa), peuvent avoir un rôle dans la fonction de la région des A-repeats. Certaines protéines pourraient stabiliser la structure secondaire de la région des A-repeats comme PTB, KSRP ou FBP1. Les autres protéines pourraient, en se liant à la région des A-repeats, jouer soit un rôle direct dans l'inactivation du chromosome X, soit un rôle indirect en tant que plateforme pour le recrutement des protéines ayant un rôle dans l'inactivation du chromosome X (hnRNP U/SAF-A, BAZ-1A, Dnmt1, TIF1b, PARP1, Sam68, Vigiline, RHA, TIAR, les protéines des complexes SWI/SNF et DEK).

2.1.2 Les protéines qui pourraient avoir un lien avec l'inactivation du chromosome X et qui sont spécifiques aux RNP formées dans les extraits nucléaires de cellules ES de souris

Comme dans le paragraphe précédent, nous classerons ces protéines en 4 catégories : des protéines impliquées dans le métabolisme de l'ADN, des protéines impliquées dans le métabolisme des ARN, des protéines déjà décrites comme intervenant dans l'inactivation du chromosome X (qui ont fait l'objet de l'étude détaillée présentée dans l'article en particulier Suz12, RbAp46, RbAp48, Ezh2) et la catégorie des autres protéines.

### Protéines impliquées dans le métabolisme de l'ADN :

Nous avons détecté 3 des protéines histones H1 (4 peptides pour la région des A-repeats murine placée en présence d'extrait nucléaire de cellules ES J4, **Tab. III**), H2B (1 peptide pour les ARN-SLS1M-MS2 et la région des A-repeats murine entière placée dans l'extrait nucléaire de cellules ES J0, **Tab. IV**) et H4 (1 peptide pour la région des A-repeats entière, **Tab. IV**). Ces protéines ne sont détectées que pour des ARN comportant au moins 4 éléments répétés et n'étaient pas présentes dans les expériences réalisées avec l'ARN-5'M-MS2 (**Tab. IV**). De manière inattendue, nous n'avons pas identifié l'histone H3 qui est la cible des modifications H3K27me3 et H3K9me2 caractéristiques du Xi. Néanmoins, comme mentionné ci-dessus, une absence de détection par spectrométrie de masse ne signifie pas forcément que la protéine est absente. La présence d'histones dans les RNP pourrait

s'expliquer si l'ARN Xist interagit avec le chromosome X au travers d'interactions directes ou médiées par des protéines entre certaines régions de cet ARN et les histones.

Nous avons également détecté la protéine PDS5A. Elle ne se lie qu'à la région des A-repeats entière et 2 peptides de ses peptides ont été identifiés avec l'extrait nucléaire de cellules ES non différenciées (**Tab. IV**). Cette protéine pourrait donc s'associer spécifiquement à la région des A-repeats. Elle fait partie d'un complexe multi-protéique appelé COHESINE impliqué dans la cohésion des chromatides sœurs, processus essentiel à la ségrégation des chromosomes au cours de la mitose. Le complexe COHESINE correspond à un ensemble de protéines liées de manière stable SMC1, SMC3, SCC1 et SCC3, ensemble qui s'associe de manière plus transitoire aux facteurs PDS5, WAPL, SCC2, SCC4, ECO1 et à la séparase, qui en assure la régulation (Hartman *et al.*, 2000 ; Nasmyth et Haering, 2005 ; Skibbens *et al.*, 1999 ; Kueng *et al.*, 2006 ; Chen *et al.*, 2002 ; Losada *et al.*, 2005). Ce complexe pourrait réguler l'expression génique en contrôlant l'architecture globale de la chromatine, en particulier la formation de boucles inter ou intra-chromosomique. Un lien est donc envisageable avec la fonction majeure de la région des A-repeats qui est de re-localiser les gènes liés au chromosome X à inactiver dans le domaine Xist (Chaumeil *et al.*, 2006).

Nous avons également identifié 5 peptides de la protéine RIF1 associée au télomère. Cette protéine a été identifiée dans les RNP formées avec des ARN comportant au moins 4 éléments répétés (3 peptides, **Tab. IV**) et ne se lie pas sur l'ARN-5'M-MS2. Bien qu'aucun site de fixation aux ARN n'ait été identifié dans cette protéine, elle sera intéressante à étudier ultérieurement. En effet, chez les eucaryotes, les régions télomériques sont condensées en hétérochromatine par l'interaction des séquences télomériques avec plusieurs facteurs protéiques, dont RIF1. Ces facteurs forment un complexe appelé le télosome (Wright et Zakian, 1995). Ce complexe télomérique permet la mise en place de l'hétérochromatine par le recrutement de protéines à activité histone désacétylase. L'implication de RIF1 est donc envisageable. La région des A-repeats pourrait recruter les protéines impliquées dans la désacétylation de la chromatine préalable à la méthylation des histones par le complexe PRC2.

Nous avons détecté 2 peptides de la protéine Sin3a pour l'ARN-SLS1M-MS2 et 1 peptide pour la région des A-repeats entière (**Tab. IV**). La délétion de la protéine Sin3a provoque la mort de l'embryon avant même son implantation. Sin3a n'a pas été détectée dans les RNP formées avec l'ARN-5'M-MS2 ce qui suggère une spécificité possible de liaison à la région des A-repeats. Cette protéine joue un rôle de plateforme de recrutement d'histone désacétylase HDAC dans les régions transcriptionnelles. Le complexe ainsi recruté est extrêmement conservé de la levure jusqu'à l'homme et est formé grâce à Sin3a de 8 protéines chez les mammifères : Sin3a, HDAC1, HDAC2, RbAp46, RbAp48, SP30, SAP18 et SDS3 (pour revue, Silverstein et Ekwall, 2005). Nous avons identifié 3 des 8 composants de ce

complexe dans les RNP formées avec les ARN-SLS1M-MS2 et ARN-AM-MS2 : Sin3a, RbAp46 et RbAp48. Notons que les protéines RbAp46 et RbAp48 font également partie du complexe PRC2 décrit dans l'article. Nous avons également identifié la protéine BRM1L qui a une forte homologie de séquence avec la protéine SDS3 du complexe décrit ci-dessus (1 peptide pour les RNP formées avec l'ARN-SLS1M-MS2). On pourrait imaginer que la région des A-repeats participe au recrutement du complexe Sin3a/HDAC vraisemblablement par l'intermédiaire d'une autre protéine se liant à l'ARN, puisqu'aucun site de fixation aux ARN n'a été identifié dans Sin3a et donc que l'interaction, si elle existe, n'est sans doute pas directe. Il faut aussi noter qu'aucune protéine HDAC n'a été détectée. Nous ne pouvons pas exclure que Sin3a soit trouvée dans les RNP du fait de son interaction avec les protéines RbAp46 et RbAp48 qui font aussi partie du complexe PRC2.

### **Protéines impliquées dans le métabolisme des ARN :**

De nombreuses modifications de nucléotides telles que les 2'O-méthylations et les pseudouridylations sont formées au sein des ARNr. Nous avons détecté plusieurs des protéines impliquées dans la catalyse de ces modifications (NHP2, Fibrillarine et NOP56). Il est peu probable que leur détection corresponde au fait que la région des A-repeats subisse de telles modifications puisque ces modifications ont lieu dans le nucléole ou les corps de Cajal or l'ARN Xist ne transite pas par ces corpuscules nucléaires.

Nous avons détecté 2 peptides de la protéine DIMT1 dans les RNP formées avec la région des A-repeats entière en extrait nucléaire J4 (**Tab. III**). Cette protéine est une diméthyladénosine transférase putative. On pourrait donc envisager que la région des A-repeats subisse des méthylations au niveau de certaines de ces adénosines.

### **Autres protéines :**

Deux peptides de la protéine Rod1 (Regulator of differentiation 1) ont également été détectés dans les RNP formées avec la région des A-repeats entière en extrait nucléaire J4 (**Tab. III**). Cette protéine, découverte en 1999, est impliquée dans la régulation de la différenciation cellulaire (Yamamoto *et al.*, 1999). Rod1 est un paralogue de PTB, elle est majoritairement exprimée dans les cellules hématopoïétiques. Aucun rôle ne lui a encore été attribué. De façon intéressante, les embryons dans lesquels le gène codant Rod1 a été délété présentent une mortalité à E3,5, ce qui n'est pas incompatible avec une implication dans les mécanismes d'inactivation du chromosome X.

Notre détection de la protéine lin28 pourrait également être intéressante (4 peptides pour les RNP formées avec la région des A-repeats en extrait nucléaire de cellules ES J2) bien que nous l'ayons aussi détectée pour les RNP formées avec l'ARN-5'M-MS2. Cette protéine de 28 kDa est surexprimée dans les cellules ES en prolifération (Polesskaya *et al.*, 2007). Elle

est fortement exprimée chez l'embryon à E6,5 et son expression diminue au cours de la différenciation des cellules ES. Elle est donc exprimée dans la fenêtre de temps correspondant à l'initiation de l'inactivation aléatoire, ce qui souligne son recrutement potentiel par la région des A-repeats.

Le **Tableau V** présente la liste des protéines qui, d'après nos critères de sélection, paraissent être les plus pertinentes :

<b>hnRNP</b>	hnRNPU/SAF-A
	PTB
	KSRP
	FBP1
<b>Métabolisme de l'ADN</b>	SWI/SNF
	BAZ-1A
	Dnmt1
	Tif1b
	PARP
	histones
	PDS5A
	RIF1
	BRM1L
	Sin3a
	<b>Métabolisme de l'ARN</b>
Vigiline	
RNA hélicase A	
TIAR	
DEK	
DIMT1	
<b>Autres protéines</b>	Rod1
	Lin28

**Table V : Liste des protéines dont il serait intéressant d'étudier l'effet de l'extinction par ARN interférence sur l'inactivation du chromosome X**

Le seul moyen de vérifier l'implication de certaines de ces protéines dans le mécanisme d'inactivation du chromosome X sera de mettre au point un système de siRNA permettant d'éteindre l'expression des gènes de ces protéines dans les cellules ES de souris femelles, ceci durant un laps de temps assez long couvrant la fenêtre pendant laquelle l'inactivation est initiée par l'ARN Xist et ensuite, de suivre l'effet de l'ARN interférence sur l'inactivation du chromosome X. Ceci nécessite de longues mises au point que P Clerc réalise actuellement dans l'équipe de P Avner à l'institut Pasteur (Paris).

En attendant, nous avons décidé de réaliser un certain nombre d'expériences complémentaires permettant d'avoir des informations supplémentaires sur certaines des protéines identifiées qui avaient paru intéressantes d'après les données de la littérature.

Un premier point important était d'avoir des informations plus quantitatives sur leurs taux relatifs dans les différentes RNP étudiées. La méthode de Western-blot était la mieux adaptée à cet objectif.

Une autre information intéressante était la connaissance de la localisation des protéines étudiées au sein des noyaux de cellules ES. Une implication dans l'inactivation ou la maintenance de l'inactivation du chromosome X implique une co-localisation avec l'ARN Xist ou le domaine Xist. L'emploi d'approche d'immunofluorescence en microscopie confocale était adapté à ce second objectif.

Ces deux approches étaient tributaires de l'obtention d'anticorps dirigés contre les protéines étudiées qui soient fonctionnels en Western-blot mais aussi en imagerie.

Enfin, en ce qui concernait les 2 protéines PTB et KSRP que nous supposons interagir directement avec la région des A-repeats, nous pouvions débiter des études d'interaction avec la région des A-repeats par production de protéines recombinantes et transcription d'ARN *in vitro*.

Nous allons décrire ces analyses complémentaires dans le chapitre suivant.

## 2.2 Analyses complémentaires sur les protéines identifiées

### 2.2.1 Etude des taux relatifs de certaines protéines dans les différentes RNP étudiées par Western-blot

Nous avons pu réunir des anticorps pour les protéines suivantes : nucléoline, KSRP, PTB, DEK, hnRNP/SAF-A, PARP-1, FBP1, TIAR, Rod1, Sin3a, Lin28, Suz12, RbAp46, RbAp48, Eed et Ezh2. Leurs origines sont données dans le **Tableau I** de la partie Matériel et Méthodes.

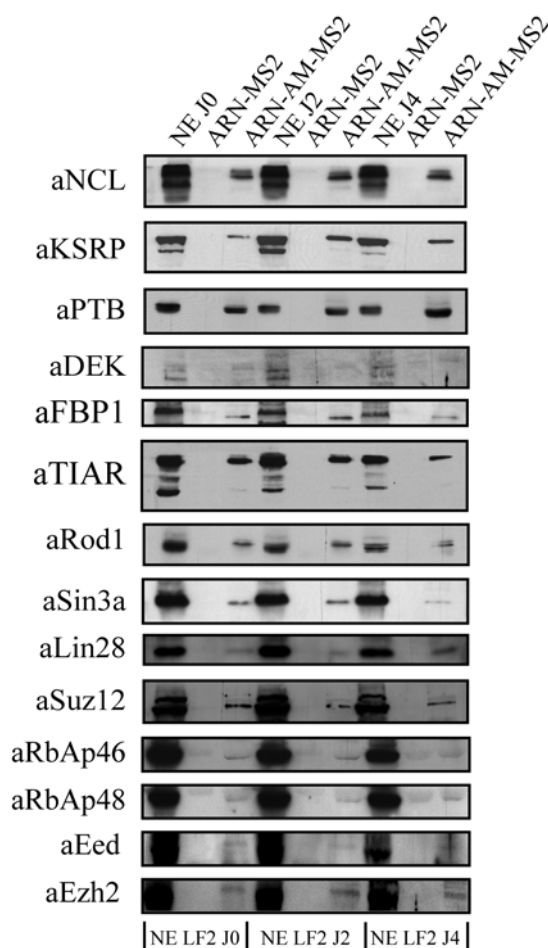
Dans un premier temps, les analyses ont été réalisées sur les RNP formées avec la région des A-repeats murine entière en extraits nucléaires de cellules ES, récoltées à différents stades de différenciation (J0, J2 et J4). D'abord, il est important de signaler que pour toutes les protéines testées, une détection par Western-blot a été obtenue, quel que soit l'extrait nucléaire (J0, J2 et J4) utilisé pour la formation des RNP.

Les protéines PTB, TIAR et à un moindre degré KSRP étaient présentes à un taux élevé dans les RNP en comparaison de leur taux dans les extraits nucléaires J0, J2 et J4 ce qui est en accord avec les données de spectrométrie de masse (**Fig. 65**). Nous avons noté une légère augmentation de KSRP dans les RNP formées avec les extraits de cellules en cours de différenciation et pour PTB, pour les extraits recueillis après 4 jours de différenciation, mais ces résultats ont besoin d'être vérifiés en répétant les expériences. Notons que Rod1 était détectée à un taux assez élevé dans tous les complexes, par contre les autres protéines identifiées étaient détectées à des taux plus faibles. Les protéines DEK, Sin3a, Lin28, nucléoline, FBP1 et les protéines du complexe PRC2 à l'exception de Suz12 étaient détectées



à un faible taux. Notons que la quantité détectée par Western-blot pour la protéine Suz12 était plus importante que pour celles détectées pour les autres protéines du complexe PRC2.

Ces expériences nous ont permis de vérifier la présence d'une partie des protéines détectées par spectrométrie de masse mais ne nous ont pas montré de différences majeures dans la présence de certaines d'entre elles selon le stade de différenciation des cellules ES à partir desquelles les extraits nucléaires avaient été préparés.



**Figure 65 : Analyse par Western-blot de protéines détectées dans les RNP formées avec la région des A-repeats de l'ARN Xist murin en extraits nucléaires de cellules ES après purification par chromatographie d'affinité**

L'ARN-MS2 correspond aux 3 motifs MS2. L'ARN-AM-MS2 correspond à la région des A-repeats murine entière. Les taux des protéines Nucléoline (NCL), KSRP, PTB, DEK, FBP1, TIAR, Rod1, Sin3a, Lin28, Suz12, RbAp46-48, Eed et Ezh2 ont été étudiés par la technique de Western-blot après fractionnement en gel SDS-PAGE des protéines des RNP purifiées par chromatographie d'affinité. Les RNP formées étaient formées dans un extrait nucléaire de cellules ES de souris non différenciées (NE LF2 J0), ou dans des extraits nucléaires préparés à partir de cellules ES récoltées après 2 jours (NE LF2 J2) ou 4 jours (NE LF2 J4) après initiation de la différenciation. Un échantillon d'extrait nucléaire NE correspondant à 5% de la quantité d'extrait nucléaire utilisée pour la formation des RNP a été déposé en parallèle.

Le même type d'expériences a ensuite été réalisé pour les RNP formées avec la région des A-repeats entière (ARN-AM-MS2) et pour ses sous-fragments correspondant à 1, 2 et 4 répétitions (ARN-1R-MS2, ARN-2RM-MS2, ARN-SLS1M-MS2) en extrait nucléaire de cellules ES non différenciées. Les résultats obtenus pour le complexe PRC2 et pour PTB ont été donnés dans l'article. Nous donnerons ici les résultats correspondant aux autres protéines pour lesquelles nous disposons d'anticorps (**Fig. 66**).

Pour deux des protéines testées, la nucléoline et DEK, une fixation était détectée pour tous les ARN testés sauf l'ARN-MS2. En d'autres termes, un motif répété était suffisant pour assurer leur présence dans les RNP purifiées (**Fig. 66**).

Les résultats obtenus pour Lin28 doivent être confirmés car une fixation était détectée pour une répétition et pas pour 2. Il faut souligner que comme elle n'avait pas été détectée dans d'autres RNP étudiées au laboratoire, elle reste une protéine dont l'étude mérite d'être approfondie. En spectrométrie de masse, elle n'a été détectée qu'avec la région des A-repeats entière, il semble en effet qu'elle soit plus présente dans les complexes formés avec la région des A-repeats entière (**Tab. III**, et avec l'ARN-5'M-MS2 **Tab. IV**).

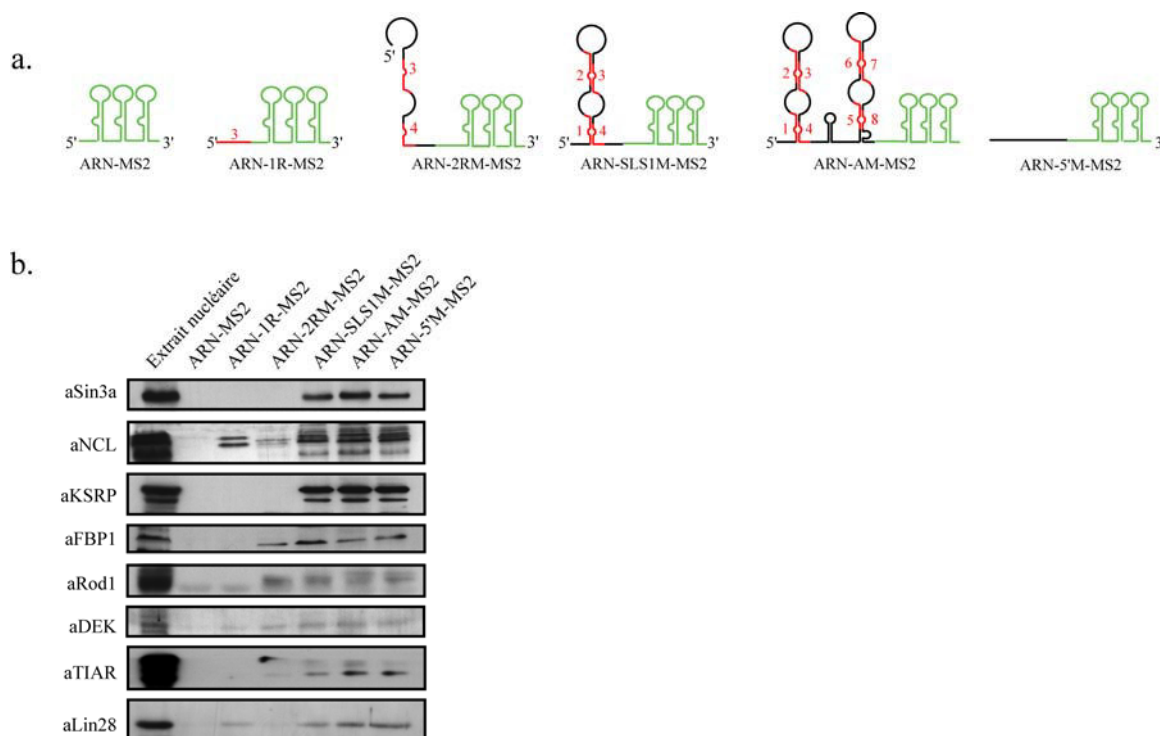
L'analyse par Western-blot montrait aussi que 3 des protéines avaient la capacité de s'associer lorsqu'au moins 2 éléments répétés et leurs espaceurs étaient présents : FBP1, Rod1 et TIAR, ce qui suggère que l'association de ces protéines nécessite sans doute la présence d'au moins une séquence simple brin riche en résidus U correspondant à une séquence espaceur (**Fig. 66**). Notons que pour TIAR, le taux de protéine est très faible pour l'ARN-2RM-MS2 et qu'il augmente fortement lorsque la région des A-repeats est entière. En spectrométrie de masse, FBP1 n'a été détectée que pour des ARN contenant au moins 4 éléments répétés (**Tab. IV**).

Enfin les 2 autres protéines testées, KSRP et Sin3a, n'étaient détectées par Western-blot que dans des complexes formés avec un ARN contenant au moins 4 éléments répétés et formant la grande structure tige-boucle SLS1 (**Fig. 66**). Ces données sont en accord avec la détection de ces protéines par spectrométrie de masse uniquement dans les RNP formées avec les ARN-SLS1M-MS2 et ARN-AM-MS2 (**Tab. IV**).

En ce qui concerne KSRP que nous supposons pouvoir intervenir dans la stabilisation de la structure 2D de l'ARN, il s'agissait d'une information très intéressante méritant une étude approfondie basée sur l'emploi de la protéine recombinante.

Enfin, la grande spécificité de Sin3a vis-à-vis du type de structure adoptée par l'ARN est très intéressante et renforce l'idée qu'elle pourrait avoir un rôle possible dans l'activité de la région des A-repeats. Rappelons que Sin3a est impliquée dans la régulation transcriptionnelle en tant que protéine plateforme de recrutement d'histone désacétylase (HDAC).

L'analyse par Western-blot montrait également que toutes les protéines testées étaient retrouvées dans les RNP formées avec l'ARN-5'M-MS2. Ceci était en accord avec les données de spectrométrie de masse sauf pour les protéines Sin3a et TIAR qui n'étaient pas détectées dans les RNP formées avec l'ARN-5'M-MS2. Ces observations renforcent le fait de réaliser les analyses complémentaires par Western-blot.



**Figure 66 : Analyse par Western-blot de protéines contenues dans les RNP formées par la région des A-repeats, ses sous-fragments ou l'extrémité 5' terminale de l'ARN Xist murin dans un extrait nucléaire de cellules ES non différenciées, complexes qui ont été purifiés par chromatographie d'affinité MS2/MBP**

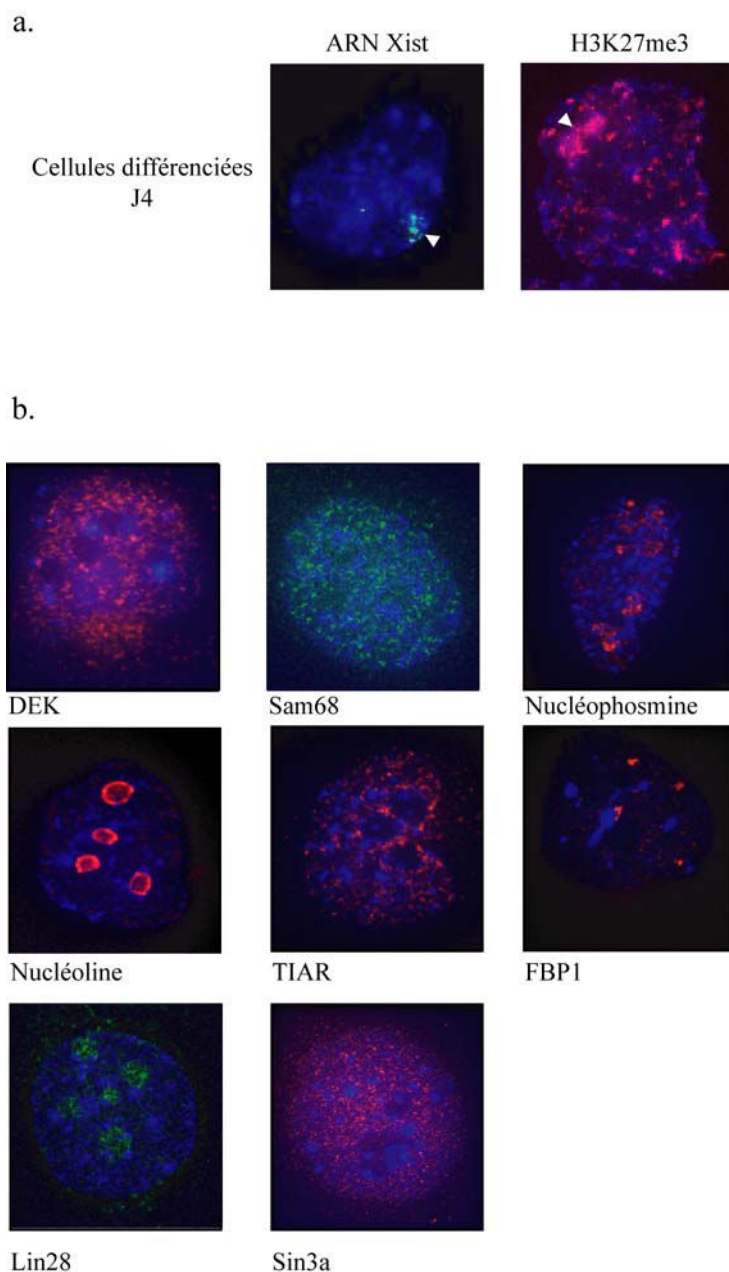
a. Représentation schématique des ARN utilisés lors de cette expérience. L'ARN-MS2 utilisé comme contrôle contient les 3 motifs MS2. Les ARN-1R-MS2, ARN-2RM-MS2, ARN-SLS1M-MS2 et ARN-AM-MS2 comportent respectivement 1, 2, 4 et 8 éléments répétés. L'ARN-5'M-MS2 correspond à l'extrémité 5' terminale de l'ARN Xist murin.

b. La présence des protéines Sin3a, Nucléoline (NCL), KSRP, FBP1, Rod1, DEK, TIAR et Lin28 a été testée par Western-blot.

### 2.2.2 Etude de la localisation intra-cellulaire de certaines des protéines identifiées par immunofluorescence

Si les protéines identifiées avaient bien un lien avec l'activité de la région des A-repeats, on pouvait s'attendre à observer des changements de localisation d'au moins une partie de ces protéines en fonction du stade de différenciation et des caractéristiques sexuelles des cellules ES utilisées. D'autre part, on pouvait imaginer qu'une partie de certaines de ces protéines soit localisée dans le domaine Xist après différenciation (J4). Lors de mes séjours au laboratoire de P Avner (Institut Pasteur), j'ai appris à cultiver et à différencier les cellules ES sur lames, puis à réaliser des études de localisation de protéines par immunofluorescence. Ce travail réalisé avec P Clerc a nécessité dans un premier temps la mise au point des conditions expérimentales d'utilisation des anticorps. Nous avons testé plusieurs dilutions des anticorps s'échelonnant de 1/50 à 1/2000. Les anticorps dirigés contre les protéines suivantes (DEK, Sam68, nucléophosmine, nucléoline, TIAR, FBP1, Lin28 et Sin3a) se sont avérés satisfaisants

pour des études de localisation par immunofluorescence. Le protocole retenu est décrit dans la partie Matériel et Méthode. Une fois les conditions expérimentales déterminées, nous avons pu comparer la localisation des protéines étudiées au cours de la différenciation des cellules ES. En parallèle, j'ai réalisé des localisations du Xi ainsi que des H3K27me3 sur d'autres cellules ES (**Fig. 67a**). La **Figure 67b** présente les résultats obtenus pour les 8 protéines étudiées. Plus d'une centaine de cellules ont été examinées pour chaque protéine étudiée et pour chaque stade de la différenciation. Nous avons observé comme attendu que les protéines Lin28, Nucléoline et Nucléophosmine avaient une localisation majoritairement nucléolaire. Les protéines TIAR, Sin3a, DEK et Sam68 avaient une localisation diffuse dans le nucléoplasme avec une exclusion des nucléoles. La protéine FBP1 semblait être majoritairement localisée au niveau des corps de Cajal. Bien que des co-localisations n'aient pas été réalisées avec l'ARN Xist ou avec les H3K27me3, il ne semblait pas que nous ayons une forte concentration d'une partie de ces protéines dans une zone qui puisse correspondre au domaine Xist. De plus, aucun changement de localisation de ces protéines n'était observé aussi bien au cours de la différenciation des cellules ES, qu'en fonction des caractéristiques sexuelles des cellules utilisées. C'est d'ailleurs pourquoi, seules les images recueillies 4 jours après différenciation sont présentées. Ces travaux préliminaires nous ont néanmoins permis de tester si les anticorps que nous possédions étaient utilisables pour ce type de localisation. Seules des expériences de localisation de ces protéines par immunofluorescence devront être répétées en les couplant à la détection de l'ARN Xist par RNA-FISH.



**Figure 67 : Etude de localisation de 8 protéines présentes dans les RNP au sein de cellules ES de souris 4 jours après l'initiation de la différenciation**

a. Localisation du domaine Xist par détection de l'ARN Xist (RNA-FISH) et par immunolocalisation des marques H3K27me3 dans des cellules ES de souris 4 jours après l'initiation de la différenciation. Les 2 images présentent le noyau de 2 cellules différentes.

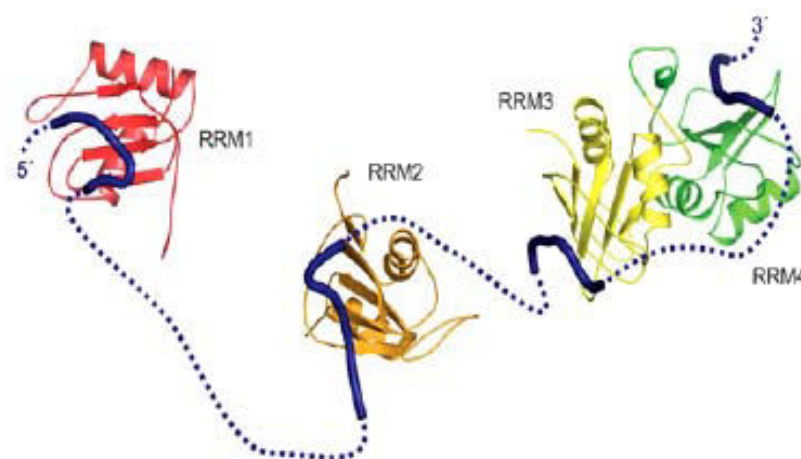
b. Immunolocalisation des protéines DEK, Sam68, Nucléophosmine, Nucléoline, TIAR, FBP1, Lin28, Sin3a dans des cellules ES 4 jours après initiation de la différenciation

### 3. Etude de l'interaction possible de PTB et KSRP avec la région des A-repeats

L'ensemble de nos données était en accord avec un rôle possible de PTB et KSRP dans la structure secondaire de la région des A-repeats. Avant de décrire les études d'interaction que nous avons réalisées, nous donnerons quelques informations sur l'architecture de ces protéines.

- PTB

Comme nous l'avons déjà mentionné, la protéine PTB est impliquée dans de nombreux mécanismes 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). Son poids moléculaire est de 57 kDa. Elle renferme 4 domaines de fixation aux 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 ; Petoukhov *et al.*, 2006). 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 (**Fig. 68**) (Oberstrass *et al.*, 2005 ; Simpson *et al.*, 2004 ; Petoukhov *et al.*, 2006). Par la fixation de ces multiples domaines de fixation 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 (Mitchell *et al.*, 2003 ; Petoukhov *et al.*, 2006). Ces données renforcent l'hypothèse que nous avons émise concernant l'implication possible de la protéine PTB dans la stabilisation de la structure 2D de la région des A-repeats.



**Figure 68 : Modèle de structure de la protéine PTB et interaction avec les ARN**

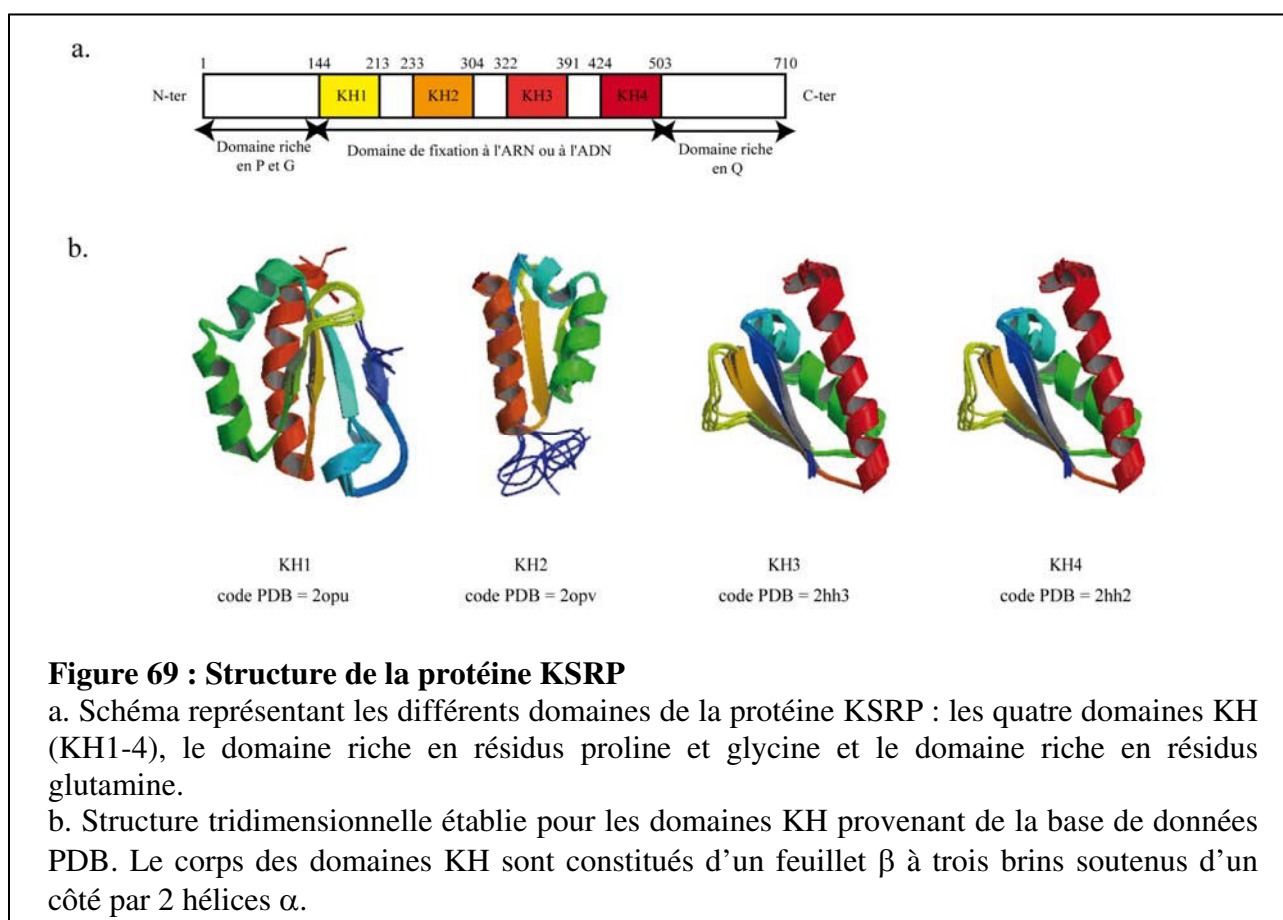
La structure et l'arrangement spatial des domaines RRM de PTB sont montrés. La conformation de l'ARN lié à chaque domaine RRM est représentée par une ligne bleue. Les domaines RRM sont constitués d'un feuillet de quatre brins  $\beta$  soutenu de chaque côté par une hélice  $\alpha$ .

D'après les revues, Sawicka *et al.*, 2008 et Cléry *et al.*, 2008

- KSRP

Le poids moléculaire de la protéine KSRP humaine est de 73 kDa. Elle comporte 4 domaines KH (hnRNP-K homology, KH 1-4), qui sont des domaines de fixation à l'ARN. Elle contient également une extrémité N-terminale riche en résidus proline et glycine et une extrémité C-terminale riche en résidus glutamine (**Fig. 69a**). Les structures des quatre domaines KH de la protéine KSRP sont représentées dans la **Figure 69b**. Elle a été identifiée comme un constituant du complexe protéique régulant l'épissage de l'ARN pré-messager c-src (Min *et al.*, 1997). KSRP est aussi impliquée dans la régulation de la dégradation des ARNm comportant des séquences riches en U et A (région ARE)(Gherzi *et al.*, 2004). Les domaines KH3-KH4 sont essentiels à la reconnaissance des ARNm et au recrutement de la PARN (Poly(A)-specific ribonuclease) et de l'exosome (Garcia-Mayoral *et al.*, 2007 ; Gherzi *et al.*, 2004) qui dégradent les ARNm. Récemment, il a été démontré que la protéine KSRP fait également partie des complexes Dicer et Drosha et régule la biogenèse des miRNA dans les cellules eucaryotes (Trabucchi *et al.*, 2009). Les séquences de reconnaissance AUUUA de la protéine KSRP sont généralement situées dans de grandes régions riches en U comme c'est le cas pour les séquences espaceurs de la région des A-repeats. Elle pourrait donc, au même titre que PTB, stabiliser les structures tige-boucle en s'y fixant.

Nous avons donc décidé de réaliser une étude approfondie de l'interaction de PTB et KSRP avec la région des A-repeats.



**Figure 69 : Structure de la protéine KSRP**

a. Schéma représentant les différents domaines de la protéine KSRP : les quatre domaines KH (KH1-4), le domaine riche en résidus proline et glycine et le domaine riche en résidus glutamine.

b. Structure tridimensionnelle établie pour les domaines KH provenant de la base de données PDB. Le corps des domaines KH sont constitués d'un feuillet  $\beta$  à trois brins soutenus d'un côté par 2 hélices  $\alpha$ .

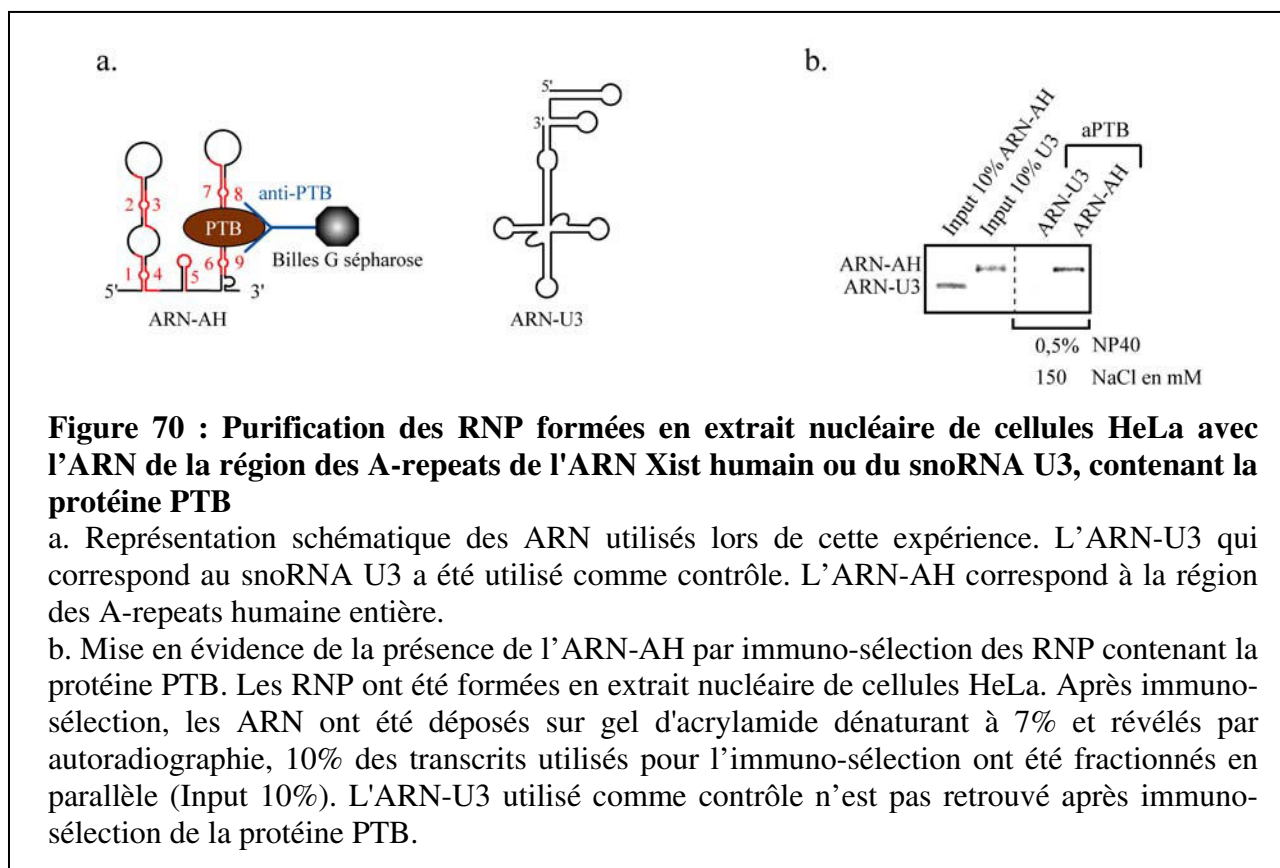


### 3.1 PTB

#### 3.1.1 Vérification par l'approche d'immuno-sélection de l'interaction de PTB avec la région des A-repeats

Comme nous l'avons fait pour les protéines du complexe PRC2, nous avons utilisé des anticorps anti-PTB fixés sur des billes de G sépharose pour retenir les complexes ARN-protéines formés en extrait nucléaire de cellules HeLa et contenant la protéine PTB. Pour cela, les ARN ont été produits par transcription *in vitro* sous forme radioactive puis incubés en présence d'un extrait nucléaire de cellules HeLa dans les mêmes conditions que celles décrites dans l'article.

Le snoRNA U3 a servi de contrôle négatif. Nous avons pu ainsi montrer qu'en présence de NaCl 150 mM et de 0,5% de NP40 en tampon Tris-HCl pH 8, nous avons bien une rétention de l'ARN-AH correspondant à la région des A-repeats humaine entière mais pas du snoRNA U3 (**Fig. 70**).



#### 3.1.2 Vérification de l'existence d'une interaction directe entre PTB et la région des A-repeats au sein d'un extrait nucléaire par pontage aux UV

Lorsqu'une protéine interagit directement avec un ARN, on peut généralement former un ou des pontages covalents entre cet ARN et la protéine par traitement aux UV à 254 nm. Nous avons pour cela utilisé une source d'UV délivrant une énergie de 3000 J/cm<sup>2</sup> (Biolink). L'intérêt de cette approche est aussi qu'en comparant les taux de pontages formés dans différents extraits nucléaires, on peut avoir une information sur le taux relatif d'association de

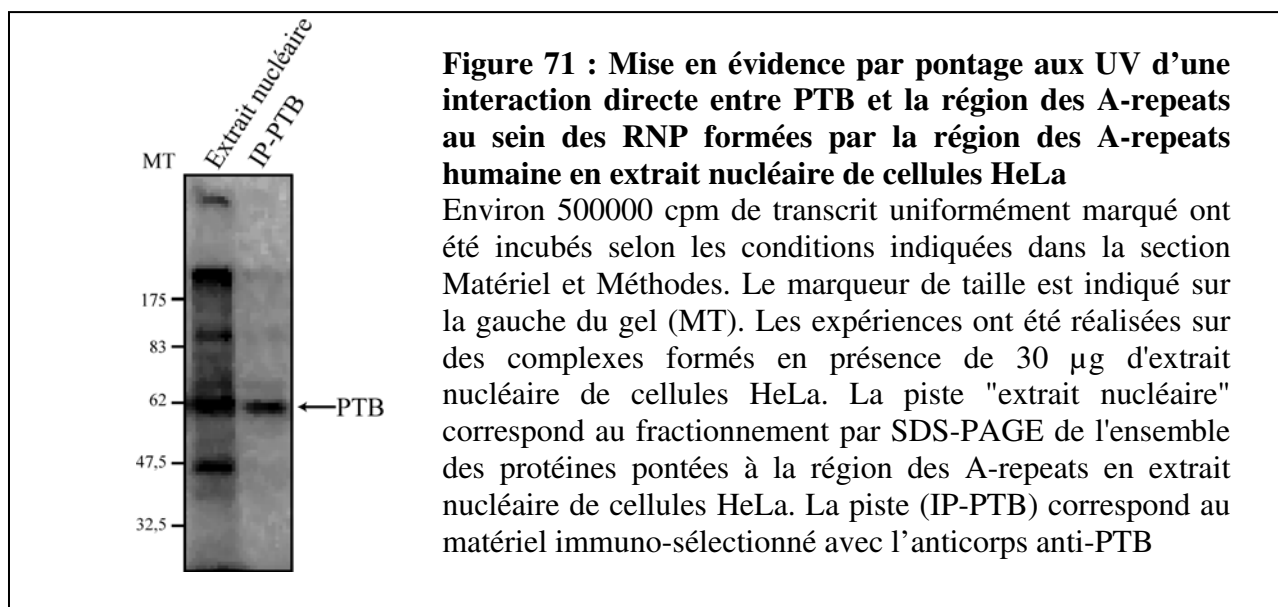


la protéine étudiée suivant l'identité de l'ARN étudié et du type d'extrait nucléaire utilisé pour former les complexes. C'est donc une méthode permettant d'obtenir des données quantitatives, ce que nous n'avons pas avec les méthodes de spectrométrie de masse utilisées.

La mise au point de l'approche a été réalisée sur les complexes formés par la région des A-repeats humaine entière (ARN-AH radiomarqué), incubée en présence d'un extrait nucléaire de cellules HeLa. Les complexes formés ont été soumis à un pontage aux UV dans une plaque 96 puits dans les conditions décrites dans la partie Matériel et Méthodes. Afin de ne conserver que de courts fragments d'ARN radioactifs pontés aux protéines PTB, les échantillons ont été soumis à l'action de la RNase A dans les conditions décrites dans la partie Matériel et Méthodes. Pour tester le pontage de PTB à l'ARN, les 9/10<sup>ième</sup> du matériel hydrolysé ont ensuite été incubés avec des billes G sépharose auxquelles l'anticorps anti-PTB avait été fixé. Après lavage, le matériel retenu sur les billes a été élué, puis fractionné par électrophorèse en gel SDS-PAGE à 10%. La présence de protéines PTB pontées à des nucléotides radioactifs a été révélée par autoradiographie.

Le 10<sup>ième</sup> du matériel restant, résultant de l'hydrolyse par la RNase A, a été soumis directement au fractionnement par électrophorèse en parallèle du matériel élué des billes de sépharose portant l'anticorps anti-PTB. Ceci nous a en plus permis d'avoir une idée de la masse moléculaire et du nombre de protéines pontées à la région des A-repeats au sein d'un extrait nucléaire de cellules HeLa. Ce nombre donne aussi une idée du nombre de protéines au sein des RNP interagissant directement avec la région des A-repeats (**Fig. 71**). Le profil de fractionnement obtenu comportait 5 bandes de forte intensité que l'on pouvait estimer avoir 47.5, 62, 68, 83 et 250 kDa. Il était toutefois difficile de tirer des conclusions sur leur identité uniquement à partir de ces masses moléculaires. Nous pouvions néanmoins présumer de l'identité de certaines protéines. La bande à 250 kDa pouvait correspondre à la protéine Rif1 (274 kDa). Nous supposons que la bande au environ de 62 kDa pouvait être PTB (57-59 kDa), ce qui a pu être vérifié par le fait que dans la piste du gel correspondant au matériel immuno-sélectionné par l'anticorps anti-PTB (**Fig. 71**, piste anti-PTB), nous avons une bande au même niveau. Nous confirmons bien qu'au sein d'un extrait nucléaire, PTB est capable de se lier directement à la région des A-repeats.

Cette méthode pourra être utilisée avec un certain nombre d'anticorps dont nous disposons afin de mettre en évidence des interactions directes des protéines avec la région des A-repeats et d'essayer d'identifier celles que nous avons détectées au cours des expériences de pontage aux UV. Par ailleurs, nous envisageons, en collaboration avec l'équipe du Dr Urlaub, d'identifier les interactions ARN-protéines directes par pontages aux UV suivies d'une analyse par spectrométrie de masse. H Urlaub avec qui nous collaborons, a mis au point une méthode permettant de détecter les protéines pontées covalamment à une molécule d'ARN.



### 3.1.3 Clonage, production et purification de la protéine PTB recombinante

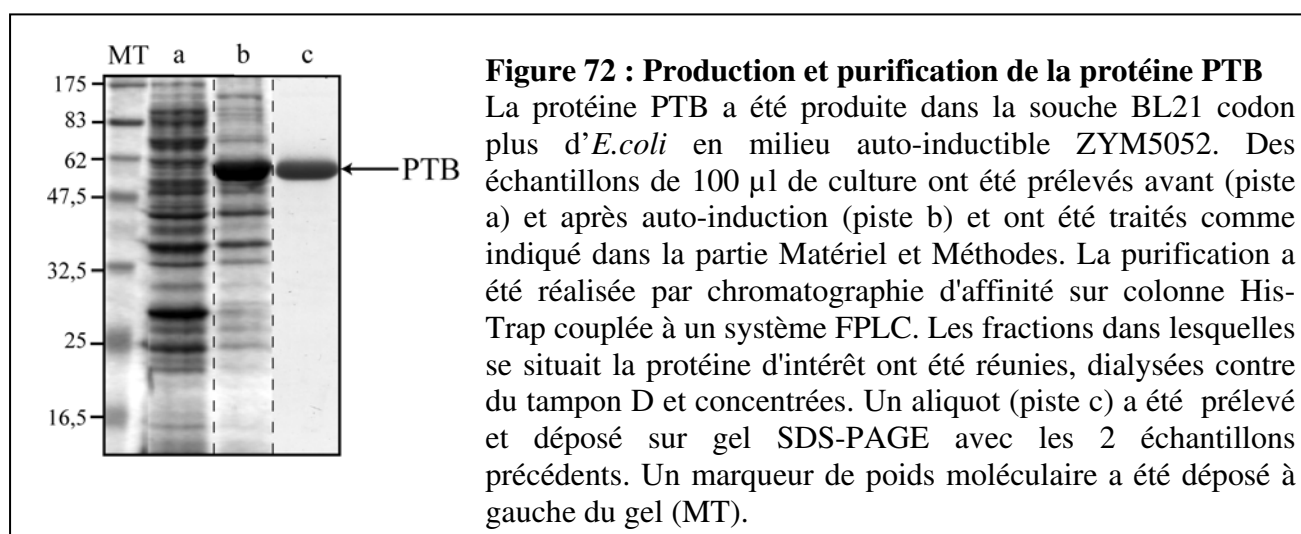
Pour pouvoir réaliser une étude directe de l'interaction entre la protéine PTB et la région des A-repeats, nous devons disposer de la protéine recombinante. J'ai donc entrepris le clonage de l'ORF PTB dans un vecteur d'expression permettant son expression chez *E. coli*, puis sa purification.

- Clonage, production et purification de la protéine PTB

Les protéines PTB de souris et d'homme sont très homologues, elles présentent 92% d'identité. Dans les deux espèces, 3 isoformes sont produits par épissage alternatif (PTBP1, PTBP2 et PTBP4). PTBP1 (531 acides aminés) est l'isoforme la plus courte, PTBP2 et PTBP4 contiennent, respectivement, 19 et 26 acides aminés supplémentaires localisés entre les domaines RRM2 et RRM3 (pour revue, Sawicka et al., 2008). Les expériences de Western-blot, que nous avons réalisées sur les RNP purifiées aussi bien à partir d'extraits nucléaires de cellules ES de souris que de cellules HeLa, montraient que les isoformes étaient présentes dans les RNP purifiées. Nous pouvions donc choisir de produire l'une ou l'autre des protéines. Pour des raisons de disponibilité de clone, nous avons travaillé avec la plus petite des 3, c'est-à-dire PTBP1 (531 acides aminés, 57 kDa).

Nous avons utilisé un plasmide fourni par la société RZPD (Hs. 172550) qui renfermait l'ADNc codant la protéine PTBP1 humaine. Nous avons amplifié la région codant la protéine complète en utilisant des amorces permettant d'insérer des sites de restriction compatibles avec le sous-clonage dans le vecteur d'expression pET28b (*Bam*HI). Ce vecteur permettait de produire la protéine recombinante fusionnée à son extrémité N-terminale avec une étiquette «6xHis», ceci sous le contrôle d'un promoteur T7. Nous avons choisi la souche d'*E. coli* BL21 codon<sup>+</sup> comme souche hôte.

Après plusieurs tests d'expression, nous avons pu établir le protocole expérimental suivant : les bactéries transformées avec le plasmide d'expression pET28a-PTB sont cultivées en milieu auto-inductible (ZYM5052), à 37°C, pendant une nuit (Studier, 2005). Ce procédé repose sur l'emploi d'un milieu contenant plusieurs sources de carbone, dont le glucose et le lactose. Il présentait l'avantage de ne pas nécessiter d'inducteur externe. En effet, la population bactérienne consomme le lactose lorsque le glucose est épuisé. Cette consommation du lactose permet l'induction de l'expression de l'ARN polymérase T7 placée sous le contrôle du promoteur LacUV5. Le gène d'intérêt placé sous contrôle du promoteur T7 est alors exprimé. La **Figure 72** présente les résultats de sur-expression obtenus. On distingue dans la piste (b), correspondant à la culture récupérée après induction, la présence d'une bande de forte intensité d'une masse moléculaire correspondant à celle attendue pour la protéine PTBP1. Cette bande était absente dans l'extrait témoin non induit (a), ce qui démontrait l'efficacité de l'induction. Après centrifugation et sonication des cellules comme décrit dans la partie Matériel et Méthodes, la protéine PTB a été enrichie par purification sur colonne Ni-sépharose (FPLC). Après concentration de la protéine PTB, nous avons obtenu une quantité de 3 mg de protéine PTB pure pour 1 litre de culture. La qualité de l'échantillon protéique a été vérifiée par SDS-PAGE (**Fig. 72**).

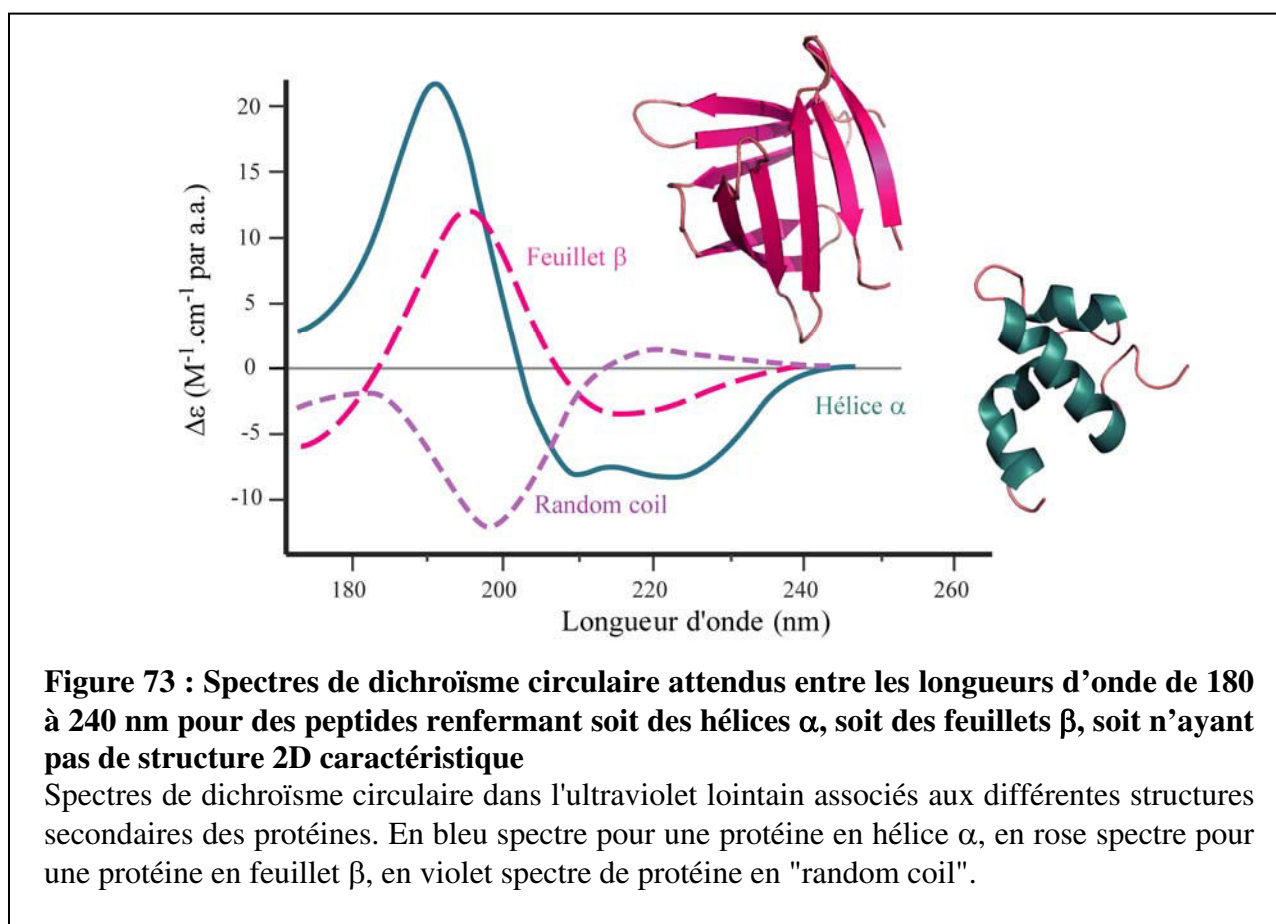


- Contrôle de la qualité de l'échantillon protéique par dichroïsme circulaire
  - i. Principe de la méthode

L'emploi du dichroïsme circulaire permet d'obtenir des informations globales sur le degré et la nature des structures secondaires des protéines. Nous verrons par la suite que cette méthodologie est également applicable à l'étude de la structure 2D des ARN.

Par leurs asymétries, les structures secondaires (hélice  $\alpha$ , feuillet  $\beta$ , « random coiled ») ont la propriété d'absorber différemment la lumière selon qu'elle est polarisée circulairement à droite ou polarisée circulairement à gauche. L'absorbance mesurée par l'appareil de

dichroïsme circulaire (dichrographe) correspond à la différence de ces deux composantes : l'absorption de la lumière polarisée circulairement à droite et l'absorption de la lumière polarisée circulairement à gauche. Cette différence d'absorption mesurée par le dichrographe entre 170 nm et 330 nm (UV lointain) est ensuite normalisée en tenant compte de la concentration de l'échantillon ( $\Delta\epsilon$ ). Pour les protéines, chaque type de structure secondaire correspond à un spectre de dichroïsme circulaire caractéristique dans l'UV lointain entre 180 et 240 nm. La **Figure 73** présente les spectres de dichroïsme circulaire entre 180 et 240 nm correspondant respectivement aux hélices  $\alpha$  et aux feuillets  $\beta$  des protéines. Le spectre des hélices  $\alpha$  se caractérise par la présence d'un pic positif à 195 nm et de deux pics négatifs aux voisinages de 208 et 222 nm. Celui des feuillets  $\beta$  (rose) présente un pic positif à 198 nm et un pic négatif à 215 nm. Les protéines en "random coil" (violet) ont des spectres présentant un pic négatif à 196 nm. Le spectre obtenu pour une protéine comportant différents types de structures secondaires correspond à la somme des contributions de chaque structure secondaire présente. Dans notre cas, cette méthode a permis d'évaluer la structuration secondaire de la protéine PTB recombinante purifiée.



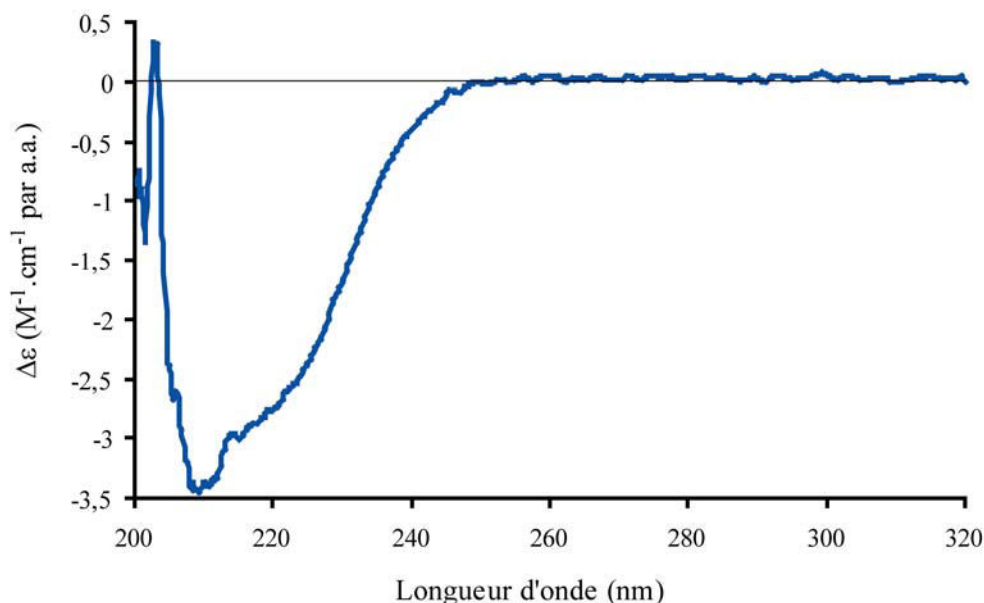
De plus, le dichroïsme circulaire nous a permis d'apprécier si nous nous étions bien débarrassés des acides nucléiques au cours des purifications. En effet, au même titre que les protéines, les acides nucléiques absorbent la lumière et ont la propriété d'être chiraux. Ils sont

donc susceptibles d'avoir un signal analysable par dichroïsme circulaire. En effet, la forme B de l'ADN en double hélice droite présente un spectre dichroïque avec une bande négative à 250 nm et une bande positive autour de 270 nm. Celui de la forme A présente une bande positive vers 260 nm alors que la bande négative à 250 nm est plus faible que dans le cas des hélices de forme B. La forme Z non naturelle présente quant à elle un spectre avec une bande positive à 260 nm et une bande négative à 290 nm. Pour l'ARN, le dichroïsme circulaire a démontré que lorsqu'il est structuré en une double hélice, il adopte préférentiellement une hélice de forme A.

Pour résumer, les acides nucléiques ont un spectre dichroïque caractéristique dans la région de 200 à 320 nm avec un signal positif de 240 à 320 nm. Les protéines quant à elles ont un signal négatif de 190 à 240 nm. De ce fait, entre 240 nm et 320 nm, seuls les acides nucléiques présentent un signal (notons tout de même que les acides aminés aromatiques et les ponts disulfures peuvent avoir également un signal dichroïque à ces longueurs d'ondes, mais ce signal est faible en comparaison de celui des ARN). Ainsi, un échantillon protéique dépourvu de tout contaminant correspondant aux acides nucléiques ne présente pas de signal dichroïque dans la zone de 240 à 320 nm. En plus de la mesure du rapport d'absorbance A260/A280, le dichroïsme circulaire permet donc de vérifier l'absence d'acides nucléiques. Par ailleurs comme nous le verrons plus loin, ce non-chevauchement des régions spectrales ARN/protéine nous a permis d'analyser le signal dichroïque de chacune des 2 composantes dans un même mélange d'ARN et de protéines, afin de rechercher si des changements de structuration de l'ARN avaient lieu lors de la formation de complexes RNP.

## ii. Résultats obtenus

Pour obtenir des spectres exploitables, nous avons utilisé 500  $\mu$ l de protéine PTB recombinante à une concentration de 0,2  $\mu$ M dans une cuve de 0,875 mm de trajet optique. Le spectre obtenu (**Fig. 74**) présentait deux pics négatifs à 208 et 222 nm ce qui est caractéristique d'une structuration majoritaire en hélices  $\alpha$  et en feuillets  $\beta$ . Ces résultats étaient en accord avec les structures 3D établies pour les domaines RRM présents au nombre de 4 dans cette protéine (**cf Fig. 68**). La protéine PTB semblait donc bien structurée. De plus, l'absence de signal dichroïque positif entre 240 et 320 nm démontrait l'absence d'acides nucléiques dans l'échantillon. Nous avons donc entrepris la réalisation d'expériences de retard sur gel et d'empreinte aux RNases avec la protéine PTB recombinante et la région des A-repeats.



**Figure 74 : Spectre de dichroïsme circulaire de la protéine PTB enregistré entre 200 et 320 nm**

Le spectre a été enregistré en plaçant l'échantillon dans une cuve de 0,875 cm de trajet optique, dans une solution de tampon Tris 10 mM, NaCl 150 mM, pH 7,5 filtrée à 0,22  $\mu\text{m}$ , à une concentration en protéine PTB de 0,2  $\mu\text{M}$ . Les mesures de  $\Delta$  d'absorbance ont été converties en  $\Delta\epsilon$  par acides aminés, comme expliqué dans la section Matériel et Méthodes.

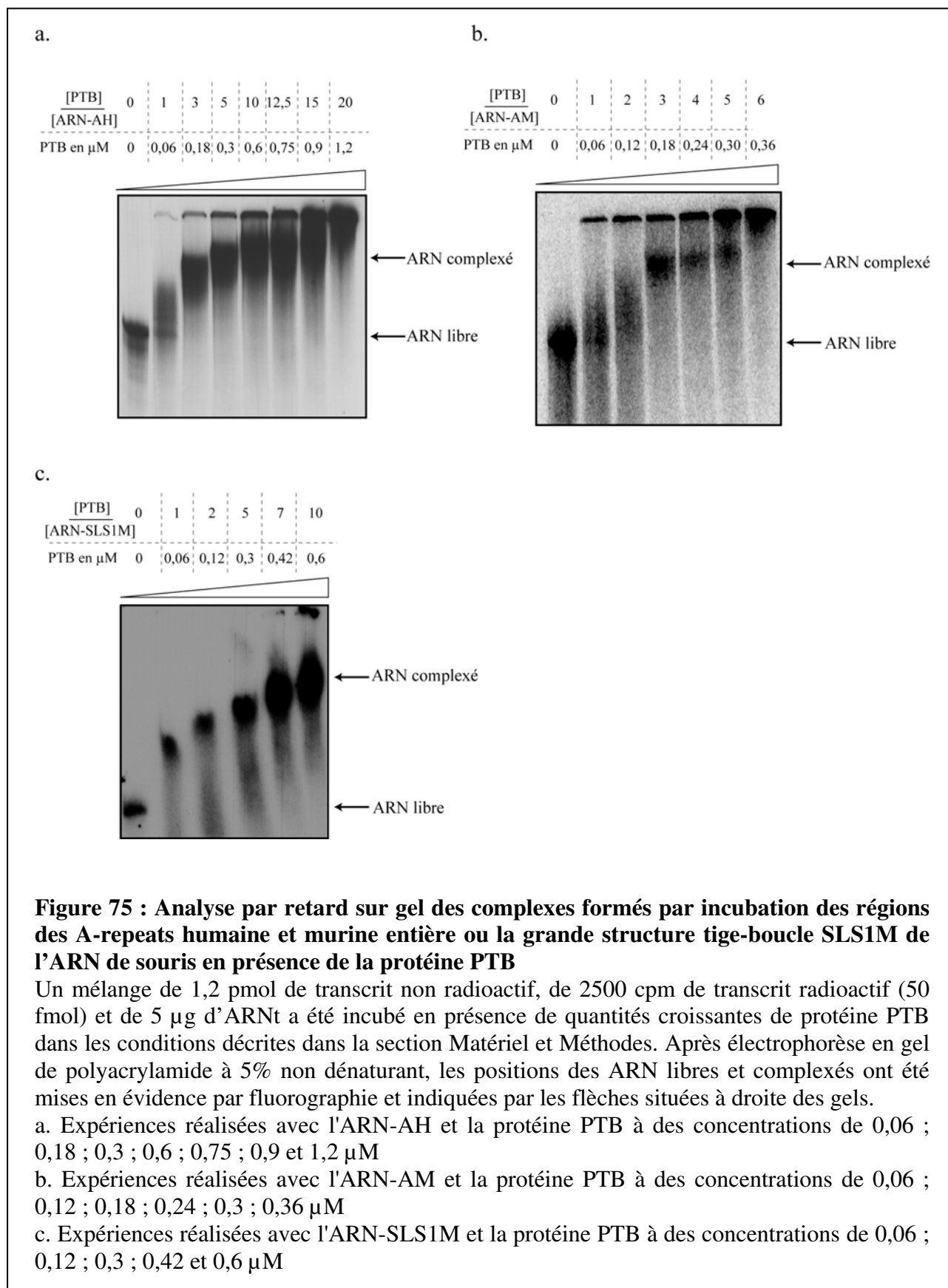
### 3.1.4 Expériences de retard sur gel

Les expériences ont été réalisées à la fois sur la région des A-repeats humaine et sur celle murine. Nous avons aussi testé la fixation de PTB sur l'ARN-SLS1M. L'objectif était double : confirmer le caractère direct de la fixation et déterminer les conditions à utiliser pour réaliser les expériences d'empreinte aux RNases. Nous avons d'abord défini les concentrations relatives en ARN et en protéine requises pour obtenir une complexation totale de l'ARN. Un mélange de 1,2 pmoles d'ARN non radio-marqué, d'une quantité très faible de ce même ARN radioactif (0,5 fmol) et de 5  $\mu\text{g}$  d'ARNt a été incubé dans 20  $\mu\text{l}$ , en présence de quantités croissantes de protéines, ceci dans les conditions décrites dans la partie Matériel et Méthodes. Après électrophorèse en gel de polyacrylamide à 5% non-dénaturant, les positions des ARN libres et des ARN complexés ont été mises en évidence par autoradiographie.

Aussi bien pour les ARN humain que murin, nous avons remarqué qu'en présence de 3,6 pmoles de protéines, c'est-à-dire pour un excès molaire de 3 fois de protéine par rapport à l'ARN, toutes les molécules d'ARN étaient complexées (**Fig. 75**). A une concentration supérieure, plusieurs molécules de PTB devaient s'associer à l'ARN, ce qui expliquait les retards de migration de plus en plus importants observés. Nous notons aussi une apparition

d'agrégats au niveau des poches aux concentrations les plus élevées dans le cas de la région des A-repeats entière.

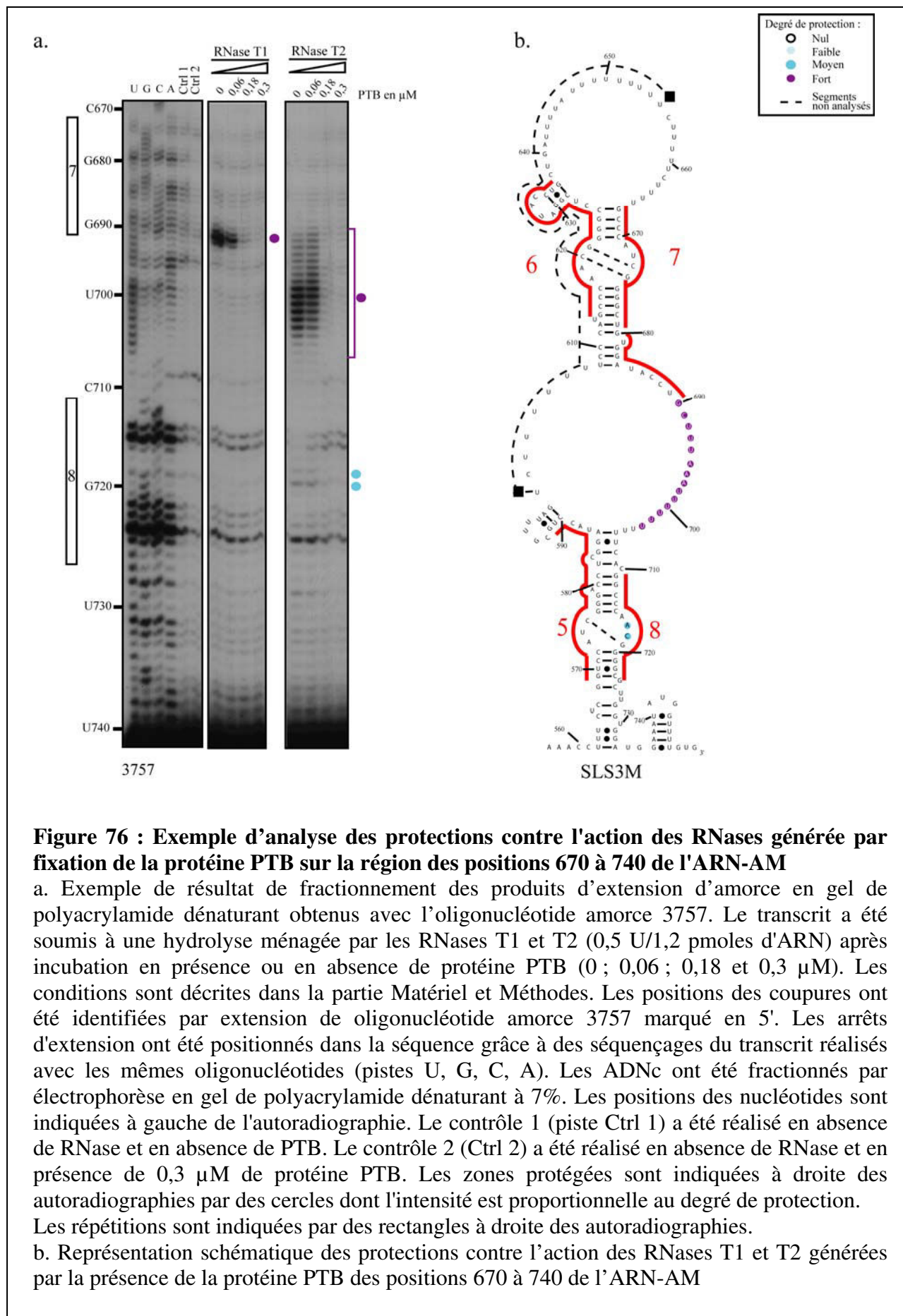
Une affinité tout aussi forte de PTB était observée pour la grande structure tige-boucle de l'ARN de souris SLS1M.



### 3.1.5 Etude par empreinte aux RNases T1 et T2 des sites de fixation de la protéine PTB 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 PTB sur la région des A-repeats murine entière. Vu la forte affinité de la protéine recombinante humaine pour l'ARN-AM de souris et sa forte similitude avec la protéine murine (92%), ce choix de l'ARN-AM était justifié. Comme nous avons observé par retard sur gel que des complexes contenant un nombre croissant de protéines PTB étaient formés lorsque l'ARN-AM était incubé en présence de concentrations croissantes en protéines PTB, les hydrolyses ont été réalisées sur des complexes formés à trois ratios différents [protéine]/[ARN] : 1, 3 et 5. Pour cela, 1,2 pmoles d'ARN-AM ont été incubés en présence de 0 ; 1,2 ; 3,6 et 6 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 protocole expérimental utilisé était identique à celui utilisé pour les expériences d'empreinte en extrait nucléaire de cellules HeLa. Ces expériences n'ont été réalisées qu'une seule fois. Il sera nécessaire de les reproduire pour tirer des conclusions définitives. Un exemple de fractionnement des produits d'extension d'amorce en gel de polyacrylamide dénaturant obtenus avec l'oligonucléotide amorce 3757 qui permet d'analyser le segment compris entre les positions 670 à 740 contenant la répétition 8 et le segment la reliant à la répétition 7 est illustré dans la **Figure 76a**. Les protections détectées à partir de ces extensions d'amorce sont représentées sur la structure SLS3M dans la **Figure 76b**. Comme nous l'attendions, une grande partie de la région en simple brin des positions 690 à 703 était protégée par la protéine PTB à partir d'un ratio [protéine]/[ARN] de 3. Bien que notre analyse n'ait été que partielle comme le montre la **Figure 77**, nous avons aussi observé la protection d'autres segments simple brin riches en résidus U. Ces observations sont donc cohérentes avec l'idée que PTB se lie aux grandes structures boucles des structures tige-boucle de la région des A-repeats et qu'elle permette la stabilisation de la structure 2D de cette région. Nous avons aussi observé une protection des résidus A717 et C718 localisés dans la petite boucle interne des structures hélicoïdales formées par deux répétitions. Bien qu'encore incomplètes, ces données suggèrent que lorsqu'elle est seule, la protéine PTB recouvre une très grande partie de la région des A-repeats.

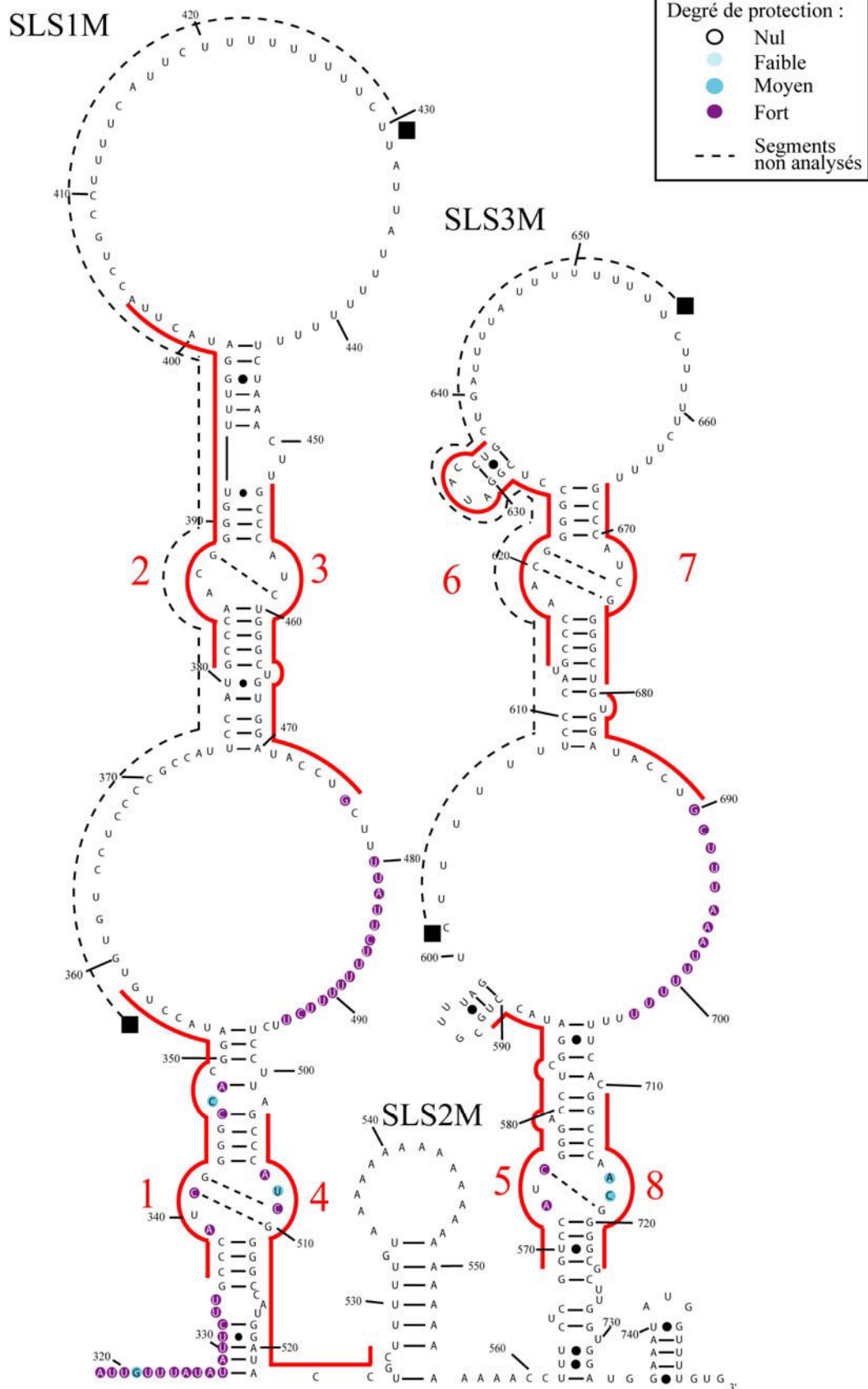




**Figure 76 : Exemple d'analyse des protections contre l'action des RNases générée par fixation de la protéine PTB sur la région des positions 670 à 740 de l'ARN-AM**

a. Exemple de résultat de fractionnement des produits d'extension d'amorce en gel de polyacrylamide dénaturant obtenu avec l'oligonucléotide amorce 3757. Le transcrit a été soumis à une hydrolyse ménagée par les RNases T1 et T2 (0,5 U/1,2 pmoles d'ARN) après incubation en présence ou en absence de protéine PTB (0 ; 0,06 ; 0,18 et 0,3  $\mu\text{M}$ ). Les conditions sont décrites dans la partie Matériel et Méthodes. Les positions des coupures ont été identifiées par extension de oligonucléotide amorce 3757 marqué en 5'. Les arrêts d'extension ont été positionnés dans la séquence grâce à des séquençages du transcrit réalisés avec les mêmes oligonucléotides (pistes U, G, C, A). Les ADNc ont été fractionnés par électrophorèse en gel de polyacrylamide dénaturant à 7%. Les positions des nucléotides sont indiquées à gauche de l'autoradiographie. Le contrôle 1 (piste Ctrl 1) a été réalisé en absence de RNase et en absence de PTB. Le contrôle 2 (Ctrl 2) a été réalisé en absence de RNase et en présence de 0,3  $\mu\text{M}$  de protéine PTB. Les zones protégées sont indiquées à droite des autoradiographies par des cercles dont l'intensité est proportionnelle au degré de protection. Les répétitions sont indiquées par des rectangles à droite des autoradiographies.

b. Représentation schématique des protections contre l'action des RNases T1 et T2 générées par la présence de la protéine PTB des positions 670 à 740 de l'ARN-AM



**Figure 77 : Représentation des protections contre l'hydrolyse par les RNases T1 et T2 générées par fixation de la protéine PTB sur la région des A-repeats murine**  
 Les zones protégées ont été indiquées sur le modèle 3 de structure 2D proposé pour l'ARN-AM par des cercles avec le code couleur indiqué dans l'encart en haut à droite. Le numéro des répétitions est indiqué en rouge.

## 3.2 KSRP

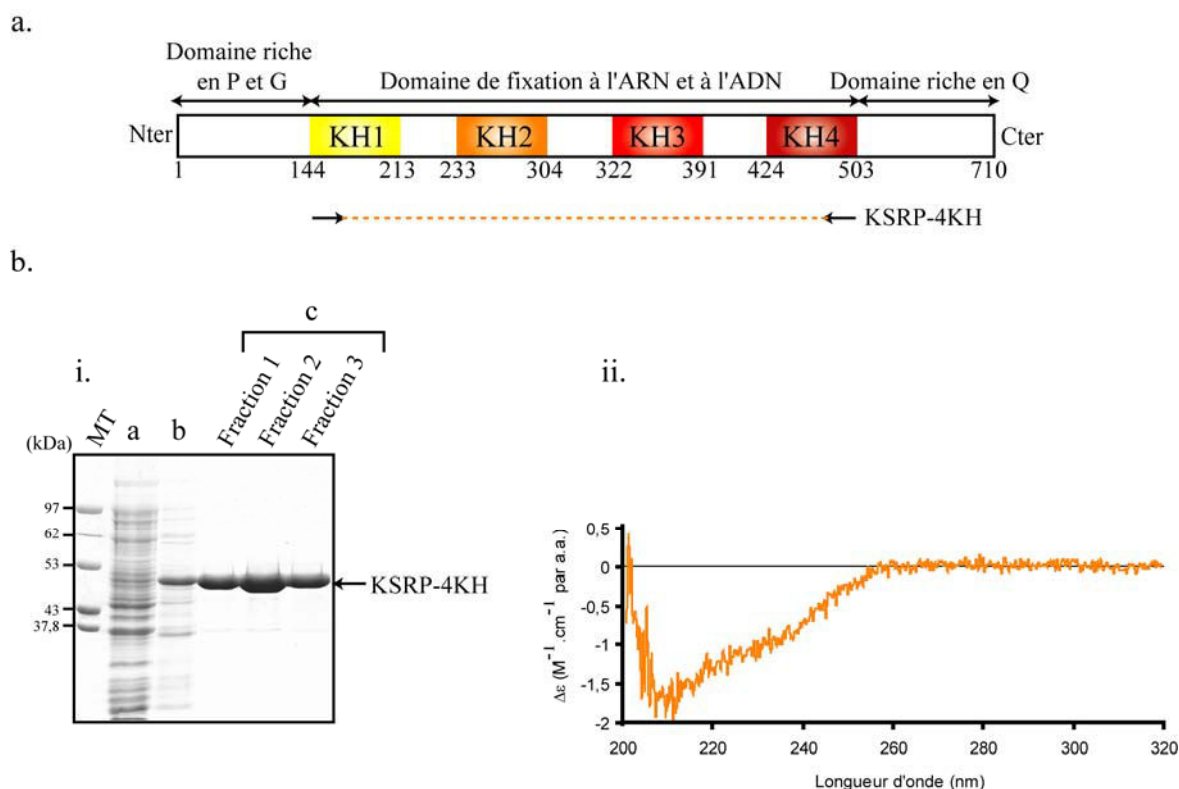
Les anticorps anti-KSRP dont nous disposions ne permettaient pas de réaliser des expériences d'immuno-sélection des RNP formées en extraits nucléaires avec la région des A-repeats contenant la protéine KSRP, probablement du fait que l'épitope ciblé par l'anticorps dont nous disposions n'était pas accessible au sein des complexes. Nous avons donc décidé d'étudier directement les interactions possibles entre KSRP et la région des A-repeats en utilisant une protéine recombinante.

### 3.2.1 Clonage, production et purification de la protéine KSRP-4KH recombinante

A Douablin du laboratoire de F Baklouti (CGMC, Lyon), qui avait effectué un stage au laboratoire, avait tenté de cloner la protéine KSRP entière. Produite sous forme entière chez *E.coli*, KSRP était retrouvée dans les corps d'inclusion. A Douablin avait donc cloné la région codant les 4 domaines KH dans le plasmide pGEX-4T1 permettant la production de protéine sous forme fusionnée à la GST (pGEX-4T1-KSRP-4KH). J'ai utilisé son clone pour réaliser la construction pET28b-KSRP-4KH, ceci de façon à produire la même région contenant les 4 domaines KH sous forme fusionnée à une séquence 6His. L'intérêt était que cette étiquette 6His est beaucoup plus petite que la GST. Cette dernière aurait pu gêner les études d'interaction. De plus, l'étiquette 6His permettait une purification efficace sur billes Ni sépharose (résine Ni-NTA).

Pour réaliser la construction, le fragment *EcoRI/XhoI* de plasmide pGEX-4T1-KSRP-4KH codant les 4 domaines KH a été cloné dans le plasmide pET28b.

La production a été réalisée dans la souche d'*E.coli* BL21 codon<sup>+</sup>, en utilisant une induction par de l'IPTG. Comme le montre la **Figure 78**, un haut degré de purification a pu être obtenu après purification sur résine Ni-NTA et l'absence d'acide nucléique a été vérifiée par mesure du rapport d'absorbance A260/A280 et par dichroïsme circulaire.



**Figure 78 : Production, purification et spectre dichroïque de la protéine KSRP-4KH**

a. Schéma représentant les différents domaines de la protéine KSRP. Les 4 domaines KH permettant la fixation aux ARN sont indiqués en couleur. Leurs positions dans la chaîne polypeptidique sont indiquées. Les oligonucléotides qui ont permis d'amplifier la région codant les 4 domaines KH sont représentés par des flèches.

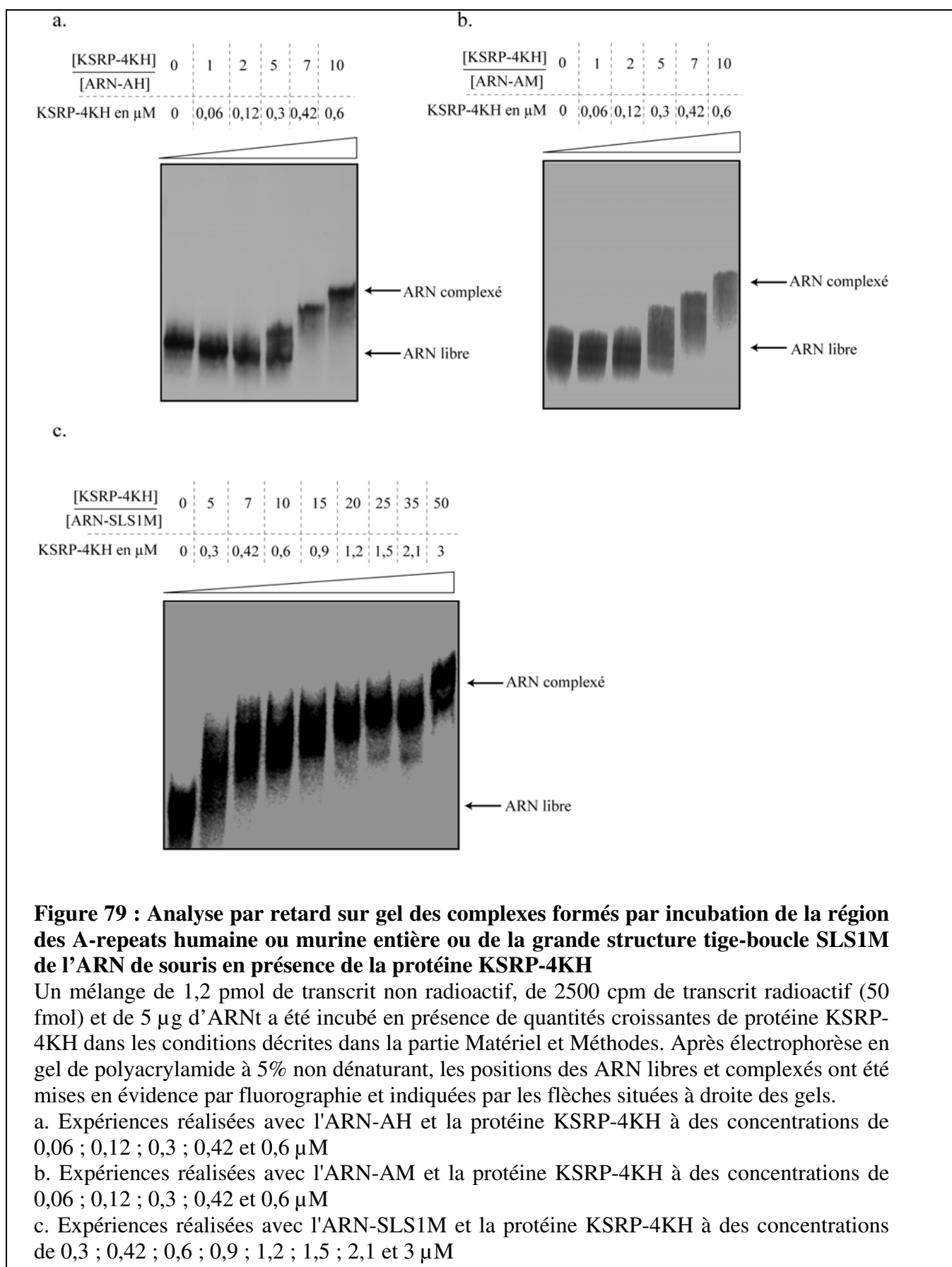
b. i. La protéine KSRP-4KH a été produite en milieu LB induit par l'IPTG. Des échantillons de 100  $\mu$ l de culture ont été prélevés avant (piste a) et après induction (piste b) et ont été traités comme indiqué dans le chapitre Matériel et Méthodes. La purification a été réalisée par chromatographie d'affinité sur colonne Ni-NTA. Un aliquot des fractions dans lesquelles se situait la protéine d'intérêt a été fractionné par SDS-PAGE, en parallèle avec les 2 échantillons précédents (pistes c). Les fractions ont ensuite été réunies, dialysées contre du tampon D et concentrées. Un marqueur de poids moléculaire a été déposé à gauche du gel (MT).

ii. Spectre de dichroïsme circulaire de la protéine KSRP-4KH. Il a été réalisé dans une cuve de 0,875 cm de trajet optique, dans une solution de tampon Tris 10 mM, NaCl 150 mM, pH 7,5 filtrée à 0,22  $\mu$ m et à une concentration en protéine KSRP-4KH de 0,2  $\mu$ M. Les mesures de  $\Delta$  d'absorbance ont été transformées en  $\Delta\epsilon$  par acides aminés, comme indiqué dans la section Matériel et Méthodes.

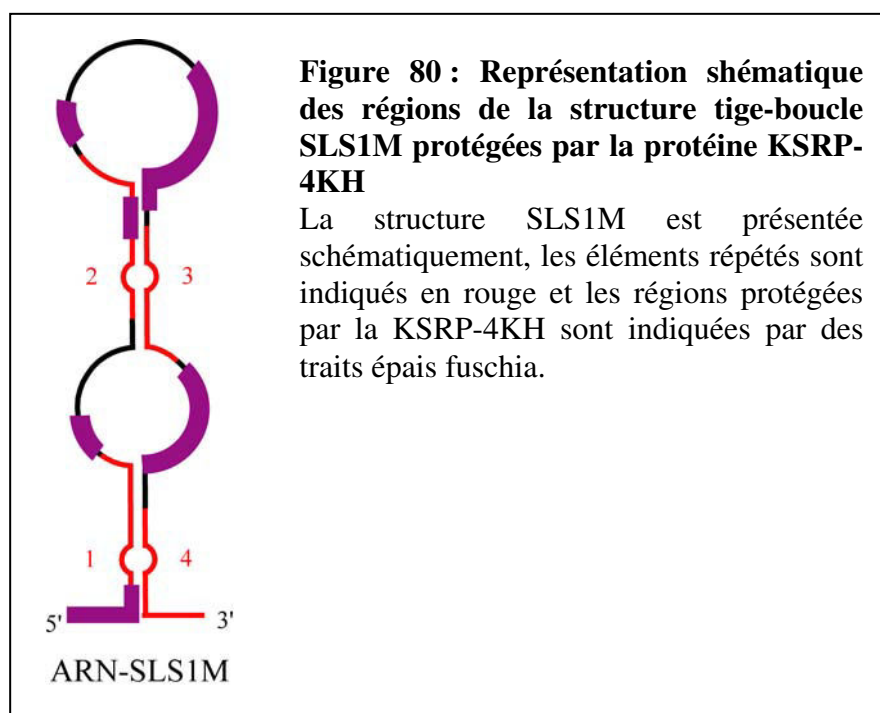
### 3.2.2 Expériences de retard sur gel

Pour vérifier le caractère direct de l'interaction entre la région des A-repeats et la protéine KSRP-4KH, nous avons ensuite réalisé des expériences de retard sur gel en utilisant les ARN-AH, ARN-AM et ARN-SLS1M et la protéine KSRP-4KH. La **Figure 79** présente les résultats obtenus. Ils montrent que la protéine KSRP-4KH interagit avec les 3 ARN testés. Un retard était observé dès l'ajout de 0,06  $\mu$ M de KSRP-4KH (correspondant à un rapport  $[\text{ARN}]/[\text{protéine}]$  de 1/1) et l'ARN libre disparaissait complètement à une concentration en protéine KSRP-4KH de 0,42  $\mu$ M (rapport  $[\text{ARN}]/[\text{protéine}]$  de 1/7). Pour des concentrations

en protéine supérieures à 0,3  $\mu\text{M}$ , le complexe formé était plus fortement retardé, ce qui indiquait que plusieurs protéines se fixaient sur une même molécule d'ARN. Ces observations permettaient d'affirmer que la protéine KSRP-4KH interagit directement avec la région des A-repeats humaine, murine et aussi avec l'ARN-SLS1M.



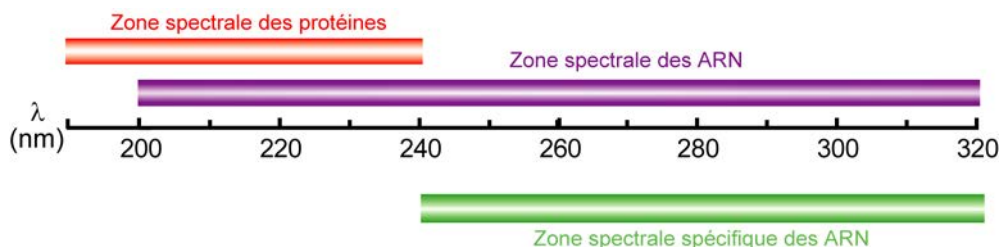
J'ai ensuite encadré Anne Savoye qui a réalisé un stage de Master2 recherche sur ma thématique de recherche. Elle a réalisé au cours de ce stage des expériences d'empreinte aux RNases sur les complexes formés entre l'ARN-AM et la protéine KSRP-4KH. Les résultats qu'elle a obtenus montrent que les sites de fixation préférentiels de la protéine KSRP-4KH se situent dans les régions en simple brin riches en U des structures tige-boucle SLS1M et SLS3M. La **Figure 80** représente de façon schématique les régions de la structure tige-boucle SLS1M protégées par la protéine KSRP-4KH. Malgré le fait que les expériences d'empreinte aux RNases avec PTB n'étaient que préliminaires, l'ensemble des données recueillies suggère fortement que les protéines PTB et KSRP-4KH se lient sur les mêmes segments de la région des A-repeats. Ces données étaient en accord avec l'hypothèse que PTB et KSRP pourraient stabiliser la structure 2D de la région des A-repeats. Nous avons cherché à apporter des informations sur cette hypothèse en utilisant le dichroïsme circulaire.



### 3.3 Etude des complexes SLS1M/PTB et SLS1M/KSRP-4KH par dichroïsme circulaire

#### 3.3.1 Principe de la méthode

Comme déjà mentionné, les acides nucléiques donnent lieu à un signal positif de dichroïsme circulaire et sont les seuls à contribuer au spectre dans la région comprise entre des longueurs d'onde de 240 à 320 nm (**Fig. 81**). Nous avons donc utilisé cette méthodologie pour tester l'impact de l'association des protéines PTB et KSRP-4KH à l'ARN-SLS1M, sur la structuration de la grande structure tige-boucle SLS1M.



**Figure 81 : Représentation des zones spectrales des protéines et des ARN**

La zone spectrale des protéines est représentée en rouge, celle des ARN en violet. Les acides nucléiques ont un spectre dichroïque caractéristique dans la région de 200 à 320 nm avec un signal positif de 240 à 320 nm. Les protéines donnent lieu à un signal négatif de 190 à 240 nm. Dès lors, de 240 à 320 nm (indiqué en vert), seuls les acides nucléiques présentent un signal spécifique. Cette propriété donne la possibilité d'analyser le signal dichroïque spécifique des ARN et donc d'apprécier ses changements structuraux en présence de protéine.

Si l'interaction de PTB et de KSRP stabilisait les régions hélicoïdales de la structure SLS1M, on pouvait s'attendre à détecter une stabilisation du signal de dichroïsme circulaire à 268 nm, ce qui reflète le degré d'hélicité lorsque la température est augmentée. Il fallait déjà dans un premier temps tester quel était l'effet de l'ajout de PTB et de KSRP sur la structure tige-boucle SLS1M à 20°C. Pour cela nous avons tiré profit du fait que des cuves du dichrographe ayant une séparation interne peuvent être utilisées.

Dans les expériences de dichroïsme circulaire, nous ne pouvions pas utiliser d'Hepes qui est contenu dans le tampon D que nous avons utilisé pour les expériences de retard sur gel, car il a la propriété d'absorber dans la région spectrale comprise entre les longueurs d'onde de 200 à 320 nm. Nous avons donc utilisé le tampon suivant : Tris 10 mM, NaCl 150 mM, pH 7,5 filtré à 0,22 µm.

Un autre point dont il fallait tenir compte, est que l'obtention de spectre de dichroïsme circulaire nécessite l'utilisation de grandes quantités de matériel. L'ARN et la protéine devaient avoir une concentration élevée pour que le signal soit exploitable. Ainsi, nous avons constaté qu'il fallait au moins une concentration de l'ordre 0,2 µM d'ARN-SLS1M dans une cuve de 500 µl pour pouvoir enregistrer un spectre correct de l'ARN, soit 7,5 µg d'ARN-SLS1M.

Nous avons enregistré les spectres de dichroïsme circulaire dans la région comprise entre les longueurs d'onde 200 à 320 nm.

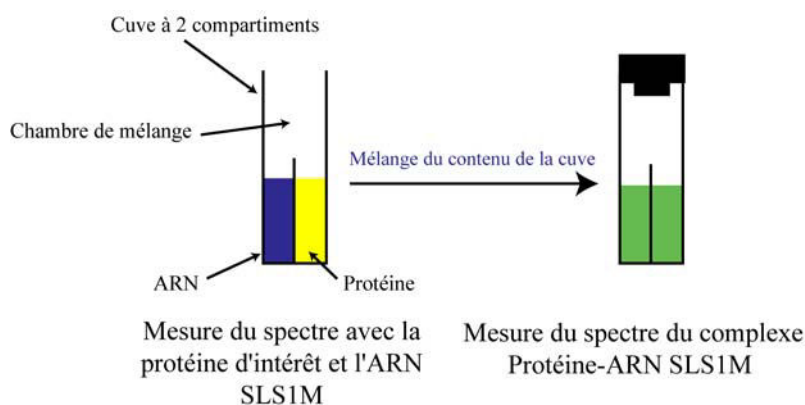
Pour étudier l'effet des protéines sur la structure de l'ARN, l'expérience se déroulait en 2 étapes :

Premièrement, la somme des spectres de l'ARN et de la protéine était enregistrée « avant interaction ». Pour cela, l'ARN-SLS1M et la protéine testée étaient chacun placés dans un des deux compartiments de la cuve en quartz (**Fig. 82**). Comme nous l'avons dit précédemment, l'ARN et la protéine étaient chacun en solution dans du tampon Tris 10 mM,



NaCl 150 mM, pH 7,5 à une concentration respective de 0,2  $\mu\text{M}$  et de 0,4  $\mu\text{M}$  car nous avons décidé d'utiliser un rapport molaire [ARN]/[protéine] de  $\frac{1}{2}$  car nous savions que tout l'ARN était complexé à cette concentration.

Deuxièmement, le spectre dichroïque du complexe [SLS1M-Protéine] (spectre « après interaction ») était enregistré toujours pour la même fenêtre de longueurs d'onde 200 à 320 nm, après mélange par inversion du contenu de la cuve et une incubation dans les mêmes conditions que celles utilisées pour les expériences de retard sur gel (Fig. 82). La comparaison des deux spectres avant et après le mélange permettait d'évaluer l'effet de la fixation des protéines testées sur la structuration globale de l'ARN-SLS1M (Fig. 82).



**Figure 82 : Schéma représentant les différentes étapes de la mesure des spectres de dichroïsme circulaire visant à tester l'effet de l'interaction entre l'ARN-SLS1M et la protéine testée sur la structure 2D de l'ARN**

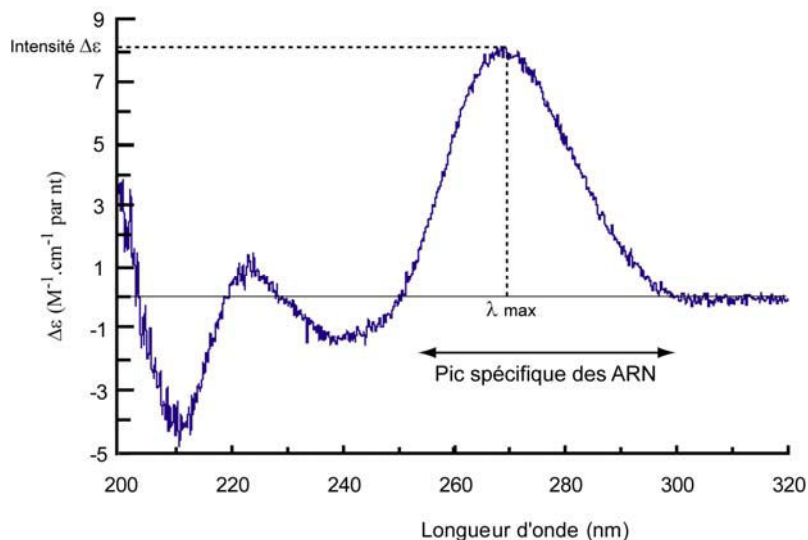
Pour chaque essai, 10 spectres étaient enregistrés. Les résultats sont présentés en  $\Delta\epsilon$  correspondant à la différence d'épsilon molaire de la lumière polarisée circulairement à droite et de la lumière polarisée circulairement à gauche par nucléotide et ceci au  $\lambda_{\text{max}}$ , longueur d'onde à laquelle l'intensité du pic spécifique des ARN est maximale. Les variations du  $\lambda_{\text{max}}$  et de l'intensité du  $\Delta\epsilon$  ( $\delta\Delta\epsilon$ ) sont respectivement considérées dans le cas où la molécule d'ARN a une architecture simple comme caractéristique d'une modification de la courbure et de l'empilement des bases de l'ARN.

### 3.3.2 Résultats

- ARN-SLS1M seul

Le spectre de l'ARN seul était caractéristique des acides nucléiques. Il présentait un pic positif à 267,6 nm ( $\lambda_{\text{max}}$ ) et son intensité  $\Delta\epsilon$  était de 8,15  $\text{M}^{-1}\cdot\text{cm}^{-1}$  par nt (Fig. 83). Ce pic montrait que l'ARN-SLS1M produit et purifié *in vitro* formait bien des hélices.





**Figure 83 : Spectre de dichroïsme de l'ARN-SLS1M**

Le spectre a été réalisé dans une cuve de 0,875 cm de trajet optique, dans une solution de tampon Tris 10 mM, NaCl 150 mM, pH 7,5 filtrée à 0,22  $\mu\text{m}$  et à une concentration de l'ARN-SLS1M de 0,2  $\mu\text{M}$ . Les mesures de  $\Delta$  d'absorbance ont été transformées en  $\Delta\epsilon$  par nucléotide comme indiqué dans la section Matériel et Méthodes.

- Complexe [ARN-SLS1M/protéine testée]

#### **PTB :**

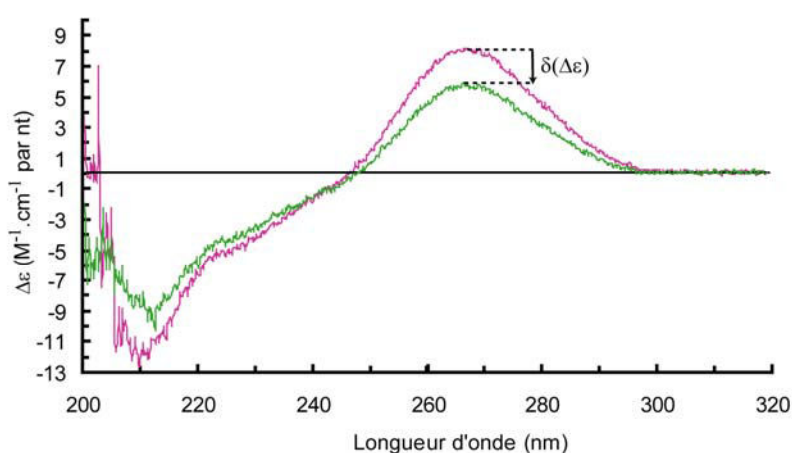
L'expérience a été répétée 2 fois avec des lots d'ARN et de protéines différents. Nous constatons que l'intensité du pic positif proche de 265 nm était de  $5,50 \text{ M}^{-1}.\text{cm}^{-1}$  par nt, ce qui représentait une diminution de  $\Delta\epsilon$  ( $\delta\Delta\epsilon$ ) de l'ordre de  $2,65 \text{ M}^{-1}.\text{cm}^{-1}$  par nt, ceci avec une précision de  $0,026 \text{ M}^{-1}.\text{cm}^{-1}$  par nt. Cette diminution du pic au  $\lambda_{\text{max}}$  résultait vraisemblablement d'une diminution de l'empilement des bases lors de l'interaction avec la protéine (**Fig. 84a**). Par contre, le  $\lambda_{\text{max}}$  ne variait pas de manière significative. Il semblait donc que la fixation de la protéine PTB entraînait l'ouverture de certains empilements de bases de l'ARN-SLS1M, ce qui n'allait pas dans le sens de notre hypothèse d'une stabilisation de la structure 2D.

#### **KSRP :**

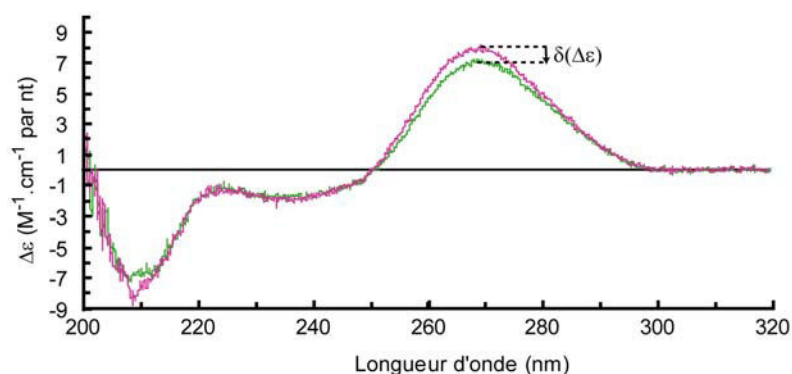
Là encore l'expérience a été répétée 2 fois avec des lots d'ARN et de protéines différents. L'intensité du pic positif proche de 265 nm ne diminuait, lors de la formation du complexe que de 8,15 à  $7,15 \text{ M}^{-1}.\text{cm}^{-1}$  par nt ; c'est-à-dire de façon moins prononcée que dans le cas de PTB. On pouvait donc supposer que la rupture d'empilement des bases entraînée par KSRP-4KH était inférieure à celle obtenue pour PTB (**Fig. 84b**). Le  $\lambda_{\text{max}}$  variait de 267,9 à 269,2 nm. Dans le cas d'un ARN formant une seule hélice, ce déplacement pourrait être interprété comme reflétant le fait que la fixation de la protéine provoquait une modification de la courbure de la double hélice. Néanmoins, la structure SLS1M renferme 4 segments en hélice. L'interprétation de cette variation n'est donc pas si simple.

Un point important était que la formation du complexe [SLS1M-KSRP-4KH] induisait un changement d'empilement des bases de l'ARN-SLS1M plus faible que celui induit par la protéine PTB ( $\delta\Delta\epsilon$  de  $2,65 \text{ M}^{-1} \cdot \text{cm}^{-1}$  par nt pour PTB contre 1 pour KSRP-4KH). Ceci pouvait aussi être lié au fait qu'alors que PTB se lie à un ARN ne contenant que 2 répétitions et leurs espaceurs, KSRP nécessite la présence de 4 éléments répétés et leurs espaceurs pour se lier. Ceci nous avait amené à conclure que KSRP nécessite la formation d'une grande structure tige-boucle de type SLS1 pour se fixer. Il n'est donc pas illogique que la structure SLS1 soit plus préservée lors de la fixation de KSRP que de PTB. On peut imaginer que PTB induise la transition de la structure SLS1 vers une autre des structures possibles de cette région.

a.



b.



### Figure 84 : Etude des complexes [SLS1M/PTB] et [SLS1M/KSRP-4KH] par dichroïsme circulaire

a. Spectres de dichroïsme circulaire de l'ARN-SLS1M ( $0,2 \mu\text{M}$ ) et la protéine PTB ( $0,4 \mu\text{M}$ ) séparés (rose), de l'ARN-SLS1M et la protéine PTB mélangés en stoechiométrie 1:2 (vert). Les spectres ont été réalisés dans une cuve de  $0,875 \text{ cm}$  de trajet optique, dans une solution de tampon Tris  $10 \text{ mM}$ , NaCl  $150 \text{ mM}$ , pH  $7,5$  filtrée à  $0,22 \mu\text{m}$ . Les mesures de  $\Delta$  d'absorbance ont été transformées en  $\Delta\epsilon$  par nucléotides comme indiqué dans la section Matériel et Méthodes.

b. Spectres de dichroïsme circulaire de l'ARN-SLS1M ( $0,2 \mu\text{M}$ ) et la protéine KSRP ( $0,4 \mu\text{M}$ ) séparés (rose), de l'ARN-SLS1M et la protéine KSRP mélangés en stoechiométrie 1:2 (vert).

Les spectres ont été réalisés dans une cuve de 0,875 cm de trajet optique, dans une solution de tampon Tris 10 mM, NaCl 150 mM, pH 7,5 filtrée à 0,22  $\mu\text{m}$ . Les mesures de  $\Delta$  d'absorbance ont été transformées en  $\Delta\epsilon$  par nucléotides comme indiqué dans la section Matériel et Méthodes.

- Stabilité des complexes [ARN-SLS1M/PTB ou KSRP-4KH] lors d'une augmentation de la température

Comme nous l'avons mentionné précédemment, si PTB et KSRP stabilisaient les régions hélicoïdales de la structure SLS1M, on pouvait s'attendre à détecter une stabilisation des empilements de bases à température élevée et donc un maintien du signal au  $\lambda_{\text{max}}$  en présence de ces protéines. Les tests de l'effet de la température n'ont été réalisés qu'une seule fois et devront donc être répétés.

Les spectres dichroïques de l'ARN-SLS1M seul et complexé avec les protéines testées ont été enregistrés à différentes températures (20°C, 37°C, 45°C, 55°C et 65°C), le dichrographe étant muni d'un système permettant de moduler la température. Les expériences ont été réalisées selon le protocole utilisé précédemment.

### ARN-SLS1M

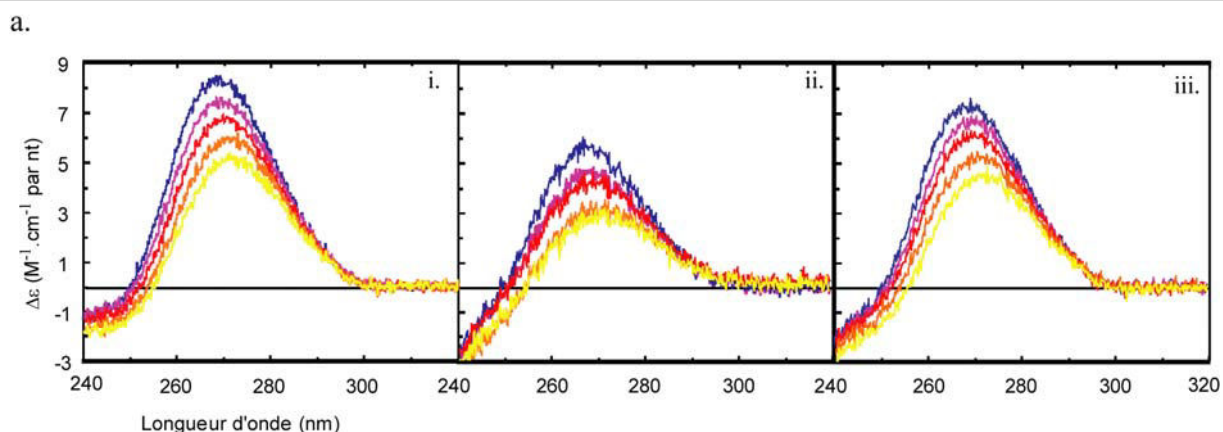
Dans le cas de l'ARN-SLS1M seul, l'intensité du pic proche de 265 nm diminuait au cours de l'augmentation de la température. Les valeurs de  $\Delta\epsilon$  s'échelonnaient de 8,15  $\text{M}^{-1} \cdot \text{cm}^{-1}$  par nt (20°C) à 5,21  $\text{M}^{-1} \cdot \text{cm}^{-1}$  par nt (65°C) (**Fig. 85**). Nous avons remarqué également que le  $\lambda_{\text{max}}$  était décalé vers les plus hautes longueurs d'ondes d'une moyenne de 0,68 nm par saut de température (**Fig. 85**). Ces diminutions de  $\Delta\epsilon$  et décalages de  $\lambda_{\text{max}}$  étaient significatives de la déstructuration globale de l'ARN-SLS1M causée par l'augmentation de la température. Néanmoins, la nécessité de chauffer à 65°C pour avoir une déstructuration partielle de la structure confirmait la stabilité de la structure SLS1M en accord avec les  $\Delta G$  calculés.

### [SLS1M-PTB]

Pour le complexe [SLS1M-PTB], les valeurs de  $\Delta\epsilon$  s'échelonnaient de 5,50 à 2,87  $\text{M}^{-1} \cdot \text{cm}^{-1}$  par nt et les  $\lambda_{\text{max}}$  étaient à nouveau décalés vers les plus hautes longueurs d'ondes d'une moyenne de 0,84 nm. En plus, les diminutions de  $\Delta\epsilon$  semblaient se faire par palier, d'abord à 37°C puis à 55°C.

### [SLS1M-KSRP-4KH]

Pour le complexe [SLS1M-KSRP-4KH] les valeurs de  $\Delta\epsilon$  s'échelonnait de 7,15  $\text{M}^{-1} \cdot \text{cm}^{-1} \cdot \text{nt}$  à 4,42  $\text{M}^{-1} \cdot \text{cm}^{-1}$  par nt c'est-à-dire une diminution  $\Delta\epsilon$  un peu plus faible que pour l'ARN seul (**Fig. 85**).



b.

	$\lambda_{\max}$ (nm)			$\Delta\epsilon$ (M <sup>-1</sup> .cm <sup>-1</sup> par nt)		
	SLS1M	PTB-SLS1M	KSRP-SLS1M	SLS1M	PTB-SLS1M	KSRP-SLS1M
20°C	267,6	267,6	267,6	8,15	5,50	7,15
37°C	269,2	268,4	268,4	7,50	4,73	6,54
45°C	269,4	269,2	269,2	6,73	4,10	6,01
55°C	270	271,2	271,2	5,88	3,16	5,10
65°C	271	271,8	271,8	5,21	2,87	4,42

**Figure 85 : Influence de la présence des protéines PTB et KSRP-4KH sur la stabilité thermique de la structure 2D de l'ARN-SLS1M**

a. i. Spectres de dichroïsme circulaire de l'ARN-SLS1M à 20°C (bleu), 37°C (violet), 45°C (rouge), 55°C (orange), 65°C (jaune).

ii. Spectres de dichroïsme circulaire de l'ARN-SLS1M complexé avec PTB à 20°C (bleu), 37°C (violet), 45°C (rouge), 55°C (orange), 65°C (jaune).

iii. Spectres de dichroïsme circulaire de l'ARN-SLS1M complexé avec KSRP-4KH à 20°C (bleu), 37°C (violet), 45°C (rouge), 55°C (orange), 65°C (jaune).

b. Présentation des valeurs de  $\lambda_{\max}$  et des  $\Delta\epsilon$  pour l'ARN-SLS1M seul et complexé aux différentes températures utilisées.

En conclusion, ces expériences, dans leur état, nous ont déjà apporté beaucoup d'informations. Tout d'abord, nous confirmons que la structure SLS1M est thermodynamiquement stable. Deuxièmement, nous observons un comportement très différent des protéines PTB et KSRP vis-à-vis de la structure SLS1. PTB, à la concentration à laquelle nous l'avons utilisée, semble modifier de manière importante la structure 2D de l'ARN. Bien que les expériences de l'effet de la température sur les complexes n'aient été faites qu'une fois, le phénomène de fusion de la structure 2D semble très différent pour PTB et pour KSRP-4KH. Pour KSRP-4KH, on assiste comme pour l'ARN libre à une fusion progressive de la structure de l'ARN. Alors que pour PTB, des palliers de fusion sont observés. Ceci est en accord avec le fait que KSRP doit se lier à la structure SLS1 impliquant les 4 motifs répétés, motifs qui doivent donc se dissocier progressivement de manière similaire en présence ou en absence de KSRP-4KH. Par contre, il est possible qu'à la concentration utilisée, PTB induit un changement de conformation de la région SLS1M. Ce changement de conformation pourrait avoir pour conséquence la génération de deux zones

distinctes de structures qui sont détruites à des températures différentes, ce qui expliquerait les palliers observés.

Ces observations sont très intéressantes car elles soulèvent la question du rôle possible que PTB et KSRP pourraient avoir *in cellulo* vis-à-vis de la structure de la région des A-repeats. Ces données de dichroïsme circulaire ne confortent pas l'hypothèse selon laquelle ces 2 protéines pourraient structurer la structure tige-boucle à 4 éléments répétés. Par contre, comme KSRP ne peut se lier qu'à cette structure, elle pourrait favoriser son maintien, alors que PTB pourrait au contraire favoriser un autre type de structuration.

Ces données ouvrent de nouvelles pistes de recherche que nous présenterons dans la discussion.

Pour compléter ce travail, il faudra répéter les expériences visant à tester l'effet de la température sur les complexes formés avec PTB et KSRP-4KH. Au cours de ces expériences, il sera aussi important de tester l'effet de la concentration en protéine sur les variations de  $\Delta\epsilon$  observées, de définir par retard sur gel la température à laquelle le complexe ARN/protéine est dissocié. Nous pouvons aussi envisager de réaliser des expériences d'empreinte à la RNase T1 à différentes températures, sachant que cette enzyme résiste à des températures élevées.

Enfin, il sera intéressant d'enregistrer des spectres de dichroïsme circulaire pour un mélange ARN-SLS1M et protéine PTB et KSRP, ceci en faisant varier la concentration relative de ces 2 protéines.

Il sera également intéressant de compléter l'ensemble de ces études réalisées avec la protéine PTB entière et la protéine KSRP-4KH par des études similaires (retard sur gel, empreintes aux RNases, dichroïsme circulaire) réalisées avec des sous-fragments de PTB contenant 1, 2 ou 3 RRM et des sous-fragments de KSRP-4KH.

## E. Constructions d'outils génétiques en vue d'étudier ultérieurement la relation structure/fonction de la région des A-repeats

### 1. Objectifs et stratégie envisagée

Ayant défini la structure 2D adoptée par la région des A-repeats *in vitro*, il nous paraissait nécessaire de développer des outils génétiques nous permettant de tester l'importance fonctionnelle des motifs structuraux mis en évidence.

Par ailleurs, d'une manière générale, il nous paraissait nécessaire de développer des outils génétiques permettant de tester le rôle de la région des A-repeats en tant que région ARN. En effet, comme nous l'avons vu dans l'introduction, toutes les études visant à comprendre la fonction de la région des A-repeats ont été réalisées par délétion d'ADN, ce qui touchait à la fois l'ADN et l'ARN. Or, cette région d'ADN, en plus de coder l'ARN *Xist* et *Tsix*, peut elle-même jouer un rôle dans l'inactivation par ses propriétés d'activation de la transcription du gène *Xist* (Allaman-Pillet *et al.*, 2000). Ainsi, chez l'embryon murin, la délétion mono-allélique paternelle de la région des A-repeats provoque un défaut d'inactivation due à une chute d'expression de *Xist* (Hoki *et al.*, 2008). L'importance fonctionnelle de la région des A-repeats semble donc dépendre relever à la fois de la composante ADN et de la composante ARN. Nous avons pensé, afin de tester l'importance fonctionnelle de la région des A-repeats au niveau de l'ARN et ceci dans son contexte naturel, que la meilleure solution serait d'introduire dans le gène *Xist* des sites donneur et accepteur d'épissage permettant une élimination de la région des A-repeats par épissage après transcription par l'ARN polymérase II. On sait que le transcrit *Xist* subit plusieurs événements d'épissage ; l'idée était donc d'en rajouter un supplémentaire éliminant la région des A-repeats.

Nos objectifs ont donc été : d'une part de réaliser des constructions en vue, à terme, d'échanger au sein de cellules ES de souris femelles le segment ADN correspondant à la région des A-repeats par un segment muté. D'autre part, nous avons essayé de générer un site 5' et un site 3' d'épissage dans un grand fragment d'ADN contenant la région des A-repeats afin de tester si nous avions bien un épissage efficace de cette région, ceci tout en introduisant le nombre le plus limité possible de mutations dans la région ADN.

La suite des expériences devait donc consister à générer les mutations créant des sites d'épissage et les mutations modifiant la structure 2D de la région des A-repeats dans un très grand fragment d'ADN incluant la partie 5' du gène *Xist*, et d'insérer ce fragment à la place de la séquence normale et ceci uniquement dans un des deux chromosomes X de cellules ES de souris femelles. Nous devions ainsi pouvoir tester l'effet des mutations générées dans leur contexte. Ce qui n'avait jamais été réalisé par d'autres équipes lorsque nous avons débuté ce travail. Depuis, une équipe (Hoki *et al.*, 2008) a réalisé une délétion de la région des A-repeats comme mentionné ci-dessus, mais ne l'a pas remplacée par une séquence mutée.

Nous présenterons d'abord la préparation de la lignée cellulaire portant une délétion dans le chromosome X, puis la préparation des mutants.

## 2. Génération de la lignée cellulaire

La génération d'une lignée transgénique présentant les propriétés envisagées était une tâche ardue. Nous avons pour objectif d'utiliser la méthode d'échange de cassette médiée par une recombinase (RMCE) (pour revue, Bode *et al.*, 2000).

### 2.1 Principe et stratégie de la méthode

La méthode comportait plusieurs étapes.

1) La première étape consistait en la génération d'un plasmide nommé plasmide de sélection appelée LoxP-FRT-PTK. Ce plasmide de sélection contient la cassette de sélection appelée P-TK (puromycine/thymidine-kinase) servant à sélectionner les clones « positifs » au cours des différentes étapes de la stratégie.

Cette cassette de sélection est flanquée de sites inversés LoxP-FRT. Les séquences LoxP sont de courtes séquences de 34 pb reconnus par la recombinase Cre. Cette dernière catalyse la recombinaison entre les deux sites de reconnaissances. Les sites FRT (FLP recognition target) sont longs de 34 pb et correspondent aux séquences-cibles reconnues par la recombinase FLP (pour Flip-Flop). L'expression de la FLP dans la lignée portant la cassette de sélection devait induire la suppression de la cassette.

Les sites de Lox-FRT sont eux-mêmes bordés par des séquences ADN bordant la région à substituer dans le chromosome où la substitution doit être faite. Ces séquences sont appelées bras d'homologie droit (BrasD) et bras d'homologie gauche (BrasG).

2) La seconde étape consistait en la transfection de cellules ES de souris par le plasmide de sélection. Ceci devait nous permettre de remplacer par recombinaison homologue le locus d'intérêt par la cassette de sélection.

#### Choix de la lignée de cellules ES

Nous avons choisi d'utiliser la lignée de cellules ES XX de type HP3-10. Elle devait nous permettre de cibler un seul allèle (et non les deux allèles) pour l'événement de recombinaison homologue. Cette lignée est hétérozygote pour le locus XIC (*Mus musculus/Mus domesticus*, forme allélique 129/Pgk du XIC) (Clerc et Avner, 1998). Ceci permettait par le choix des bras d'homologie de ne cibler qu'un seul locus XIC pour la recombinaison homologue. Le fond génétique 129 a été choisi pour les bras d'homologie. Le chromosome correspondant était donc ciblé.

### Choix des bras d'homologie

Les bras G et D choisis correspondaient respectivement aux séquences ADN situées aux positions - 7011 à - 2245 et +2034 à +9529, en prenant comme référence +1, le site principal d'initiation de la transcription du gène *Xist*. La recombinaison homologue devait permettre le remplacement d'une région de 4,2 kb qui s'étend des positions -2244 à +2033 au locus *Xist*, par la cassette de sélection. La région délétée incluait la région des A-repeats mais aussi les promoteurs du gène *Xist*. Cette taille pouvait paraître démesurée par rapport à celle de la région des A-repeats, mais cette lignée devait également permettre l'étude des éléments régulant l'expression de *Xist* (étude menée par C Rougeulle et A Oldfield).

3) L'étape suivante consistait à substituer le locus sauvage par la cassette de sélection (FRT-LoxP-PTK), ce qui nécessitait ensuite la sélection des clones dans lesquels l'événement de recombinaison avait eu lieu. Cette sélection reposait sur la présence d'un gène de résistance à la puromycine présent dans la cassette de sélection P-TK. Les clones de cellules ES ayant subi un événement stable de recombinaison étaient donc viables en présence de puromycine.

4) Enfin, il s'agissait d'utiliser la cassette LoxP-FRT-PTK pour insérer la séquence mutée. Pour cela les cellules ES portant la délétion étaient co-transfectées avec un plasmide contenant les fragments mutés flanqués de sites loxP et avec un second plasmide permettant l'expression de la recombinaise cre. L'expression de cette dernière devait permettre d'échanger la cassette LoxP-FRT-PTK par les fragments mutés de la région des A-repeats portés par le plasmide donneur, ceci grâce aux sites de recombinaison LoxP de la cassette de sélection.

L'expression de la FLP dans la lignée portant la cassette de sélection devait induire la suppression sans autre réinsertion de la cassette de sélection, ce qui devait nous permettre d'avoir une délétion complète de la région de 4,2 kb.

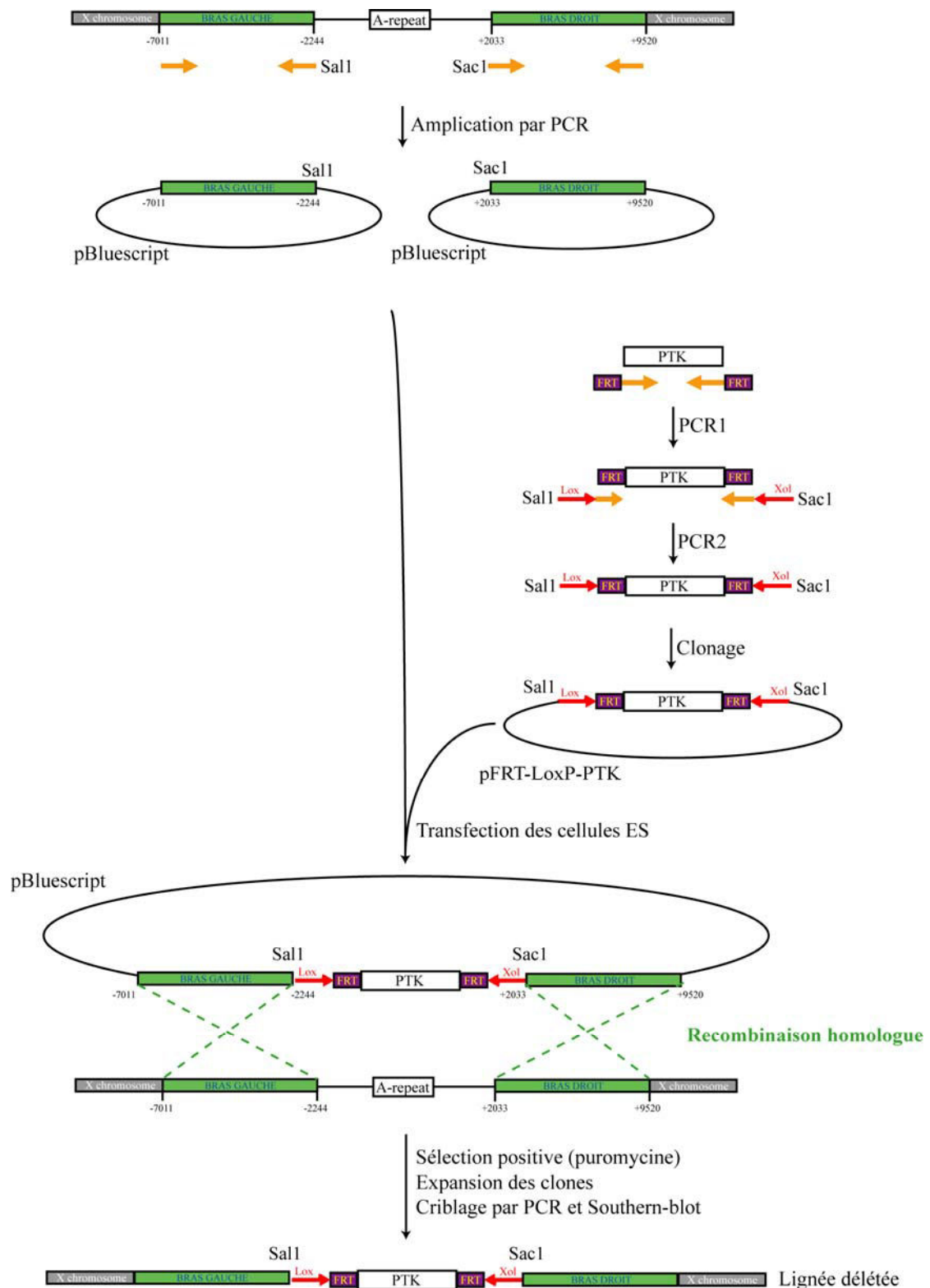
La validation de l'événement de recombinaison médiée par l'une ou l'autre des deux recombinaises faisait intervenir le deuxième gène de sélection présent dans la cassette PTK. Il s'agit du gène de la thymidine kinase du virus HSV (TK). Cette enzyme est capable de phosphoryler le gancyclovir (ajouté au milieu de culture) qui devient alors toxique pour les cellules. Cette sensibilité au gancyclovir devait nous permettre de sélectionner les clones ne portant plus la cassette sans qu'il y ait eu insertion, ainsi que les clones qui avaient intégré les fragments mutés de la région des A-repeats.

## 2.2 Résultats

J'ai pris en charge avec A Visvikis la production de la cassette de sélection flanquée des sites FRT et LoxP. Par 2 étapes d'amplification successives (PCR1 et 2), nous avons



ajouté à un segment d'ADN codant, la cassette P-TK (fournie par l'équipe de P Avner), les sites de recombinaison FRT puis les sites loxP couplés aux sites de restrictions *SalI* et *SacI* (Fig. 86). En parallèle, les bras D et G ont été amplifiés, clonés et séquencés par Andrew Oldfield sous la direction de C Rougeulle et P Clerc (équipe de P Avner). Une ligation à 3 partenaires lui a permis de construire le vecteur de sélection.



**Figure 86 : Stratégie utilisée pour générer la lignée de cellules ES de souris femelles présentant une délétion de 4,2 kb au locus *Xist***

Se référer au texte pour les détails

Etant donné leur expertise dans le domaine, l'équipe de P Avner a pris en charge la préparation finale de la lignée portant la délétion. A Oldfield a donc réalisé les transfections des cellules ES XX de type HP3-10 par le plasmide pFRT-LoxP-PTK. En utilisant la résistance à la puromycine, apparaissaient les clones qui présentaient la recombinaison. Les clones résistants à la puromycine ont été repiqués et amplifiés dans des plaques de 96 puits afin de générer suffisamment de matériel pour poursuivre le criblage. En effet, il fallait confirmer la présence de la cassette dans le génome, mais aussi vérifier que la recombinaison avait eu lieu au bon locus. L'ADN génomique de près de 2000 clones a été préparé. La présence et la position de la cassette de sélection ont été vérifiées par amplification par PCR et Southern-blot. Malheureusement, aucun phénomène de recombinaison homologue n'a été détecté au locus ciblé. L'équipe de P Avner est actuellement en train de tester différentes conditions afin de favoriser la recombinaison homologue au locus d'intérêt.

### 3. Génération de sites d'épissage bordant la région des A-repeats

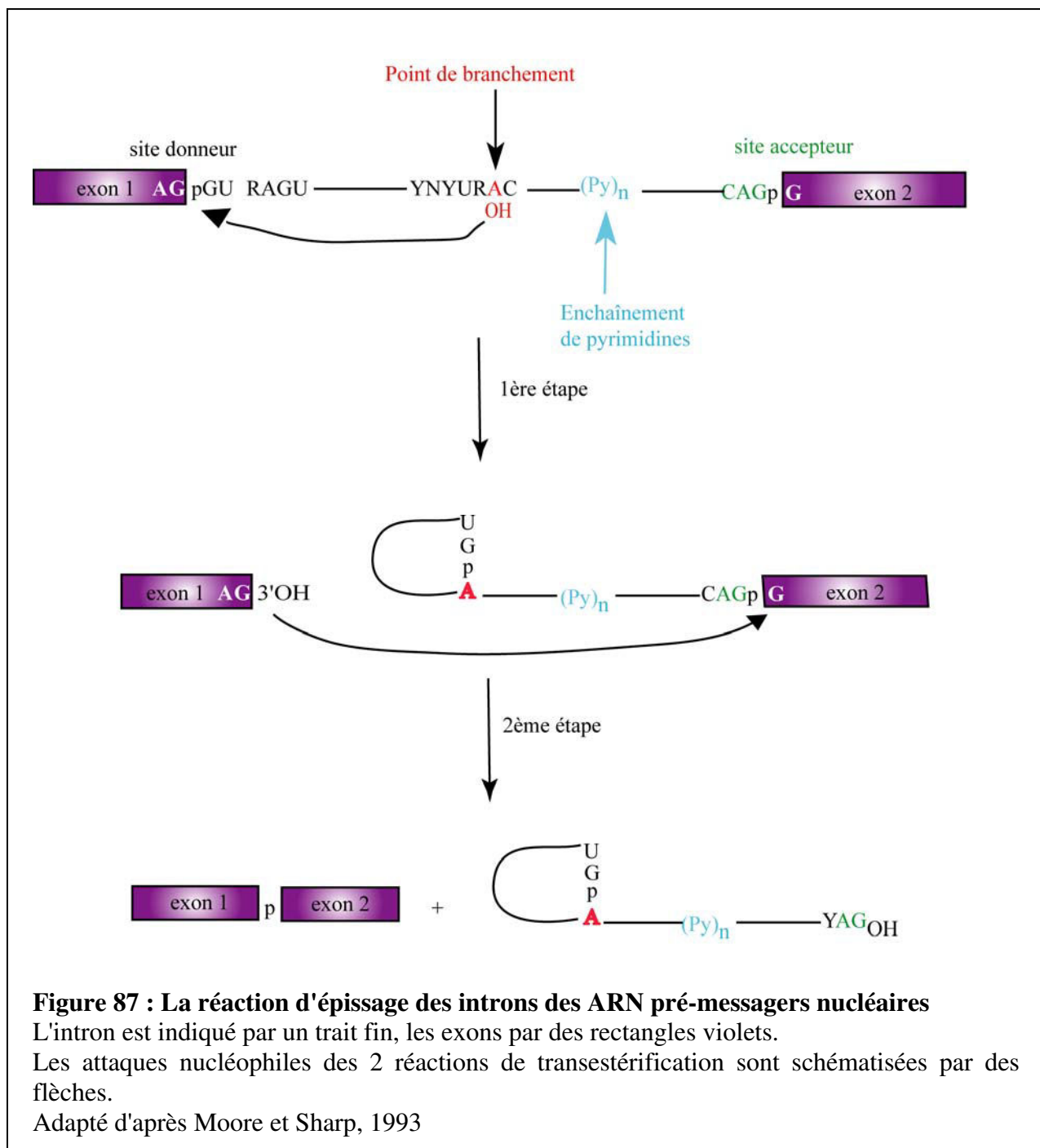
#### 3.1 Présentation de l'épissage

Nous décrivons d'abord ce que sont les sites 5' et 3' d'épissage, et comment ils interviennent dans le processus. L'élimination des introns au sein des ARN pré-messagers est réalisée par un macro-complexe renfermant 5 UsnRNP (U small Nuclear RiboNucleo Particle) (snRNP U1, U2, U4/U6.U5), qui s'associent de manière séquentielle sur l'ARN pré-messager, ceci en reconnaissant les séquences conservées des sites d'épissage. Le macro-complexe formé par l'association de ces UsnRNP à l'ARN pré-messager, avec en plus, un grand nombre des protéines nucléaires, a la capacité de catalyser les deux étapes de la réaction d'épissage.

Des séquences conservées sont présentes à la fois au niveau des sites 5' donneurs et des sites 3' accepteurs (**Fig. 87**) : au niveau du site donneur, la séquence AG/GURAGU (où R=purine) est conservée et est reconnue par la snRNP U1 et un certain nombre de facteurs protéiques permettant la définition de ce site. Au niveau du site accepteur, trois éléments de séquence sont importants. Le premier est la séquence YAG/G à la bordure intron/exon, qui est reconnue de façon spécifique par le facteur protéique U2AF<sup>35</sup>. Ce facteur facilite la fixation du facteur U2AF<sup>65</sup> sur la seconde séquence importante, située juste en amont du site 3', la séquence polypyrimidine (Zamore et Green 1989, 1991 ; Zhang *et al.*, 1992). Enfin, la troisième séquence importante est la boîte de branchement (YNYURAC, où Y=pyrimidine, R=purine), située 18 à 40 nt en amont du site 3' d'épissage (Zamore et Green 1989, 1991 ; Zhang *et al.*, 1992). L'hydroxyle en 2' du résidu A de cette séquence est responsable de la 1<sup>ère</sup> attaque nucléophile de la réaction d'épissage. Ce résidu A est aussi appelé point de branchement. Cette séquence fixe la snRNP U2 (pour revue, Hastings et Krainer, 2001).

La réaction d'épissage consiste en deux étapes de transestérification (**Fig. 87**). Dans une première étape, l'hydroxyle en position 2' de l'adénosine du point de branchement attaque le phosphate en 5' de l'intron. La liaison phosphodiester entre l'exon 1 et l'intron est alors rompue. Cette étape libère l'exon 1, ainsi qu'un produit intermédiaire en lasso composé de l'intron et de l'exon 2. La seconde étape de la réaction consiste aussi en une attaque nucléophile, cette fois du phosphate en 3' de l'intron par le groupement hydroxyle en 3' de l'exon 1 (**Fig. 87**). La liaison phosphodiester entre l'intron et l'exon 2 est rompue. Il y a alors ligation de l'exon 1 et de l'exon 2 et libération de l'intron sous forme de lasso. Cet intron est très rapidement dégradé, ce qui permet la non réversibilité de la réaction (Guthrie et Patterson, 1988 ; Moore et Sharp, 1993 ; pour revue, Graveley, 2002).

Créer un événement d'épissage splicesosomal n'existant normalement pas dans un ARN, nécessite la création de séquences correspondant à des sites 5' et 3' d'épissage et à la boîte de branchement des ARN pré-messagers. Ceci impliquait donc, dans notre cas, la génération de mutations. Cependant, nous souhaitions préserver le plus possible la séquence de la région ADN pour que les futures études de l'effet de cet épissage additionnel dans les cellules ES de souris soient pertinentes. Après examen de la séquence bordant la région des A-repeats au sein de l'ARN Xist, nous avons sélectionné des segments qui par un nombre limité de mutations, pouvaient être convertis en sites d'épissage supposés très efficaces. En effet, il faut souligner qu'un site d'épissage conforme aux séquences consensus des sites d'épissage n'est pas forcément utilisé, car l'utilisation des sites d'épissage dépend fortement de leur contexte. Nous devons donc ensuite, après la création des sites, tester leur efficacité d'utilisation à la fois *in vitro* et *in cellulo*.



### 3.2 Choix des séquences à muter et mutagenèse

Pour générer les sites 5' et 3' d'épissage, nous avons amplifié 2 fragments d'ADN.

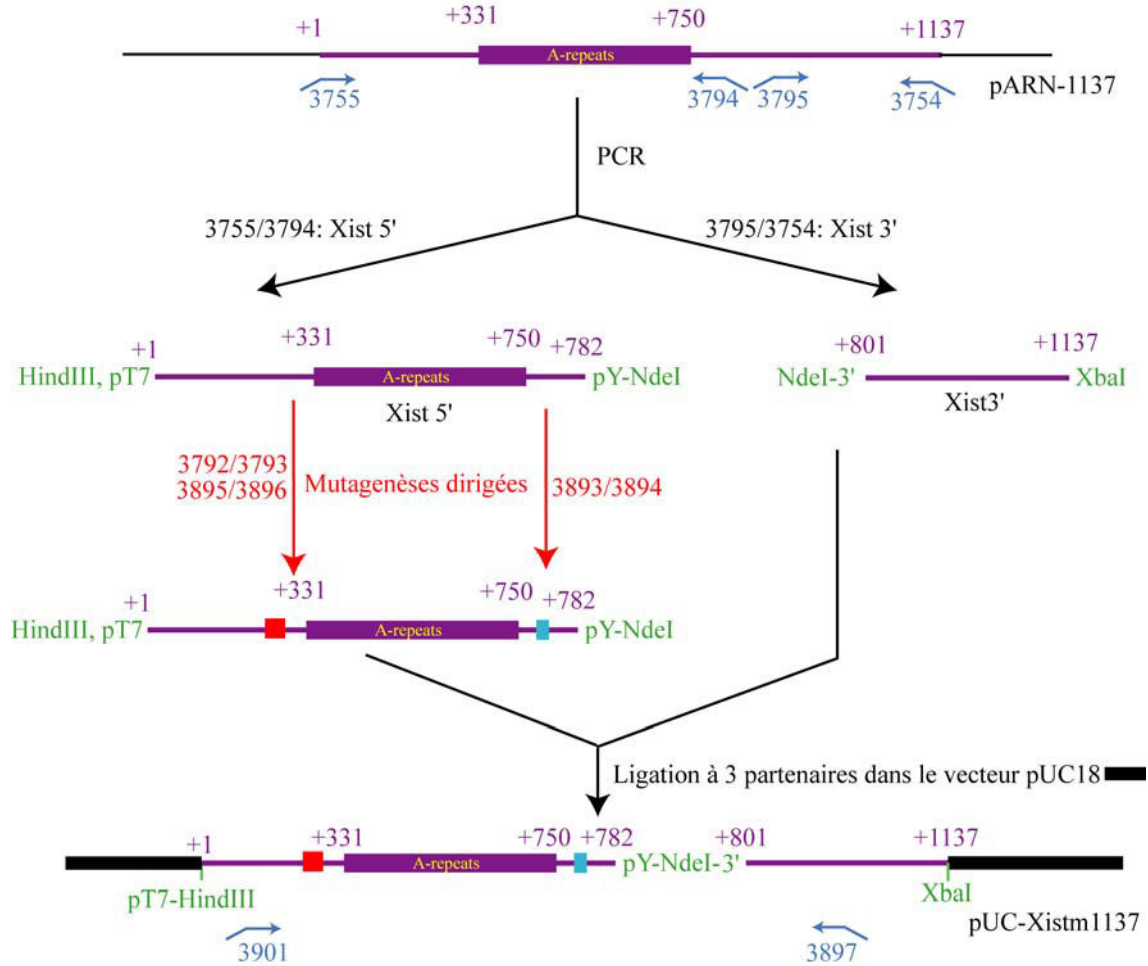
- Le premier fragment amplifié appelé Xist5' correspondait à la région du gène *Xist* correspondant au segment +1 et +782 de l'ARN *Xist*, segment qui contenait la région des A-repeats (**Fig. 88**). Ce fragment nous a permis de générer le site 5' d'épissage, la boîte de branchement et la séquence polypyrimidine. L'amplification du fragment Xist5' a été réalisée avec le couple d'oligonucléotides 3755/3794. L'oligonucléotide 3755 permettait d'introduire un site *HindIII* suivi d'une séquence correspondant au promoteur T7 en amont de la séquence de *Xist*. L'oligonucléotide 3794 permettait d'ajouter une longue séquence polypyrimidine

suivie d'un site *NdeI*. Nous verrons l'utilité de ce site *NdeI* par la suite. Ce fragment a été appelé Xist5'. Ce fragment a été cloné dans le vecteur pCR2.1 et vérifié par séquençage.

Nous avons ensuite généré sur ce fragment un site 5' d'épissage conforme à la séquence consensus des sites 5' d'épissage. Pour cela, la séquence AUUUUCAGU située entre les positions +226 à +234 (en amont de la région des A-repeats) a été convertie en site 5' d'épissage AAGGUAAGU, ceci par génération de 4 substitutions de bases par mutagenèse dirigée. Pour cela nous avons utilisé les couples d'oligonucléotides 3792/3793 et 3895/3896 (**Fig. 88 et 89**). Nous avons ensuite généré la boîte de branchement. Pour cela, la séquence UAUUGCACUA (positions +755 à +762 contenant 4 positions successives identiques à celle de la séquence consensus des boîtes de branchement UAUUAACA) a été convertie en boîte de branchement par mutagenèse dirigée en utilisant le couple d'amorces 3893/3894 (**Fig. 88 et 89**).

- Le second fragment amplifié appelé Xist3' correspondait à la région du gène *Xist* correspondant au segment de l'ARN Xist compris entre les positions +801 et +1137. Il a été amplifié avec le couple d'oligonucléotides 3795/3754. L'oligonucléotide 3795 permettait d'introduire en amont de région amplifiée du gène *Xist* un site *NdeI*, suivi d'un site 3' d'épissage conforme à la séquence consensus des sites 3' d'épissage. L'oligonucléotide 3754 permettait d'introduire un site *XbaI* en aval du fragment amplifié du gène *Xist*. Ce fragment a été cloné dans le vecteur pCR2.1 et vérifié par séquençage.

Une ligation à 3 partenaires avec comme partenaires, le fragment Xist5' préalablement soumis à une hydrolyse par les enzymes *HindIII/NdeI*, le fragment Xist3' préalablement soumis à une hydrolyse par les enzymes *NdeI/XbaI*, et le plasmide pUC18 préalablement soumis à une hydrolyse par les enzymes *HindIII/XbaI*, a permis de générer le plasmide pUC-Xistm1137 contenant la région du gène *Xist* correspondant au segment +1 à +1137 de l'ARN Xist et contenant les sites 5' et 3' d'épissage conformes aux séquences consensus. Le fragment correspondant du gène *Xist* non muté a également été cloné dans le plasmide pUC18 (plasmide pUC-Xistwt1137).

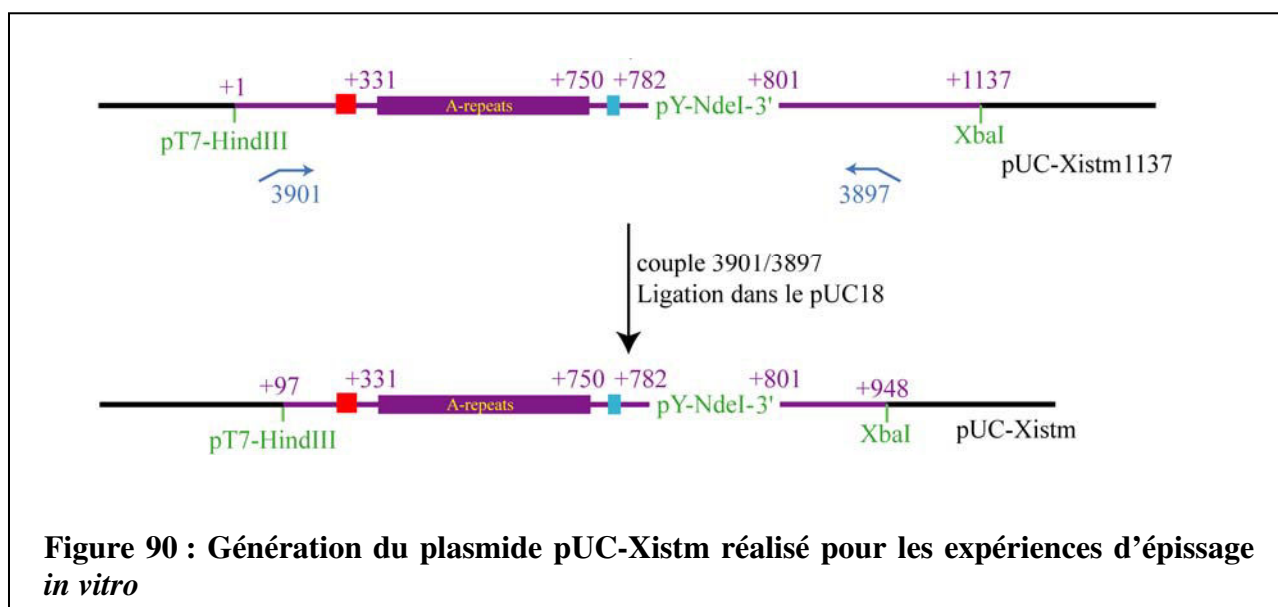


**Figure 88 : Stratégie utilisée pour l'insertion de sites d'épissage encadrant la région des A-repeats**

Les fragments Xist5' et Xist3' ont été amplifiés par PCR avec deux couples d'oligonucléotides distincts à partir du vecteur pARN-1137. Les sites de restriction, ainsi que la région polypyrimidine et le site 3' insérés sont indiqués en vert. La région des A-repeats est indiquée par un rectangle violet. Le site 5' est indiqué par un rectangle rouge. La boîte de branchement est indiquée par un rectangle bleu. Le vecteur final portant le site d'épissage complet est nommé pUC-Xistm1137.



Xist comprise entre les positions +1 à +1137. Il nous fallait donc travailler avec un fragment d'ARN plus court. Dans cet objectif, nous avons construit les plasmides pUC-Xistwt et pUC-Xistm qui produisaient par transcription *in vitro* des ARN correspondant à la région de l'ARN Xist comprise entre les positions +97 à +948, ceci sous forme sauvage ou avec les sites d'épissage que nous avons insérés. Pour cela nous avons utilisé le couple d'oligonucléotides 3901/3897 (**Fig. 90**). Les ARN correspondants ont ensuite été produits par transcription par l'ARN polymérase du phage T7, taille d'ARN compatible avec les expériences d'épissage *in vitro*.



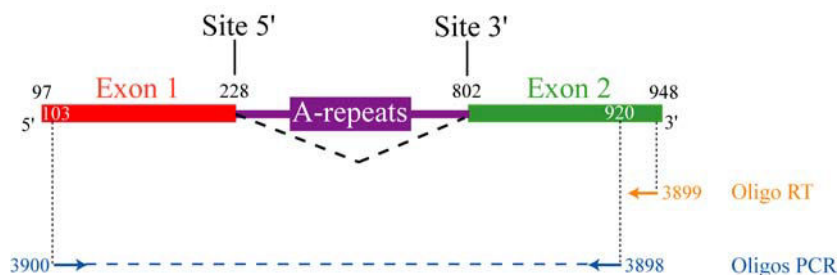
### 3.3.2 Épissage et mise en évidence des produits d'épissage

Ces deux ARN (ARN A2Cwt pour l'ARN sauvage et l'ARN A2Cm pour l'ARN muté) ont ensuite été incubés en présence d'extrait nucléaire de cellules HeLa, dans les conditions décrites dans la partie Matériel et Méthodes. Ces extraits nucléaires commerciaux sont couramment utilisés au laboratoire et sont fonctionnels pour l'épissage d'ARN pré-messagers humain mais aussi de souris.

L'apparition des produits d'épissage a été testée par RT-PCR en utilisant l'oligonucléotide 3899 comme amorce pour la transcription inverse, et en utilisant le couple d'oligonucléotides 3900/3898 pour la réaction de PCR. La position d'hybridation de ces oligonucléotides sur les transcrits utilisés est représentée dans la **Figure 91**. Les produits d'amplification attendus devaient être de 817 pb lorsque l'ARN n'était pas épissé et de 243 pb lorsque l'ARN était épissé. Les expériences de RT-PCR réalisées avec l'ARN sauvage (ARN A2Cwt) nous permettaient en tant que contrôle d'avoir la taille du fragment amplifié lorsque l'ARN n'était pas épissé.



Les produits d'amplification ont été fractionnés par électrophorèse en gel d'agarose. Une expérience témoin a été réalisée sur l'ARN sauvage (ARN A2Cwt) non incubé dans un extrait nucléaire.



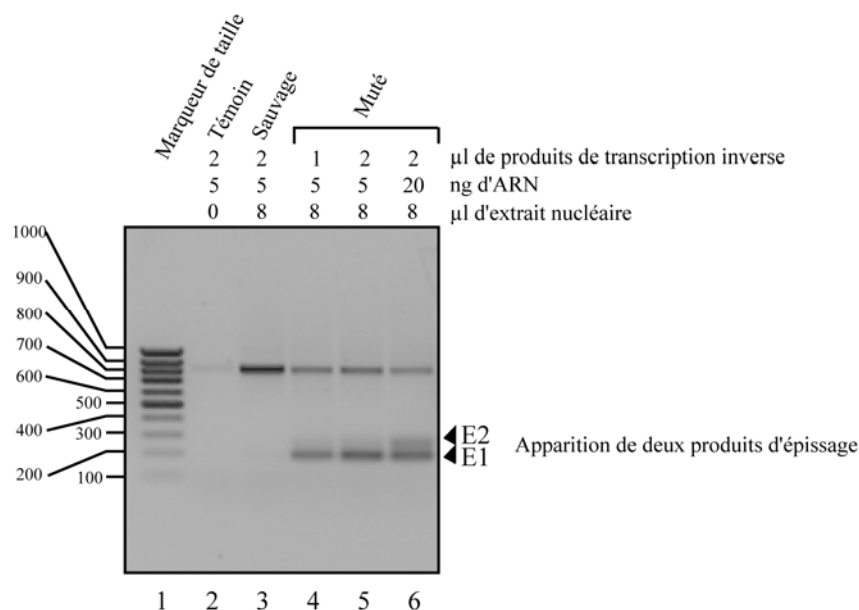
**Figure 91 : Représentation schématique de l'ARN Xist muté et des positions d'hybridation des oligonucléotides utilisés pour les expériences de RT-PCR**

Les sites d'épissage, leurs positions, le nom des oligonucléotides utilisés pour les réactions de transcription inverse et de PCR, ainsi que leurs positions d'hybridation sont indiqués sur le schéma. L'exon 1 est en rouge et l'exon 2 en vert.

### 3.3.3 Résultats

La taille du fragment amplifié lorsque l'ARN A2Cwt était en absence ou en présence d'extrait nucléaire, était la même. Ceci nous permettait de vérifier que l'ARN A2Cwt ne contenait pas de couple de sites d'épissage fonctionnels qui n'avaient pas été décrit. Lorsque l'ARN A2Cmut a été traité dans les mêmes conditions, nous avons bien obtenu 2 produits d'amplification. Un fragment d'environ 820 pb correspondant à l'ARN non épissé et un fragment E1, d'environ 200 pb (**Fig. 92**). Néanmoins, ce fragment (200 pb) avait une taille inférieure à celle attendue pour le produit d'épissage (243 pb). Une faible bande E2 contenant un fragment de taille légèrement supérieure à celle du fragment E1 était aussi détecté. Nous avons mieux visualisé cette bande additionnelle en réalisant la réaction de PCR avec 2 µl de produit de transcription inverse ou en augmentant la quantité d'ARN A2Cmut incubée avec l'extrait nucléaire. Pour identifier les 2 produits PCR E1 et E2, nous les avons clonés dans le plasmide pCR2.1 afin de les séquencer (**Fig. 93**). D'après le séquençage les fragments E1 et E2 correspondaient tous les deux à des ARN épissés au site 3' que nous avons inséré. Le fragment E2, obtenu en faible quantité, correspondait à l'utilisation du site 5' que nous avons généré. Par contre, le fragment le plus abondant E1 correspondait à l'utilisation d'un site 5' présent naturellement dans l'ARN de Xist, qui pourrait donc être un site 5' cryptique (**Fig. 93**). Ce site est situé entre les positions 174 et 184 et correspond à la séquence (exon1)CU/GGUAAGAU(intron). Ainsi, en dépit de 2 différences par rapport à la séquence consensus, ce site était préférentiellement utilisé *in vitro* par rapport au site 5' que nous avons généré, qui lui, était conforme à la séquence consensus. Ceci peut s'expliquer comme

mentionné plus haut par l'effet du contexte. Un point important était la fonctionnalité du site 3' d'épissage généré *in vitro*. Il était important de montrer qu'il en était de même *in cellulo*.



**Figure 92 : Analyse par RT-PCR du devenir des ARN-A2C sauvages et mutés après incubation dans un extrait nucléaire**

5 ng ou 20 ng d'ARN ont été incubés en présence de 8 μl d'extrait nucléaire, comme décrit dans la partie Matériel et Méthodes.

Après transcription inverse, 1 à 2 μl de mélange issus de la transcription inverse ont servi à la réalisation de l'amplification par PCR.

Fractionnement des produits de RT-PCR sur gel d'agarose 1%.

Piste 1 : Marqueur de taille

Piste 2 : Témoin, la réaction de transcription inverse a été effectuée sur l'ARN-A2Cwt sans incubation en présence d'extrait nucléaire de cellules HeLa

Piste 3 : ARN-A2Cwt

Piste 4, 5 et 6 : ARN-A2Cmut

Les quantités des différents produits sont indiquées au dessus du gel.



### 3.4 Tests d'épissage *in cellulo*

#### 3.4.1 Stratégie utilisée

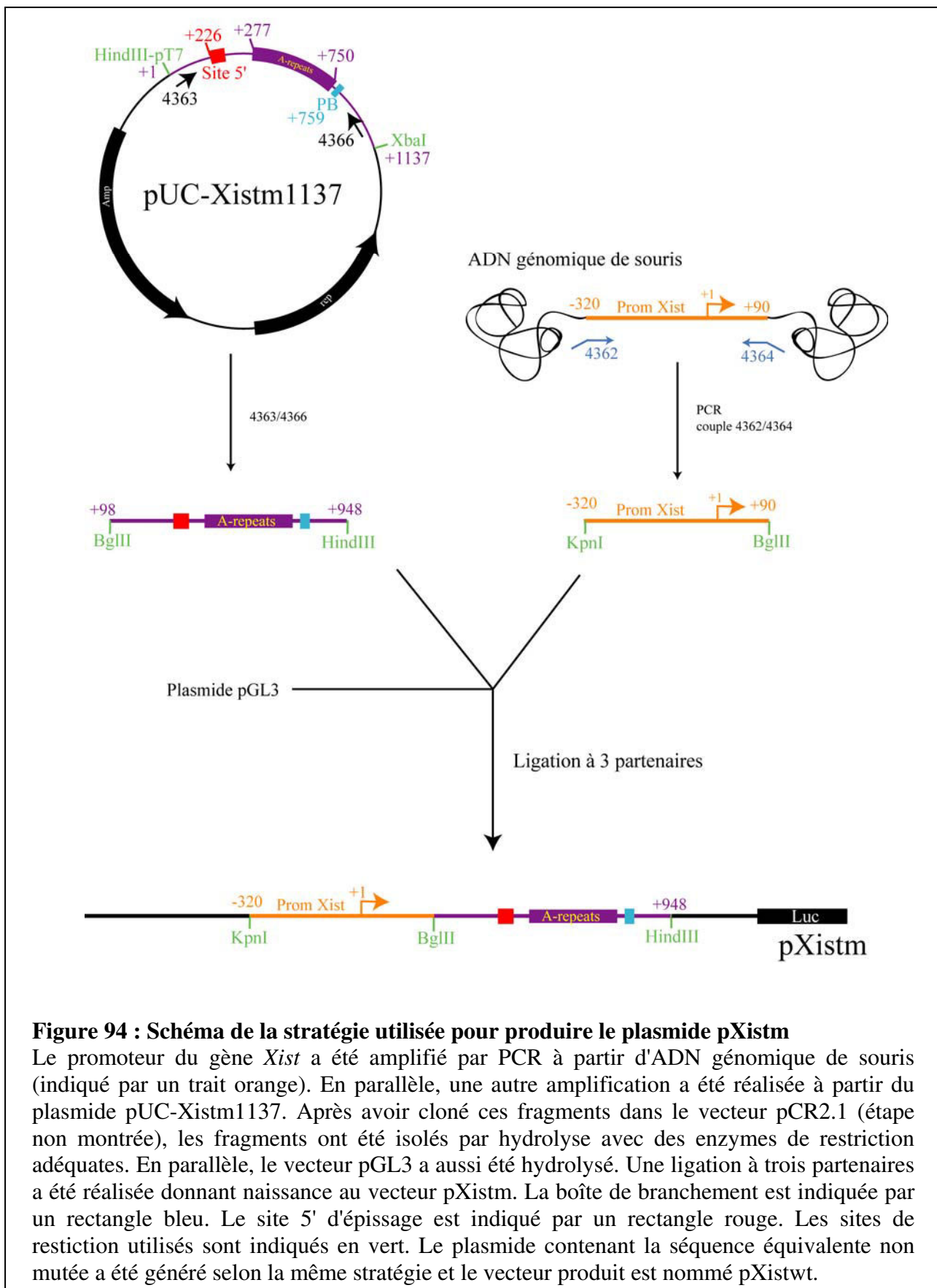
Pour tester l'efficacité de l'épissage *in cellulo*, nous avons utilisé des cellules de mammifères en culture : les cellules HeLa étaient adaptées pour cet objectif. Il nous fallait d'abord préparer des plasmides compatibles avec une transfection de cellules de mammifères qui contenaient la région de l'ARN *Xist* avec ou sans les sites 5' et 3'. Puis, les cellules étaient transfectées par les plasmides selon le protocole décrit dans le chapitre Matériel et Méthodes. Enfin, les ARN totaux des cellules transfectées pouvaient être extraits et analysés par expériences de RT-PCR.

#### 3.4.2 Préparation des plasmides contenant ou non les sites 5' et 3' d'épissage qui puissent être transfectés dans des cellules HeLa

Il nous fallait intégrer les fragments du gène *Xist* contenant ou non les sites 5' et 3' d'épissage dans un vecteur utilisé pour la transfection de cellules eucaryotes. Nous avons utilisé le plasmide pGL3. Ce plasmide porte un gène codant la luciférase, protéine qui nous a permis d'apprécier l'efficacité de transfection.

Pour se placer dans les conditions les plus proches possibles du contexte naturel, nous avons placé les fragments du gène *Xist* sous contrôle de la région promotrice de ce gène. Il a en effet été démontré que l'efficacité d'épissage peut dépendre de l'identité du promoteur utilisé. Pour cela, la région promotrice du gène *Xist* comprise entre les positions -320 à +90 a donc été amplifiée à partir d'ADN génomique de souris avec le couple d'oligonucléotides 4362/4364 (**Fig. 94**). L'oligonucléotide 4362 permettait d'ajouter un site de restriction *KpnI* en amont du fragment amplifié et l'oligonucléotide 4364 permettait d'ajouter un site de restriction *BglII* à l'extrémité opposée. En parallèle, nous avons amplifié la région du gène *Xist* mutée ou non à partir des plasmides pUC-*Xistm*1137 ou pUC-*Xistwt*1137, ceci avec le couple d'oligonucléotide 4363/4366 (**Fig. 94**). L'oligonucléotide 4363 permettait d'ajouter un site de restriction *BglII* en amont du fragment amplifié et l'oligonucléotide 4366 d'ajouter un site de restriction *HindIII* à l'autre extrémité (**Fig. 94**).

La ligation à 3 partenaires entre le plasmide pGL3 préalablement hydrolysé par les enzymes *KpnI/HindIII*, la région du gène *Xist* comprise entre les positions +98 et +948 contenant ou pas les sites 3' et 5' d'épissage préalablement hydrolysée par les enzymes *BglII/HindIII* et la région promotrice du gène *Xist* comprise entre les positions -320 à +90 préalablement hydrolysée par les enzymes *KpnI/BglII*, a permis de produire les plasmides appelés p*Xistwt* et p*Xistm*. p*Xistwt* contenait la séquence sauvage et p*Xistm* contenait la séquence contenant les sites 5' et 3' d'épissage (**Fig. 94**).



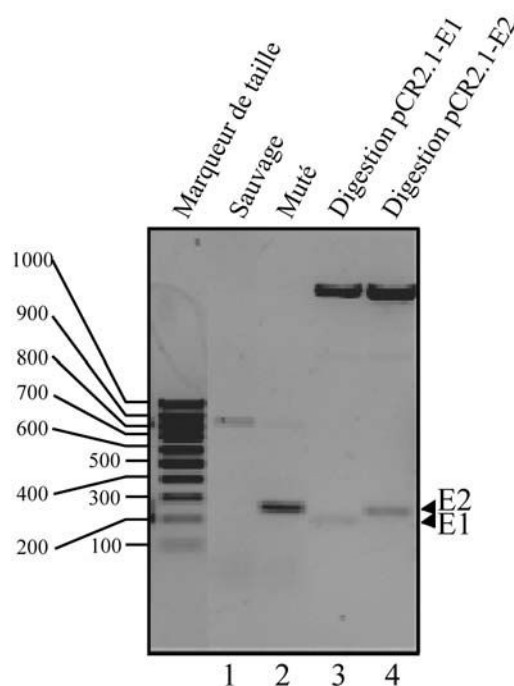
### 3.4.3 Analyse de l'épissage *in cellulo*

Les cellules HeLa ont été transfectées par les plasmides pXistm et pXistwt selon le protocole explicité dans le chapitre Matériel et Méthode. Les ARN totaux ont été extraits 48 h après la transfection et ont été analysés par RT-PCR. Ces expériences de RT-PCR ont été

réalisées dans des conditions semi-quantitatives ce qui permettait de détecter des différences de taux d'amplification des produits d'épissage, et ceci avec les mêmes oligonucléotides que ceux utilisés dans les expériences d'épissage *in vitro* (3899, 3900/3898) (**Fig. 91**). Toutes les expériences de transcription inverse ont été réalisées avec 4 µg d'ARN totaux et les amplifications par PCR avec 1 µl de produit de transcription inverse.

Pour les cellules transfectées par le plasmide pXistwt, un seul fragment d'ADN était amplifié et ce fragment ayant une taille d'environ 820 pb correspondait au fragment attendu pour l'ARN non épissé. Un fragment de plus petite taille (environ 250 pb) était obtenu pour les cellules transfectées par le vecteur pXistm (**Fig. 95**). La comparaison de la taille des fragments d'ADN obtenus lors des expériences d'épissage *in vitro* et *in cellulo* montrait que la bande obtenue *in cellulo* pouvait correspondre à un épissage de l'ARN Xist entre les deux sites générés par mutagenèse. Ceci était en accord avec la mobilité électrophorétique des produits d'hydrolyse des plasmides pCR2.1-E1 et pCR2.1-E2, contenant les fragments amplifiés à partir des produits d'épissage E1 et E2 obtenus *in vitro*. Nous en avons conclu que les sites 5' et 3' d'épissage introduits par mutagenèse dans l'ARN Xist conduisaient bien à l'épissage efficace *in cellulo*, éliminant la région des A-repeats.

Sur cette base, nous pensons donc pouvoir, lorsque la cassette LoxP-FRT-PTK aura été insérée dans un des deux chromosomes X de cellules ES de souris femelles, éliminer spécifiquement la région des A-repeats au niveau de l'ARN par insertion du fragment portant les sites 5' et 3' d'épissage.



**Figure 95 : Analyse par RT-PCR des produits d'épissage *in cellulo* des fragments de l'ARN Xist sauvage et muté**

Les cellules HeLa ont été transfectées soit par le vecteur pXistwt (piste 1), soit par le vecteur pXistm (piste 2). 48 h après transfection, les ARN totaux ont été extraits comme décrit dans la partie Matériel et Méthodes, 4 µg d'ARN totaux ont été utilisés pour la transcription inverse

avec comme amorce l'oligonucléotide 3899. Les réactions de PCR ont été effectuées à partir de 1  $\mu$ l de produit de transcription inverse avec le couple 3900/3898. La piste 1 correspond au marqueur de taille. Les pistes 3 et 4 correspondent respectivement au dépôt des produits de digestion des plasmides pCR2.1-E1 et pCR2.1-E2 par l'enzyme de restriction *EcoRI*.

**CONCLUSION GÉNÉRALE**

**ET**

**PERSPECTIVES**



Lorsque j'ai débuté ce travail de thèse, on savait que la région des A-repeats de l'ARN Xist était importante. Néanmoins, le mécanisme de son fonctionnement n'était pas connu. L'ARN Xist était supposé recruter le complexe PRC2 responsable de la mise en place des modifications H3K27me3 lors de l'inactivation du chromosome X et il était envisageable que la région des A-repeats soit responsable de ce recrutement (Plath *et al.*, 2003 ; Kohlmaier *et al.*, 2004 ; Plath *et al.*, 2004). Aucune étude expérimentale n'avait été réalisée en vue de déterminer la structure 2D de la région des A-repeats. Des travaux publiés pendant que je réalisais mon travail de thèse ont confirmé et précisé le rôle de la région des A-repeats dans l'inactivation du chromosome X (Chaumeil *et al.*, 2006). Ces travaux montraient que la région des A-repeats n'intervient pas dans la fixation de l'ARN Xist sur le futur chromosome X inactif mais contribue au recrutement des gènes du chromosome X à inactiver dans le domaine inactif. Très récemment, des travaux suggéraient fortement que la région des A-repeats puisse interagir directement avec certains composants du complexe PRC2 (Zhao *et al.*, 2008).

De notre côté, nous avons défini la structure secondaire adoptée par la région des A-repeats en solution. Ce travail a été difficile du fait de la présence des répétitions et nous avons dû utiliser à la fois des sondes chimiques et enzymatiques en solution et une approche de FRET pour réussir à identifier parmi toutes les structures possibles celle adoptée par cette région en solution.

Il n'est pas possible d'exclure le fait que cette région subisse des transitions structurales au cours de son fonctionnement, en particulier, lors de la fixation des protéines. Par dichroïsme circulaire, nous avons en effet observé que cela pourrait être le cas lors de la fixation de la protéine PTB. Néanmoins, nous avons observé aussi par dichroïsme circulaire que la grande structure tige-boucle SLS1, qui peut être formée aussi bien dans l'ARN de souris que dans l'ARN humain, est une structure stable. De plus, nos recherches de protéines partenaires de la région des A-repeats et de ses sous-fragments montrent qu'un nombre important de protéines, et en particulier de protéines fonctionnelles, nécessite la présence de ce type de structures tige-boucle comportant 4 éléments répétés pour se lier. C'est le cas des protéines FBP1, BRML1, KSRP, SWI/SNF, BAZ-1A, Dnmt1, TIF1b, Sam68, Vigiline, RHA, TIAR, DEK, H1, H2B, H4, PDS5A, RIF1, Sin3a, DIMT1, Rod1, Lin28 et des protéines du complexe PRC2.

Nos données sur la structure secondaire de la région des A-repeats en solution ne sont en accord qu'avec des structures où les répétitions sont appariées 2 à 2, ce qui va à l'encontre de la prédiction initiale de structure 2D de la région des A-repeats selon laquelle chaque répétition se replie sur elle-même. Récemment, il a été montré par RMN qu'un élément répété

de la région des A-repeats (26 nt) pris individuellement est capable de former la première des deux structures tige-boucle prédite dans le cas d'un repliement interne. Il a été proposé par cette équipe que le deuxième élément participe à la dimérisation de l'ARN mais pas selon le modèle de structure 2D que nous proposons (Duszczuk *et al.*, 2008). Il faut noter que la répétition choisie a été étudiée par RMN hors de son contexte et donc, que la validité des données obtenues dans ces conditions est faible.

Il sera très difficile d'envisager une analyse directe de la structure 2D adoptée par la région des A-repeats *in cellulo* pour les raisons suivantes :

- 1) le taux d'ARN Xist dans les cellules est faible,
- 2) *in cellulo*, il doit être associé avec de nombreuses protéines comme le suggère nos études des complexes formés en extrait nucléaire de cellules HeLa. Les segments de la région des A-repeats accessibles à des sondes *in cellulo* sont sans doute peu nombreux, d'où la difficulté d'obtenir une information structurale, avec en plus le fait que la présence d'un grand nombre de segments répétés complexifie l'analyse des résultats.

La seule solution qui nous paraît performante pour obtenir des informations sur l'importance fonctionnelle de la structure 2D établie sera de produire des ARN Xist mutés de façon : soit à abolir cette structure, soit à la renforcer, ou encore à renforcer d'autres structures possibles, ceci en utilisant la lignée de cellules ES de souris femelles dans laquelle la cassette LoxP-FRT-PTK aura été introduite dans un des deux chromosomes X. Il sera ensuite possible de tester la fonctionnalité des ARN Xist mutés. L'équipe de P Avner tente toujours à ce jour de générer cette lignée. Deux stratégies sont envisagées :

changer la taille de la région à éliminer ou déplacer légèrement ses bornes pour favoriser la recombinaison homologue au locus d'intérêt,

changer le fond génétique des bras d'homologie, ce qui permettrait de cibler l'autre allèle, en l'occurrence l'allèle *pgk*. La recombinaison serait peut-être plus efficace.

Jusque là, seules quelques protéines ont été proposées comme étant partenaires de la région des A-repeats. Nous avons été les premiers à réaliser une étude systématique des partenaires potentiels. Néanmoins, l'une des difficultés rencontrées dans l'analyse des résultats obtenus est l'absence de quantification relative des protéines par la méthode de spectrométrie de masse. L'emploi de méthode de spectrométrie de masse permettant la quantification est possible, mais est très lourd. Ainsi, nous avons identifié beaucoup de protéines pouvant avoir une fonction dans l'activité de la région des A-repeats et il nous faudra tester lesquelles le sont réellement. En plus des protéines du complexe PRC2, c'est le

cas des protéines SWI/SNF, BAZ-1A, Dnmt1, TIF1b, Sam68, Vigiline, RHA, TIAR, DEK, H1, H2B, H4, PDS5A, RIF1, Sin3a, BRML1, DIMT1, Rod1, Lin28 qui ont soit des rôles dans le métabolisme de l'ARN (Sam68, Vigiline, RHA, TIAR, DEK et DIMT1), soit dans le métabolisme de l'ADN (SWI/SNF, RIF1, BAZ-1A, Dnmt1, TIF1b, PARP1, H1, H2B et H4, PDS5A, RIF1, BRML1, Sin3a), soit sont des protéines hnRNP (hnRNPU/SAF-A, PTB, FBP1, KSRP), soit ont d'autres rôles ( Rod1 et Lin28). Parmi les protéines qui pourraient être importantes, nous soulignerons la protéine PDS5A qui, en association avec les protéines du complexe COHESINE, pourrait réguler l'expression génique en contrôlant l'architecture globale de la chromatine, en particulier la formation de boucles inter ou intra-chromosomiques. Un lien est en effet envisageable avec la fonction majeure de la région des A-repeats qui est de re-localiser les gènes liés au chromosome X à inactiver dans le domaine Xist (Chaumeil *et al.*, 2006). Nous retiendrons aussi la protéine RIF1 qui, en association avec les protéines du complexe télomérique, permet la mise en place de l'hétérochromatine par le recrutement de protéines à activité histone désacétylase. La région des A-repeats pourrait recruter RIF1 et permettre la désacétylation de la chromatine préalable à la méthylation des histones par le complexe PRC2. Nous soulignerons également la présence de la protéine Sin3a qui joue un rôle de plateforme de recrutement d'histone désacétylase HDAC dans les régions transcriptionnelles, ce qui est compatible avec un recrutement possible par la région des A-repeats.

La mise au point par P Clerc de l'approche siRNA applicable à l'étude de l'inactivation du chromosome X au sein des cellules ES de souris est un impératif afin de définir lesquelles parmi les protéines détectées sont réellement impliquées dans le processus d'inactivation.

En attendant, nous pouvons envisager plusieurs approches pour une meilleure sélection des protéines à étudier en particulier la recherche d'une co-localisation de ces protéines avec l'ARN Xist par imagerie confocale, ceci en utilisant sur les mêmes cellules à la fois des anticorps spécifiques de certaines des protéines identifiées et l'approche de FISH pour localiser l'ARN Xist, voir même la région des A-repeats. Nous avons déjà montré que certains anticorps dirigés contre certaines des protéines sont fonctionnels pour la détection de ces protéines par immunofluorescence.

Une autre approche utilisable est la méthode de RIP, basée sur un pontage aux UV *in cellulo* suivi d'une hydrolyse ménagée des ARN pontés puis d'immuno-précipitation des fragments d'ARN grâce à l'utilisation d'un anticorps dirigé contre la protéine étudiée. Dans notre cas, une recherche systématique de la région des A-repeats serait réalisée.

Nous pourrions également réaliser des expériences de pontage aux UV sur les RNP formées avec la région des A-repeats en extrait nucléaire de cellules ES récoltées à différents

stades de différenciation, les protéines et résidus ARN pontés pouvant ensuite être identifiés par des expériences de spectrométrie de masse. Ces expériences nous permettraient d'identifier les protéines qui se lient directement à la région des A-repeats et d'identifier leurs sites exacts d'interaction sur la région des A-repeats (collaboration avec U Henning au Max Plank Institut de Göttingen).

Nous avons montré que deux des protéines identifiées par spectrométrie de masse se lient à la région des A-repeats au niveau des régions simple brin riches en résidus U. Nous pensions qu'elles pourraient participer à la stabilisation des structures tige-boucle formées par 4 répétitions et leurs éléments espaceurs. Mais nos données de dichroïsme circulaire montrent que la fixation de PTB diminue l'hélicité de l'ARN SLS1M, ce qui suggère qu'elle déstabilise la structure SLS1M. Par contre, la fixation de la protéine KSRP-4KH a une influence bien moins marquée sur l'hélicité de l'ARN-SLS1M. On peut donc imaginer que certaines protéines comme KSRP aient un rôle dans le maintien de la structure à 4 éléments répétés, alors que d'autres comme PTB puisse induire un changement de structuration, ce qui pourrait moduler l'activité de la région des A-repeats en changeant le type de protéines recrutées par cette région.

Il sera important à ce stade de l'étude de savoir si les complexes formés par la région des A-repeats en extrait nucléaire de cellules ES contiennent à la fois PTB et KSRP, ou si la présence de ces 2 protéines est exclusive. Pour cela, nous testerons si dans les complexes formés en extrait nucléaire de cellules ES de souris par la région des A-repeats (purification par le système MS2/MBP), les RNP contiennent aussi KSRP et vice-versa. Si les fixations de PTB et KSRP sur la région des A-repeats sont bien compétitives, nous pourrions savoir si la fixation de l'une ou l'autre de ces protéines a lieu à un stade défini de différenciation des cellules ES. Ce type de données permettrait de savoir si des changements structuraux de la région des A-repeats pourraient intervenir au cours de la mise place de l'inactivation.

Dans ce cadre, l'application de la méthode RIP à différents stades de la différenciation des cellules ES pourra aussi nous apporter une information sur la présence de l'une ou l'autre des 2 protéines sur la région des A-repeats en fonction du stade de différenciation et donc de l'évolution du phénomène d'inactivation.

Il est important de noter que lors des expériences de purification des RNP, nous avons utilisé un ARN correspondant à l'extrémité 5' terminale de l'ARN Xist et que les données obtenues montrent que cet ARN est capable de fixer un nombre important de protéines qui se lient aussi à la région des A-repeats. Il sera intéressant d'explorer ces résultats ultérieurement, car aucune donnée ne permet d'affirmer que cette région n'a pas de rôle dans l'inactivation du chromosome X. En effet, les délétions réalisées par A Wutz visant à caractériser l'importance

fonctionnelle de la région des A-repeats supprimeraient également l'extrémité 5' terminale du gène *Xist* (Wutz *et al.*, 2002). Un point commun entre la région 5' du gène *Xist* et la région des A-repeats est la présence de grandes structures tige-boucle riches en résidus U, d'où encore un rôle possible des protéines PTB et KSRP. On pourrait imaginer que l'une ou l'autre de ces protéines, et peut-être plus particulièrement KSRP, pourrait jouer un rôle de plateforme permettant de recruter un certain nombre des autres protéines trouvées associées avec la région des A-repeats.

Dans ce cadre, il sera intéressant de produire une protéine KSRP portant une séquence additionnelle de type TAP-TAG dans les cellules ES de souris, ceci afin de purifier et d'identifier toutes les protéines associées à KSRP à différents stades de différenciation des cellules ES de souris. Il faut noter également que d'après les données de Zhao *et al.* (2008), la région 5' de l'ARN *Xist* pourrait ne pas toujours être associée à la région des A-repeats puisqu'un ARN plus court dont l'extrémité 5' se situerait juste en amont de la région des A-repeats pourrait être exprimé dans un premier temps avant la production d'un transcrit *Xist* complet.

Comme nous l'avons mentionné dans l'introduction, plusieurs grands ARN interviennent dans les régulations épigénétiques de la chromatine. C'est le cas des ARN *Air* et *Kcnq1ot1*. Il a été montré que l'ARN *Kcnq1ot1* recrute le complexe PRC2 (pour revue, Wu et Bernstein, 2008). Nous nous sommes demandé si *Kcnq1ot1* pourrait présenter une région se repliant de manière homologue à la région des A-repeats. L'examen de la structure 2D potentielle des ARN *Air* et *Kcnq1ot1* que nous avons réalisé, ne nous a pas permis de mettre en évidence d'homologie marquée avec la région des A-repeats. Il sera intéressant de poursuivre l'étude afin de définir les déterminants ARN requis pour la fixation du complexe PRC2.





- Agrelo, R., Souabni, A., Novatchkova, M., Haslinger, C., Leeb, M., Komnenovic, V., Kishimoto, H., Gresh, L., Kohwi-Shigematsu, T., Kenner, L. and Wutz, A. (2009). SATB1 defines the developmental context for gene silencing by Xist in lymphoma and embryonic cells. *Dev Cell* **16**, 507-16.
- Allaman-Pillet, N., Djemai, A., Bonny, C. and Schorderet, D. F. (2000). The 5' repeat elements of the mouse Xist gene inhibit the transcription of X-linked genes. *Gene Expr* **9**, 93-101.
- Allen, E., Horvath, S., Tong, F., Kraft, P., Spiteri, E., Riggs, A. D. and Marahrens, Y. (2003). High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. *Proc Natl Acad Sci U S A* **100**, 9940-5.
- Allen, M. D., Buckle, A. M., Cordell, S. C., Lowe, J. and Bycroft, M. (2003). The crystal structure of AF1521 a protein from *Archaeoglobus fulgidus* with homology to the non-histone domain of macroH2A. *J Mol Biol* **330**, 503-11.
- Alvarez, J. D., Yasui, D. H., Niida, H., Joh, T., Loh, D. Y. and Kohwi-Shigematsu, T. (2000). The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes Dev* **14**, 521-35.
- Ame, J. C., Spenlehauer, C. and de Murcia, G. (2004). The PARP superfamily. *Bioessays* **26**, 882-93.
- Angelopoulou, R., Lavranos, G. and Manolakou, P. (2008). Regulatory RNAs and chromatin modification in dosage compensation: a continuous path from flies to humans? *Reprod Biol Endocrinol* **6**, 12.
- Atchison, L., Ghias, A., Wilkinson, F., Bonini, N. and Atchison, M. L. (2003). Transcription factor YY1 functions as a PcG protein in vivo. *Embo J* **22**, 1347-58.
- Augui, S., Fillion, G. J., Huart, S., Nora, E., Guggiari, M., Maresca, M., Stewart, A. F. and Heard, E. (2007). Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic. *Science* **318**, 1632-6.
- Austin, C. R. and Amoroso, E. C. (1957). Sex chromatin in early cat embryos. *Exp Cell Res* **13**, 419-21.
- Auweter, S. D. and Allain, F. H. (2008). Structure-function relationships of the polypyrimidine tract binding protein. *Cell Mol Life Sci* **65**, 516-27.
- Avner, P. and Heard, E. (2001). X-chromosome inactivation: counting, choice and initiation. *Nat Rev Genet* **2**, 59-67.
- Avner, P., Prissette, M., Arnaud, D., Courtier, B., Cecchi, C. and Heard, E. (1998). Molecular correlates of the murine Xce locus. *Genet Res* **72**, 217-24.
- Bacher, C. P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., Eils, R. and Heard, E. (2006). Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat Cell Biol* **8**, 293-9.
- Bachler, M., Schroeder, R. and von Ahsen, U. (1999). StreptoTag: a novel method for the isolation of RNA-binding proteins. *Rna* **5**, 1509-16.
- Bai, X., Alekseyenko, A. A. and Kuroda, M. I. (2004). Sequence-specific targeting of MSL complex regulates transcription of the roX RNA genes. *Embo J* **23**, 2853-61.
- Bailey, J. A., Carrel, L., Chakravarti, A. and Eichler, E. E. (2000). Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc Natl Acad Sci U S A* **97**, 6634-9.
- Barlow, D. P., Stoger, R., Herrmann, B. G., Saito, K. and Schweifer, N. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature* **349**, 84-7.
- 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**, 676.
- Barr, M.L. (1963) Some properties of the sex chromosomes and their bearing on normal and abnormal development. *Second International Conference on Congenital Malformations*
- Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-37.
- Baumann, C., Schmidtman, A., Muegge, K. and De La Fuente, R. (2008). Association of ATRX with pericentric heterochromatin and the Y chromosome of neonatal mouse spermatogonia. *BMC Mol Biol* **9**, 29.
- Beletskii, A., Hong, Y. K., Pehrson, J., Egholm, M. and Strauss, W. M. (2001). PNA interference mapping demonstrates functional domains in the noncoding RNA Xist. *Proc Natl Acad Sci U S A* **98**, 9215-20.
- Bell, A. C. and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482-5.
- Berglund, F. M. and Clarke, P. R. (2009). hnRNP-U is a specific DNA-dependent protein kinase substrate phosphorylated in response to DNA double-strand breaks. *Biochem Biophys Res Commun* **381**, 59-64.
- Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., 3rd, Gingeras, T. R., Schreiber, S. L. and Lander, E. S. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169-81.
- Bernstein, E. and Allis, C. D. (2005). RNA meets chromatin. *Genes Dev* **19**, 1635-55.
- Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E. and Allis, C. D. (2006). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol Cell Biol* **26**, 2560-9.



- Bernstein, E., Muratore-Schroeder, T. L., Diaz, R. L., Chow, J. C., Changolkar, L. N., Shabanowitz, J., Heard, E., Pehrson, J. R., Hunt, D. F. and Allis, C. D. (2008). A phosphorylated subpopulation of the histone variant macroH2A1 is excluded from the inactive X chromosome and enriched during mitosis. *Proc Natl Acad Sci U S A* **105**, 1533-8.
- Bestor, T. H. (2000). The DNA methyltransferases of mammals. *Hum Mol Genet* **9**, 2395-402.
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21.
- Blewitt, M. E., Gendrel, A. V., Pang, Z., Sparrow, D. B., Whitelaw, N., Craig, J. M., Apedaile, A., Hilton, D. J., Dunwoodie, S. L., Brockdorff, N., Kay, G. F. and Whitelaw, E. (2008). SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat Genet* **40**, 663-9.
- Blewitt, M. E., Vickaryous, N. K., Paldi, A., Koseki, H. and Whitelaw, E. (2006). Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. *PLoS Genet* **2**, e49.
- Bode, J., Schlake, T., Iber, M., Schubeler, D., Seibler, J., Snezhkov, E. and Nikolaev, L. (2000). The transgeneticist's toolbox: novel methods for the targeted modification of eukaryotic genomes. *Biol Chem* **381**, 801-13.
- 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. *Nat Genet* **30**, 73-6.
- Borggreffe, T., Wabl, M., Akhmedov, A. T. and Jessberger, R. (1998). A B-cell-specific DNA recombination complex. *J Biol Chem* **273**, 17025-35.
- Borsani, G., Tonlorenzi, R., Simmler, M. C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C., Willard, H., Avner, P. and Ballabio, A. (1991). Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351**, 325-9.
- Boumil, R. M. and Lee, J. T. (2001). Forty years of decoding the silence in X-chromosome inactivation. *Hum Mol Genet* **10**, 2225-32.
- Bourc'his, D. and Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96-9.
- Boyd, Y. and Fraser, N. J. (1990). Methylation patterns at the hypervariable X-chromosome locus DXS255 (M27 beta): correlation with X-inactivation status. *Genomics* **7**, 182-7.
- Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A. and Jaenisch, R. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-53.
- Branco, M. R. and Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol* **4**, e138.
- Brockdorff, N. (2002). X-chromosome inactivation: closing in on proteins that bind Xist RNA. *Trends Genet* **18**, 352-8.
- Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S. and 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**, 515-26.
- Brown, C. J. and Baldry, S. E. (1996). Evidence that heteronuclear proteins interact with XIST RNA in vitro. *Somat Cell Mol Genet* **22**, 403-17.
- Brown, C. J. and Greally, J. M. (2003). A stain upon the silence: genes escaping X inactivation. *Trends Genet* **19**, 432-8.
- Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, 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**, 527-42.
- Brown, C. J. and Willard, H. F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. *Nature* **368**, 154-6.
- Brown, J. L., Mucci, D., Whiteley, M., Dirksen, M. L. and Kassis, J. A. (1998). The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Mol Cell* **1**, 1057-64.
- Brown, S. D. (1991). XIST and the mapping of the X chromosome inactivation centre. *Bioessays* **13**, 607-12.
- Brzovic, P. S., Keeffe, J. R., Nishikawa, H., Miyamoto, K., Fox, D., 3rd, Fukuda, M., Ohta, T. and Klevit, R. (2003). Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex. *Proc Natl Acad Sci U S A* **100**, 5646-51.
- Brzovic, P. S., Rajagopal, P., Hoyt, D. W., King, M. C. and Klevit, R. E. (2001). Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. *Nat Struct Biol* **8**, 833-7.
- Buhler, M., Verdel, A. and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**, 873-86.
- Burdon, T., Smith, A. and Savatier, P. (2002). Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol* **12**, 432-8.
- Burgoyne, P. S., Evans, E. P. and Holland, K. (1983). XO monosomy is associated with reduced birthweight and lowered weight gain in the mouse. *J Reprod Fertil* **68**, 381-5.

- Cai, S., Han, H. J. and Kohwi-Shigematsu, T. (2003). Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat Genet* **34**, 42-51.
- Cao, R., Tsukada, Y. and Zhang, Y. (2005). Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol Cell* **20**, 845-54.
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S. and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-43.
- Cao, R. and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr Opin Genet Dev* **14**, 155-64.
- Caparros, M. L., Alexiou, M., Webster, Z. and Brockdorff, N. (2002). Functional analysis of the highly conserved exon IV of XIST RNA. *Cytogenet Genome Res* **99**, 99-105.
- Carrel, L., Cottle, A. A., Goglin, K. C. and Willard, H. F. (1999). A first-generation X-inactivation profile of the human X chromosome. *Proc Natl Acad Sci U S A* **96**, 14440-4.
- Carrel, L. and Willard, H. F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**, 400-4.
- Carthew, R. W. (2001). Gene silencing by double-stranded RNA. *Curr Opin Cell Biol* **13**, 244-8.
- Castelo-Branco, P., Furger, A., Wollerton, M., Smith, C., Moreira, A. and Proudfoot, N. (2004). Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol Cell Biol* **24**, 4174-83.
- Chadwick, B. P. and Willard, H. F. (2001a). Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. *Hum Mol Genet* **10**, 1101-13.
- Chadwick, B. P. and Willard, H. F. (2001b). A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J Cell Biol* **152**, 375-84.
- Chadwick, B. P. and Willard, H. F. (2002). Cell cycle-dependent localization of macroH2A in chromatin of the inactive X chromosome. *J Cell Biol* **157**, 1113-23.
- Chadwick, B. P. and Willard, H. F. (2003). Chromatin of the Barr body: histone and non-histone proteins associated with or excluded from the inactive X chromosome. *Hum Mol Genet* **12**, 2167-78.
- Chadwick, B. P. and Willard, H. F. (2004). Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome. *Proc Natl Acad Sci U S A* **101**, 17450-5.
- Chakravarthy, S., Gundimella, S. K., Caron, C., Perche, P. Y., Pehrson, J. R., Khochbin, S. and Luger, K. (2005). Structural characterization of the histone variant macroH2A. *Mol Cell Biol* **25**, 7616-24.
- Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643-55.
- Chambeyron, S. and Bickmore, W. A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev* **18**, 1119-30.
- Changolkar, L. N., Costanzi, C., Leu, N. A., Chen, D., McLaughlin, K. J. and Pehrson, J. R. (2007). Developmental changes in histone macroH2A1-mediated gene regulation. *Mol Cell Biol* **27**, 2758-64.
- 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* **295**, 345-7.
- Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. *Curr Biol* **6**, 149-62.
- Charlesworth, D., Charlesworth, B. and Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95**, 118-28.
- 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 Dev* **20**, 2223-37.
- Chaumeil, J., Okamoto, I., Guggiari, M. and Heard, E. (2002). Integrated kinetics of X chromosome inactivation in differentiating embryonic stem cells. *Cytogenet Genome Res* **99**, 75-84.
- Chen, F., Kamradt, M., Mulcahy, M., Byun, Y., Xu, H., McKay, M. J. and Cryns, V. L. (2002). Caspase proteolysis of the cohesin component RAD21 promotes apoptosis. *J Biol Chem* **277**, 16775-81.
- Cheng, X. and Blumenthal, R. M. (2008). Mammalian DNA methyltransferases: a structural perspective. *Structure* **16**, 341-50.
- Choo, J. H., Kim, J. D., Chung, J. H., Stubbs, L. and Kim, J. (2006). Allele-specific deposition of macroH2A1 in imprinting control regions. *Hum Mol Genet* **15**, 717-24.
- Chow, J. and Heard, E. (2009). X inactivation and the complexities of silencing a sex chromosome. *Curr Opin Cell Biol* **21**, 359-66.
- 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. *Proc Natl Acad Sci U S A* **104**, 10104-9.
- Chu, F., Nusinow, D. A., Chalkley, R. J., Plath, K., Panning, B. and Burlingame, A. L. (2006). Mapping post-translational modifications of the histone variant MacroH2A1 using tandem mass spectrometry. *Mol Cell Proteomics* **5**, 194-203.
- Chuang, L. S., Ian, H. I., Koh, T. W., Ng, H. H., Xu, G. and Li, B. F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* **277**, 1996-2000.

- Chureau, C., Prissette, M., Bourdet, A., Barbe, V., Cattolico, L., Jones, L., Eggen, A., Avner, P. and Duret, L. (2002). Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. *Genome Res* **12**, 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 Genet* **2**, e94.
- Ciavatta, D., Kalantry, S., Magnuson, T. and Smithies, O. (2006). A DNA insulator prevents repression of a targeted X-linked transgene but not its random or imprinted X inactivation. *Proc Natl Acad Sci U S A* **103**, 9958-63.
- Clemson, C. M., Chow, J. C., Brown, C. J. and Lawrence, J. B. (1998). Stabilization and localization of Xist RNA are controlled by separate mechanisms and are not sufficient for X inactivation. *J Cell Biol* **142**, 13-23.
- 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. *Proc Natl Acad Sci U S A* **103**, 7688-93.
- 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. *J Cell Biol* **132**, 259-75.
- 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**, 249-53.
- Clerc, P. and Avner, P. (2003). Multiple elements within the Xic regulate random X inactivation in mice. *Semin Cell Dev Biol* **14**, 85-92.
- Clery, A., Blatter, M. and Allain, F. H. (2008). RNA recognition motifs: boring? Not quite. *Curr Opin Struct Biol* **18**, 290-8.
- Cohen, D. E., Davidow, L. S., Erwin, J. A., Xu, N., Warshawsky, D. and Lee, J. T. (2007). The DXPas34 repeat regulates random and imprinted X inactivation. *Dev Cell* **12**, 57-71.
- Cooper, P., Keer, J. T., McCabe, V. M., Hamvas, R. M., Brown, S. D., 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**, 570-5.
- Copps, K., Richman, R., Lyman, L. M., Chang, K. A., Rampersad-Ammons, J. and Kuroda, M. I. (1998). Complex formation by the Drosophila MSL proteins: role of the MSL2 RING finger in protein complex assembly. *Embo J* **17**, 5409-17.
- Costanzi, C. and Pehrson, J. R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* **393**, 599-601.
- Costanzi, C. and Pehrson, J. R. (2001). MACROH2A2, a new member of the MARCOH2A core histone family. *J Biol Chem* **276**, 21776-84.
- Costanzi, C., Stein, P., Worrada, D. M., Schultz, R. M. and Pehrson, J. R. (2000). Histone macroH2A1 is concentrated in the inactive X chromosome of female preimplantation mouse embryos. *Development* **127**, 2283-9.
- Courtier, B., Heard, E. and Avner, P. (1995). Xce haplotypes show modified methylation in a region of the active X chromosome lying 3' to Xist. *Proc Natl Acad Sci U S A* **92**, 3531-5.
- Cremer, M., von Hase, J., Volm, T., Brero, A., Kreth, G., Walter, J., Fischer, C., Solovei, I., Cremer, C. and Cremer, T. (2001). Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome Res* **9**, 541-67.
- Cremer, T., Cremer, M., Dietzel, S., Muller, S., Solovei, I. and Fakan, S. (2006). Chromosome territories--a functional nuclear landscape. *Curr Opin Cell Biol* **18**, 307-16.
- Crews, D. (2003). Sex determination: where environment and genetics meet. *Evol Dev* **5**, 50-5.
- 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. *Nat Genet* **22**, 323-4.
- Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A. and Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185-96.
- Damelin, M. and Bestor, T. H. (2007). Biological functions of DNA methyltransferase 1 require its methyltransferase activity. *Mol Cell Biol* **27**, 3891-9.
- de la Cruz, X., Lois, S., Sanchez-Molina, S. and Martinez-Balbas, M. A. (2005a). Do protein motifs read the histone code? *Bioessays* **27**, 164-75.
- de la Cruz, C. C., Fang, J., Plath, K., Worringer, K. A., Nusinow, D. A., Zhang, Y. and Panning, B. (2005b). Developmental regulation of Suz 12 localization. *Chromosoma* **114**, 183-92.
- De La Fuente, R., Viveiros, M. M., Wigglesworth, K. and Eppig, J. J. (2004). ATRX, a member of the SNF2 family of helicase/ATPases, is required for chromosome alignment and meiotic spindle organization in metaphase II stage mouse oocytes. *Dev Biol* **272**, 1-14.
- de Laat, W. and Grosveld, F. (2003). Spatial organization of gene expression: the active chromatin hub. *Chromosome Res* **11**, 447-59.

- de Napoles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., Nesterova, T. B., Silva, J., Otte, A. P., Vidal, M., Koseki, H. and Brockdorff, N. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev Cell* **7**, 663-76.
- de Napoles, M., Nesterova, T. and Brockdorff, N. (2007). Early loss of Xist RNA expression and inactive X chromosome associated chromatin modification in developing primordial germ cells. *PLoS One* **2**, e860.
- Debrand, E., Chureau, C., Arnaud, D., Avner, P. and Heard, E. (1999). Functional analysis of the DXPas34 locus, a 3' regulator of Xist expression. *Mol Cell Biol* **19**, 8513-25.
- Debrand, E., Heard, E. and Avner, P. (1998). Cloning and localization of the murine Xpct gene: evidence for complex rearrangements during the evolution of the region around the Xist gene. *Genomics* **48**, 296-303.
- DeCerbo, J. and Carmichael, G. G. (2005). Retention and repression: fates of hyperedited RNAs in the nucleus. *Curr Opin Cell Biol* **17**, 302-8.
- Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C. L., Kastner, B., Stark, H., Urlaub, H. and Luhrmann, R. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol* **26**, 5528-43.
- Deng, C. X. (2006). BRCA1: cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution. *Nucleic Acids Res* **34**, 1416-26.
- Diaz-Perez, S., Ouyang, Y., Perez, V., Cisneros, R., Regelson, M. and Marahrens, Y. (2005). The element(s) at the nontranscribed Xist locus of the active X chromosome controls chromosomal replication timing in the mouse. *Genetics* **171**, 663-72.
- Dignam, J. D., Martin, P. L., Shastry, B. S. and Roeder, R. G. (1983). Eukaryotic gene transcription with purified components. *Methods Enzymol* **101**, 582-98.
- Dindot, S. V., Farin, P. W., Farin, C. E., Romano, J., Walker, S., Long, C. and Piedrahita, J. A. (2004). Epigenetic and genomic imprinting analysis in nuclear transfer derived Bos gaurus/Bos taurus hybrid fetuses. *Biol Reprod* **71**, 470-8.
- 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. *Mol Cell* **25**, 43-56.
- Downs, K. (2001). Mouse Early Development: Molecular Basis, *Encyclopedia of life sciences*
- Duret, L., Chureau, C., Samain, S., Weissenbach, J. and Avner, P. (2006). The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* **312**, 1653-5.
- 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 Res* **36**, 7068-77.
- Duthie, S. M., Nesterova, T. B., Formstone, E. J., Keohane, A. M., Turner, B. M., Zakian, S. M. and Brockdorff, N. (1999). Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. *Hum Mol Genet* **8**, 195-204.
- Edwards, C. A. and Ferguson-Smith, A. C. (2007). Mechanisms regulating imprinted genes in clusters. *Curr Opin Cell Biol* **19**, 281-9.
- Eggan, K., Akutsu, H., Hochedlinger, K., Rideout, W., 3rd, Yanagimachi, R. and Jaenisch, R. (2000). X-Chromosome inactivation in cloned mouse embryos. *Science* **290**, 1578-81.
- Eggert, H., Schulz, M., Fackelmayer, F. O., Renkawitz, R. and Eggert, M. (2001). Effects of the heterogeneous nuclear ribonucleoprotein U (hnRNP U/SAF-A) on glucocorticoid-dependent transcription in vivo. *J Steroid Biochem Mol Biol* **78**, 59-65.
- Elderkin, S., Maertens, G. N., Endoh, M., Mallery, D. L., Morrice, N., Koseki, H., Peters, G., Brockdorff, N. and Hiom, K. (2007). A phosphorylated form of Mel-18 targets the Ring1B histone H2A ubiquitin ligase to chromatin. *Mol Cell* **28**, 107-20.
- Engemann, S., Stroedicke, M., Paulsen, M., Franck, O., Reinhardt, R., Lane, N., Reik, W. and Walter, J. (2000). Sequence and functional comparison in the Beckwith-Wiedemann region: implications for a novel imprinting centre and extended imprinting. *Hum Mol Genet* **9**, 2691-706.
- Ercan, S., Giresi, P. G., Whittle, C. M., Zhang, X., Green, R. D. and Lieb, J. D. (2007). X chromosome repression by localization of the C. elegans dosage compensation machinery to sites of transcription initiation. *Nat Genet* **39**, 403-8.
- Ercan, S. and Lieb, J. D. (2009). C. elegans dosage compensation: a window into mechanisms of domain-scale gene regulation. *Chromosome Res* **17**, 215-27.
- Evans, H. J., Ford, C. E., Lyon, M. F. and Gray, J. (1965). DNA replication and genetic expression in female mice with morphologically distinguishable X chromosomes. *Nature* **206**, 900-3.
- Fackelmayer, F. O. (2005). A stable proteinaceous structure in the territory of inactive X chromosomes. *J Biol Chem* **280**, 1720-3.
- Fackelmayer, F. O. and Richter, A. (1994). hnRNP-U/SAF-A is encoded by two differentially polyadenylated mRNAs in human cells. *Biochim Biophys Acta* **1217**, 232-4.
- Falconer, D. S. (1953). [Total sex-linkage in the house mouse.]. *Z Indukt Abstamm Vererbungsl* **85**, 210-9.
- Fan, Y., Nikitina, T., Zhao, J., Fleury, T. J., Bhattacharyya, R., Bouhassira, E. E., Stein, A., Woodcock, C. L. and Skoultchi, A. I. (2005). Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* **123**, 1199-212.

- Fang, J., Chen, T., Chadwick, B., Li, E. and Zhang, Y. (2004). Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation. *J Biol Chem* **279**, 52812-5.
- Felsenfeld, G. and Groudine, M. (2003). Controlling the double helix. *Nature* **421**, 448-53.
- Filippova, G. N., Cheng, M. K., Moore, J. M., Truong, J. P., Hu, Y. J., Nguyen, D. K., Tsuchiya, K. D. and Disteche, C. M. (2005). Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. *Dev Cell* **8**, 31-42.
- Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D. and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* **17**, 1870-81.
- Fitzpatrick, G. V., Soloway, P. D. and Higgins, M. J. (2002). Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat Genet* **32**, 426-31.
- Fluckiger, A. C., Dehay, C. and Savatier, P. (2003). [Embryonic stem cells and cell replacement therapies in the nervous system]. *Med Sci (Paris)* **19**, 699-708.
- Francis, N. J., Follmer, N. E., Simon, M. D., Aghia, G. and Butler, J. D. (2009). Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. *Cell* **137**, 110-22.
- Francis, N. J., Saurin, A. J., Shao, Z. and Kingston, R. E. (2001). Reconstitution of a functional core polycomb repressive complex. *Mol Cell* **8**, 545-56.
- Fraser, A. S., Nay, T. and Turner, H. N. (1953). Growth of the mouse coat. II. Effect of sex and pregnancy. *Aust J Biol Sci* **6**, 645-56.
- Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., Chiba, T. and Nakao, M. (2003). Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. *J Biol Chem* **278**, 24132-8.
- Fuks, F. (2005). DNA methylation and histone modifications: teaming up to silence genes. *Curr Opin Genet Dev* **15**, 490-5.
- Fuks, F., Hurd, P. J., Deplus, R. and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res* **31**, 2305-12.
- Ganesan, S., Richardson, A. L., Wang, Z. C., Iglehart, J. D., Miron, A., Feunteun, J., Silver, D. and Livingston, D. M. (2005). Abnormalities of the inactive X chromosome are a common feature of BRCA1 mutant and sporadic basal-like breast cancer. *Cold Spring Harb Symp Quant Biol* **70**, 93-7.
- Ganesan, S., Silver, D. P., Drapkin, R., Greenberg, R., Feunteun, J. and Livingston, D. M. (2004). Association of BRCA1 with the inactive X chromosome and XIST RNA. *Philos Trans R Soc Lond B Biol Sci* **359**, 123-8.
- Ganesan, S., Silver, D. P., Greenberg, R. A., Avni, D., Drapkin, R., Miron, A., Mok, S. C., Randrianarison, V., Brodie, S., Salstrom, J., Rasmussen, T. P., Klimke, A., Marrese, C., Marahrens, Y., Deng, C. X., Feunteun, J. and Livingston, D. M. (2002). BRCA1 supports XIST RNA concentration on the inactive X chromosome. *Cell* **111**, 393-405.
- Garcia-Mayoral, M. F., Hollingworth, D., Masino, L., Diaz-Moreno, I., Kelly, G., Gherzi, R., Chou, C. F., Chen, C. Y. and Ramos, A. (2007). The structure of the C-terminal KH domains of KSRP reveals a noncanonical motif important for mRNA degradation. *Structure* **15**, 485-98.
- Garrick, D., Sharpe, J. A., Arkell, R., Dobbie, L., Smith, A. J., Wood, W. G., Higgs, D. R. and Gibbons, R. J. (2006). Loss of Atrx affects trophoblast development and the pattern of X-inactivation in extraembryonic tissues. *PLoS Genet* **2**, e58.
- Gartler, S. M. and Riggs, A. D. (1983). Mammalian X-chromosome inactivation. *Annu Rev Genet* **17**, 155-90.
- Gavory, G., Symmons, M. F., Krishnan Ghosh, Y., Klenerman, D. and Balasubramanian, S. (2006). Structural analysis of the catalytic core of human telomerase RNA by FRET and molecular modeling. *Biochemistry* **45**, 13304-11.
- Geiman, T. M., Sankpal, U. T., Robertson, A. K., Zhao, Y. and Robertson, K. D. (2004). DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. *Biochem Biophys Res Commun* **318**, 544-55.
- Gherzi, R., Lee, K. Y., Briata, P., Wegmuller, 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. *Mol Cell* **14**, 571-83.
- Giacalone, J., Friedes, J. and Francke, U. (1992). A novel GC-rich human macrosatellite VNTR in Xq24 is differentially methylated on active and inactive X chromosomes. *Nat Genet* **1**, 137-43.
- Giannelli, F. (1963). The pattern of X-chromosome deoxy-ribonucleic acid synthesis in two women with abnormal sex-chromosome complements. *Lancet* **1**, 863-5.
- Gibbons, R. J., McDowell, T. L., Raman, S., O'Rourke, D. M., Garrick, D., Ayyub, H. and Higgs, D. R. (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat Genet* **24**, 368-71.
- Gilbert, S. L., Pehrson, J. R. and Sharp, P. A. (2000). XIST RNA associates with specific regions of the inactive X chromatin. *J Biol Chem* **275**, 36491-4.
- Giorda, R., Bonaglia, M. C., Milani, G., Baroncini, A., Spada, F., Beri, S., Menozzi, G., Rusconi, M. and Zuffardi, O. (2008). Molecular and cytogenetic analysis of the spreading of X inactivation in a girl with microcephaly, mild dysmorphic features and t(X;5)(q22.1;q31.1). *Eur J Hum Genet* **16**, 897-905.

- Goodfellow, P. J., Mondello, C., Darling, S. M., Pym, B., Little, P. and Goodfellow, P. N. (1988). Absence of methylation of a CpG-rich region at the 5' end of the MIC2 gene on the active X, the inactive X, and the Y chromosome. *Proc Natl Acad Sci U S A* **85**, 5605-9.
- Gordon, S., Akopyan, G., Garban, H. and Bonavida, B. (2006). Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene* **25**, 1125-42.
- Goto, Y. and Takagi, N. (1998). Tetraploid embryos rescue embryonic lethality caused by an additional maternally inherited X chromosome in the mouse. *Development* **125**, 3353-63.
- Goto, Y. and Takagi, N. (1999). [Control of X chromosome inactivation in mammals]. *Tanpakushitsu Kakusan Koso* **44**, 1682-90.
- Gowher, H., Stockdale, C. J., Goyal, R., Ferreira, H., Owen-Hughes, T. and Jeltsch, A. (2005). De novo methylation of nucleosomal DNA by the mammalian Dnmt1 and Dnmt3A DNA methyltransferases. *Biochemistry* **44**, 9899-904.
- 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**, 59-74.
- Graveley, B. R. (2002). Sex, AGility, and the regulation of alternative splicing. *Cell* **109**, 409-12.
- Graves, J. A. (2006). Sex chromosome specialization and degeneration in mammals. *Cell* **124**, 901-14.
- Grewal, S. I., Bonaduce, M. J. and Klar, A. J. (1998). Histone deacetylase homologs regulate epigenetic inheritance of transcriptional silencing and chromosome segregation in fission yeast. *Genetics* **150**, 563-76.
- Grewal, S. I. and Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. *Science* **301**, 798-802.
- Gribnau, J., Luikenhuis, S., Hochedlinger, K., Monkhorst, K. and Jaenisch, R. (2005). X chromosome choice occurs independently of asynchronous replication timing. *J Cell Biol* **168**, 365-73.
- Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U. and Cavalli, G. (2006). RNAi components are required for nuclear clustering of Polycomb group response elements. *Cell* **124**, 957-71.
- Grumbach, M. M., Morishima, A. and Taylor, J. H. (1963). Human Sex Chromosome Abnormalities in Relation to DNA Replication and Heterochromatinization. *Proc Natl Acad Sci U S A* **49**, 581-9.
- Gu, W., Szauter, P. and Lucchesi, J. C. (1998). Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of *Drosophila melanogaster*. *Dev Genet* **22**, 56-64.
- Gudmundsdottir, K. and Ashworth, A. (2006). The roles of BRCA1 and BRCA2 and associated proteins in the maintenance of genomic stability. *Oncogene* **25**, 5864-74.
- Guthrie, C. and Patterson, B. (1988). Spliceosomal snRNAs. *Annu. Rev. Genet.* **22**, 387-419.
- Hadjantonakis, A. K., Cox, L. L., Tam, P. P. and Nagy, A. (2001). An X-linked GFP transgene reveals unexpected paternal X-chromosome activity in trophoblastic giant cells of the mouse placenta. *Genesis* **29**, 133-40.
- Hadjantonakis, A. K., Dickinson, M. E., Fraser, S. E. and Papaioannou, V. E. (2003). Technicolour transgenics: imaging tools for functional genomics in the mouse. *Nat Rev Genet* **4**, 613-25.
- Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U. C., Cesari, F., Lee, C., Almouzni, G., Schneider, R. and Surani, M. A. (2008). Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**, 877-81.
- Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J. and Surani, M. A. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**, 15-23.
- Hakimi, M. A., Bochar, D. A., Schmiesing, J. A., Dong, Y., Barak, O. G., Speicher, D. W., Yokomori, K. and Shiekhhattar, R. (2002). A chromatin remodelling complex that loads cohesin onto human chromosomes. *Nature* **418**, 994-8.
- Hall, L. L., Byron, M., Butler, J., Becker, K. A., Nelson, A., Amit, M., Itskovitz-Eldor, J., Stein, J., Stein, G., Ware, C. and Lawrence, J. B. (2008). X-inactivation reveals epigenetic anomalies in most hESC but identifies sublines that initiate as expected. *J Cell Physiol* **216**, 445-52.
- Hamada, F. N., Park, P. J., Gordadze, P. R. and Kuroda, M. I. (2005). Global regulation of X chromosomal genes by the MSL complex in *Drosophila melanogaster*. *Genes Dev* **19**, 2289-94.
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557-80.
- Hansen, K. H., Bracken, A. P., Pasini, D., Dietrich, N., Gehani, S. S., Monrad, A., Rappsilber, J., Lerdrup, M. and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. *Nat Cell Biol* **10**, 1291-300.
- 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**, 613-26.
- Hartmuth, K., Vornlocher, H. P. and Luhrmann, R. (2004). Tobramycin affinity tag purification of spliceosomes. *Methods Mol Biol* **257**, 47-64.
- Hashizume, R., Fukuda, M., Maeda, I., Nishikawa, H., Oyake, D., Yabuki, Y., Ogata, H. and Ohta, T. (2001). The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. *J Biol Chem* **276**, 14537-40.
- Hastings, M. L. and Krainer, A. R. (2001). Pre-mRNA splicing in the new millennium. *Curr Opin Cell Biol* **13**, 302-9.
- Heard, E. (2004). Recent advances in X-chromosome inactivation. *Curr Opin Cell Biol* **16**, 247-55.

- Heard, E. (2005). Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. *Curr Opin Genet Dev* **15**, 482-9.
- Heard, E. and Bickmore, W. (2007). The ins and outs of gene regulation and chromosome territory organisation. *Curr Opin Cell Biol* **19**, 311-6.
- Heard, E. and Distèche, C. M. (2006). Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev* **20**, 1848-67.
- 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**, 727-38.
- Hekimoglu, B. and Ringrose, L. (2009). Non-coding RNAs in polycomb/trithorax regulation. *RNA Biol* **6**.
- 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**, 173-82.
- Hendrich, B. and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* **18**, 6538-47.
- Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. A. and Bird, A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev* **15**, 710-23.
- Hernandez-Munoz, I., Lund, A. H., van der Stoep, P., Boutsma, E., Muijers, I., Verhoeven, E., Nusinow, D. A., Panning, B., Marahrens, Y. and van Lohuizen, M. (2005). Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin E3 ligase. *Proc Natl Acad Sci U S A* **102**, 7635-40.
- Higgs, D. R., Vernimmen, D., De Gobbi, M., Anguita, E., Hughes, J., Buckle, V., Iborra, F., Garrick, D. and Wood, W. G. (2006). How transcriptional and epigenetic programmes are played out on an individual mammalian gene cluster during lineage commitment and differentiation. *Biochem Soc Symp*, 11-22.
- Hodgkin, J. (1980). More sex-determination mutants of *Caenorhabditis elegans*. *Genetics* **96**, 649-64.
- Hodgkin, J. A. and Brenner, S. (1977). Mutations causing transformation of sexual phenotype in the nematode *Caenorhabditis elegans*. *Genetics* **86**, 275-87.
- Hoffman, L. M., Hall, L., Batten, J. L., Young, H., Pardasani, D., Baetge, E. E., Lawrence, J. and Carpenter, M. K. (2005). X-inactivation status varies in human embryonic stem cell lines. *Stem Cells* **23**, 1468-78.
- 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* **136**, 139-46.
- Hollenbach, A. D., McPherson, C. J., Mientjes, E. J., Iyengar, R. and Grosveld, G. (2002). Daxx and histone deacetylase II associate with chromatin through an interaction with core histones and the chromatin-associated protein Dek. *J Cell Sci* **115**, 3319-30.
- 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. *Proc Natl Acad Sci U S A* **96**, 6829-34.
- 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**, 220-4.
- Hsu, D. R. and Meyer, B. J. (1994). The dpy-30 gene encodes an essential component of the *Caenorhabditis elegans* dosage compensation machinery. *Genetics* **137**, 999-1018.
- Hu, J. F., Balaguru, K. A., Ivaturi, R. D., Oruganti, H., Li, T., Nguyen, B. T., Vu, T. H. and Hoffman, A. R. (1999). Lack of reciprocal genomic imprinting of sense and antisense RNA of mouse insulin-like growth factor II receptor in the central nervous system. *Biochem Biophys Res Commun* **257**, 604-8.
- Huang, S., Litt, M. and Felsenfeld, G. (2005). Methylation of histone H4 by arginine methyltransferase PRMT1 is essential in vivo for many subsequent histone modifications. *Genes Dev* **19**, 1885-93.
- Huynh, K. D. and Lee, J. T. (2003). Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos. *Nature* **426**, 857-62.
- Ideraabdullah, F. Y., Vigneau, S. and Bartolomei, M. S. (2008). Genomic imprinting mechanisms in mammals. *Mutat Res* **647**, 77-85.
- Jacquet, S., Mereau, A., Bilodeau, P. S., Damier, L., Stoltzfus, C. M. and Branlant, C. (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**, 40464-75.
- Jaeger, J. A., Turner, D. H. and Zuker, M. (1989). Improved predictions of secondary structures for RNA. *Proc Natl Acad Sci U S A* **86**, 7706-10.
- Jaenisch, R. (1997). DNA methylation and imprinting: why bother? *Trends Genet* **13**, 323-9.
- Jaenisch, R. and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**, 245-54.
- Jamieson, R. V., Zhou, S. X., Wheatley, S. C., Koopman, P. and Tam, P. P. (1998). Sertoli cell differentiation and Y-chromosome activity: a developmental study of X-linked transgene activity in sex-reversed X/XS<sub>xra</sub> mouse embryos. *Dev Biol* **199**, 235-44.
- 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 Dev* **4**, 1560-72.



- Jegalian, K. and Page, D. C.** (1998). A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. *Nature* **394**, 776-80.
- Jenuwein, T. and Allis, C. D.** (2001). Translating the histone code. *Science* **293**, 1074-80.
- Jin, Y., Wang, Y., Johansen, J. and Johansen, K. M.** (2000). JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex. *J Cell Biol* **149**, 1005-10.
- Johnston, C. M., Lovell, F. L., Leongamornlert, D. A., Stranger, B. E., Dermitzakis, E. T. and Ross, M. T.** (2008). Large-scale population study of human cell lines indicates that dosage compensation is virtually complete. *PLoS Genet* **4**, e9.
- 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**, 809-17.
- Johnston, P. G. and Cattanach, B. M.** (1981). Controlling elements in the mouse. IV. Evidence of non-random X-inactivation. *Genet Res* **37**, 151-60.
- Jones, C. A., Ng, J., Peterson, A. J., Morgan, K., Simon, J. and Jones, R. S.** (1998). The Drosophila esc and E(z) proteins are direct partners in polycomb group-mediated repression. *Mol Cell Biol* **18**, 2825-34.
- Jones, M. H., Hamana, N., Nezu, J. and Shimane, M.** (2000). A novel family of bromodomain genes. *Genomics* **63**, 40-5.
- 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. *Mol Cell Biol* **28**, 5583-94.
- Kafasla, P., Morgner, N., Poyry, T. 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. *Mol Cell* **34**, 556-68.
- Kageyama, Y., Mengus, G., Gilfillan, G., Kennedy, H. G., Stuckenholtz, C., Kelley, R. L., Becker, P. B. and Kuroda, M. I.** (2001). Association and spreading of the Drosophila dosage compensation complex from a discrete roX1 chromatin entry site. *Embo J* **20**, 2236-45.
- Kalantry, S. and Magnuson, T.** (2006). The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation. *PLoS Genet* **2**, e66.
- Kalantry, S., Mills, K. C., Yee, D., Otte, A. P., Panning, B. and Magnuson, T.** (2006). The Polycomb group protein Eed protects the inactive X-chromosome from differentiation-induced reactivation. *Nat Cell Biol* **8**, 195-202.
- Kalscheuer, V. M., Mariman, E. C., Schepens, M. T., Rehder, H. and Ropers, H. H.** (1993). The insulin-like growth factor type-2 receptor gene is imprinted in the mouse but not in humans. *Nat Genet* **5**, 74-8.
- Kanellopoulou, C., Muljo, S. A., Dimitrov, S. D., Chen, X., Colin, C., Plath, K. and Livingston, D. M.** (2009). X chromosome inactivation in the absence of Dicer. *Proc Natl Acad Sci U S A* **106**, 1122-7.
- Kay, G. F., Barton, S. C., Surani, M. A. and Rastan, S.** (1994). Imprinting and X chromosome counting mechanisms determine Xist expression in early mouse development. *Cell* **77**, 639-50.
- Kay, G. F., Penny, G. D., Patel, D., Ashworth, A., Brockdorff, N. and Rastan, S.** (1993). Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. *Cell* **72**, 171-82.
- 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. *Dev Biol* **180**, 618-30.
- Kim, M., Li, D., Cui, Y., Mueller, K., Cheers, W. C. and DeJong, J.** (2006). Regulatory factor interactions and somatic silencing of the germ cell-specific ALF gene. *J Biol Chem* **281**, 34288-98.
- Kim, M. K. and Nikodem, V. M.** (1999). hnRNP U inhibits carboxy-terminal domain phosphorylation by TFIIF and represses RNA polymerase II elongation. *Mol Cell Biol* **19**, 6833-44.
- Kim, M. Y., Mauro, S., Gevry, N., Lis, J. T. and Kraus, W. L.** (2004). NAD<sup>+</sup>-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1. *Cell* **119**, 803-14.
- Kim, M. Y., Zhang, T. and Kraus, W. L.** (2005). Poly(ADP-ribosyl)ation by PARP-1: 'PAR-laying' NAD<sup>+</sup> into a nuclear signal. *Genes Dev* **19**, 1951-67.
- King, I. F., Emmons, R. B., Francis, N. J., Wild, B., Muller, J., Kingston, R. E. and Wu, C. T.** (2005). Analysis of a polycomb group protein defines regions that link repressive activity on nucleosomal templates to in vivo function. *Mol Cell Biol* **25**, 6578-91.
- King, I. F., Francis, N. J. and Kingston, R. E.** (2002). Native and recombinant polycomb group complexes establish a selective block to template accessibility to repress transcription in vitro. *Mol Cell Biol* **22**, 7919-28.
- Klymenko, T., Papp, B., Fischle, W., Kocher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M. and Muller, J.** (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev* **20**, 1110-22.
- 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 Biol* **2**, E171.
- Koop, R., Di Croce, L. and Beato, M.** (2003). Histone H1 enhances synergistic activation of the MMTV promoter in chromatin. *Embo J* **22**, 588-99.
- Kornberg, R. D.** (1974). Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868-71.



- Kornberg, R. D. and Thomas, J. O. (1974). Chromatin structure; oligomers of the histones. *Science* **184**, 865-8.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* **128**, 693-705.
- 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**, 955-67.
- Kuroda, M. I., Kernan, M. J., Kreber, R., Ganetzky, B. and Baker, B. S. (1991). The maleless protein associates with the X chromosome to regulate dosage compensation in *Drosophila*. *Cell* **66**, 935-47.
- Kustatscher, G., Hothorn, M., Pugieux, C., Scheffzek, K. and Ladurner, A. G. (2005). Splicing regulates NAD metabolite binding to histone macroH2A. *Nat Struct Mol Biol* **12**, 624-5.
- Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* **16**, 2893-905.
- Lachner, M. and Jenuwein, T. (2002). The many faces of histone lysine methylation. *Curr Opin Cell Biol* **14**, 286-98.
- Lagarou, A., Mohd-Sarip, A., Moshkin, Y. M., Chalkley, G. E., Bezstarosti, K., Demmers, J. A. and Verrijzer, C. P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev* **22**, 2799-810.
- Lahn, B. T. and Page, D. C. (1997). Functional coherence of the human Y chromosome. *Science* **278**, 675-80.
- Lavrov, S. A. and Kibanov, M. V. (2007). Noncoding RNAs and chromatin structure. *Biochemistry (Mosc)* **72**, 1422-38.
- Lee, J. T., Davidow, L. S. and Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet* **21**, 400-4.
- Lee, J. T. and Jaenisch, R. (1997). Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. *Nature* **386**, 275-9.
- Lee, J. T. and Lu, N. (1999). Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell* **99**, 47-57.
- Lee, J. T., Lu, N. and Han, Y. (1999). Genetic analysis of the mouse X inactivation center defines an 80-kb multifunction domain. *Proc Natl Acad Sci U S A* **96**, 3836-41.
- 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**, 83-94.
- Lehnertz, B., Ueda, Y., Derijck, A. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T. and Peters, A. H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol* **13**, 1192-200.
- Levine, S. S., King, I. F. and Kingston, R. E. (2004). Division of labor in polycomb group repression. *Trends Biochem Sci* **29**, 478-85.
- Li, E., Bestor, T. H. and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-26.
- Li, H., Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D. and Patel, D. J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91-5.
- Li, H. Y., Savage, T., Obermoeller, R. D., Kazianis, S. and Walter, R. B. (2002). Parental 5-methylcytosine methylation patterns are stable upon inter-species hybridization of *Xiphophorus* (Teleostei: Poeciliidae) fish. *Comp Biochem Physiol B Biochem Mol Biol* **133**, 581-95.
- 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**, 265-77.
- Lin, S. P., Youngson, N., Takada, S., Seitz, H., Reik, W., Paulsen, M., Cavaille, J. and Ferguson-Smith, A. C. (2003). Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nat Genet* **35**, 97-102.
- Lingenfelter, P. A., Adler, D. A., Poslinski, D., Thomas, S., Elliott, R. W., Chapman, V. M. and Disteche, C. M. (1998). Escape from X inactivation of Smcx is preceded by silencing during mouse development. *Nat Genet* **18**, 212-3.
- Losada, A., Yokochi, T. and Hirano, T. (2005). Functional contribution of Pds5 to cohesin-mediated cohesion in human cells and *Xenopus* egg extracts. *J Cell Sci* **118**, 2133-41.
- Lou, H., Helfman, D. M., Gagel, R. F. and Berget, S. M. (1999). Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3'-terminal exon. *Mol Cell Biol* **19**, 78-85.
- Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251-60.
- Luikenhuis, S., Wutz, A. and Jaenisch, R. (2001). Antisense transcription through the Xist locus mediates Tsix function in embryonic stem cells. *Mol Cell Biol* **21**, 8512-20.
- Lukong, K. E. and Richard, S. (2003). Sam68, the KH domain-containing superSTAR. *Biochim Biophys Acta* **1653**, 73-86.
- Lyle, R., Watanabe, D., te Vruchte, D., Lerchner, W., Smrzka, O. W., Wutz, A., Schageman, J., Hahner, L., Davies, C. and Barlow, D. P. (2000). The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. *Nat Genet* **25**, 19-21.
- Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190**, 372-3.
- Lyon, M. F. (1962). Sex chromatin and gene action in the mammalian X-chromosome. *Am J Hum Genet* **14**, 135-48.

- Lyon, M. F. (1998). X-chromosome inactivation: a repeat hypothesis. *Cytogenet Cell Genet* **80**, 133-7.
- Lyon, M. F. and Rastan, S. (1984). Parental source of chromosome imprinting and its relevance for X chromosome inactivation. *Differentiation* **26**, 63-7.
- Mahy, G. (2002). Commentary: the evolution of treatment in catatonia. *J Am Acad Psychiatry Law* **30**, 377-9.
- Mahy, N. L., Perry, P. E. and Bickmore, W. A. (2002). Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J Cell Biol* **159**, 753-63.
- Maier, H., Ostraat, R., Parenti, S., Fitzsimmons, D., Abraham, L. J., Garvie, C. W. and Hagman, J. (2003). Requirements for selective recruitment of Ets proteins and activation of mb-1/Ig-alpha gene transcription by Pax-5 (BSAP). *Nucleic Acids Res* **31**, 5483-9.
- 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. *Curr Biol* **12**, 1016-20.
- 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* **303**, 666-9.
- Mancini-Dinardo, D., Steele, S. J., Levorse, J. M., Ingram, R. S. and Tilghman, S. M. (2006). Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes Dev* **20**, 1268-82.
- Marahrens, Y. (1999). X-inactivation by chromosomal pairing events. *Genes Dev* **13**, 2624-32.
- Marahrens, Y., Panning, B., Dausman, J., Strauss, W. and Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev* **11**, 156-66.
- Marchand, V., Mougin, A., Méreau, A., Branlant, C., (2005). Study of RNA-Protein interactions and RNA structure in ribonucleoprotein particles (RNP), *Hartmann: Handbook of RNA Biochemistry*
- Margot, J. B., Ehrenhofer-Murray, A. E. and Leonhardt, H. (2003). Interactions within the mammalian DNA methyltransferase family. *BMC Mol Biol* **4**, 7.
- Margueron, R., Trojer, P. and Reinberg, D. (2005). The key to development: interpreting the histone code? *Curr Opin Genet Dev* **15**, 163-76.
- Marin, I. (2008). Dosage Compensation Mechanisms: Evolution. *Encyclopedia of life sciences*
- Martens, J. H., Verlaan, M., Kalkhoven, E., Dorsman, J. C. and Zantema, A. (2002). Scaffold/matrix attachment region elements interact with a p300-scaffold attachment factor A complex and are bound by acetylated nucleosomes. *Mol Cell Biol* **22**, 2598-606.
- Martin, C. and Zhang, Y. (2005). The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* **6**, 838-49.
- Matsui, J., Goto, Y. and Takagi, N. (2001). Control of Xist expression for imprinted and random X chromosome inactivation in mice. *Hum Mol Genet* **10**, 1393-401.
- Mattick, J. S. (2001). Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep* **2**, 986-91.
- Mattick, J. S. (2003). Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* **25**, 930-9.
- Mattick, J. S. and Makunin, I. V. (2006). Non-coding RNA. *Hum Mol Genet* **15 Spec No 1**, R17-29.
- Matzke, M. A. and Birchler, J. A. (2005). RNAi-mediated pathways in the nucleus. *Nat Rev Genet* **6**, 24-35.
- Mayer, W., Niveleau, A., Walter, J., Fundele, R. and Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature* **403**, 501-2.
- McCarthy, E. E., Celebi, J. T., Baer, R. and Ludwig, T. (2003). Loss of Bard1, the heterodimeric partner of the Brcal tumor suppressor, results in early embryonic lethality and chromosomal instability. *Mol Cell Biol* **23**, 5056-63.
- McDonald, L. E., Paterson, C. A. and Kay, G. F. (1998). Bisulfite genomic sequencing-derived methylation profile of the xist gene throughout early mouse development. *Genomics* **54**, 379-86.
- McDonel, P., Jans, J., Peterson, B. K. and Meyer, B. J. (2006). Clustered DNA motifs mark X chromosomes for repression by a dosage compensation complex. *Nature* **444**, 614-8.
- McDowell, T. L., Gibbons, R. J., Sutherland, H., O'Rourke, D. M., Bickmore, W. A., Pombo, A., Turley, H., Gatter, K., Picketts, D. J., Buckle, V. J., Chapman, L., Rhodes, D. and Higgs, D. R. (1999). Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. *Proc Natl Acad Sci U S A* **96**, 13983-8.
- McGrath, J. and Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37**, 179-83.
- McNeil, J. A., Smith, K. P., Hall, L. L. and Lawrence, J. B. (2006). Word frequency analysis reveals enrichment of dinucleotide repeats on the human X chromosome and [GATA]<sub>n</sub> in the X escape region. *Genome Res* **16**, 477-84.
- Meister, G. and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343-9.
- Meller, V. H. and Rattner, B. P. (2002). The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *Embo J* **21**, 1084-91.
- Meller, V. H., Wu, K. H., Roman, G., Kuroda, M. I. and Davis, R. L. (1997). roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system. *Cell* **88**, 445-57.
- Mellor, J. (2005). The dynamics of chromatin remodeling at promoters. *Mol Cell* **19**, 147-57.

- Meneely, P. M. and Wood, W. B.** (1984). An autosomal gene that affects X chromosome expression and sex determination in *Caenorhabditis elegans*. *Genetics* **106**, 29-44.
- Meneely, P. M. and Wood, W. B.** (1987). Genetic analysis of X-chromosome dosage compensation in *Caenorhabditis elegans*. *Genetics* **117**, 25-41.
- Mermoud, J. E., Costanzi, C., Pehrson, J. R. and Brockdorff, N.** (1999). Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. *J Cell Biol* **147**, 1399-408.
- 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. *Curr Biol* **12**, 247-51.
- Mermoud, J. E., Tassin, A. M., Pehrson, J. R. and Brockdorff, N.** (2001). Centrosomal association of histone macroH2A1.2 in embryonic stem cells and somatic cells. *Exp Cell Res* **268**, 245-51.
- Meyer, B. J.** (2005). X-Chromosome dosage compensation. *WormBook*, 1-14.
- Meyer, B. J. and Casson, L. P.** (1986). *Caenorhabditis elegans* compensates for the difference in X chromosome dosage between the sexes by regulating transcript levels. *Cell* **47**, 871-81.
- Mietton, F., Sengupta, A. K., Molla, A., Picchi, G., Barral, S., Heliot, L., Grange, T., Wutz, A. and Dimitrov, S.** (2009). Weak but uniform enrichment of the histone variant macroH2A1 along the inactive X chromosome. *Mol Cell Biol* **29**, 150-6.
- Mihaly, J., Mishra, R. K. and Karch, F.** (1998). A conserved sequence motif in Polycomb-response elements. *Mol Cell* **1**, 1065-6.
- Miller, D., Summers, J. and Silber, S.** (2004). Environmental versus genetic sex determination: a possible factor in dinosaur extinction? *Fertil Steril* **81**, 954-64.
- 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 Dev* **11**, 1023-36.
- Mise, N., Goto, Y., Nakajima, N. and Takagi, N.** (1999). Molecular cloning of antisense transcripts of the mouse Xist gene. *Biochem Biophys Res Commun* **258**, 537-41.
- 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. *Mol Cell* **11**, 757-71.
- Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. and Yamanaka, S.** (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-42.
- Mitsuya, K., Meguro, M., Lee, M. P., Katoh, M., Schulz, T. C., Kugoh, H., Yoshida, M. A., Niikawa, N., Feinberg, A. P. and Oshimura, M.** (1999). LIT1, an imprinted antisense RNA in the human KvLQT1 locus identified by screening for differentially expressed transcripts using monochromosomal hybrids. *Hum Mol Genet* **8**, 1209-17.
- Mlynarczyk-Evans, S., Royce-Tolland, M., Alexander, M. K., Andersen, A. A., Kalantry, S., Gribnau, J. and Panning, B.** (2006). X chromosomes alternate between two states prior to random X-inactivation. *PLoS Biol* **4**, e159.
- Moen, P. T., Jr., Johnson, C. V., Byron, M., Shopland, L. S., de la Serna, I. L., Imbalzano, A. N. and Lawrence, J. B.** (2004). Repositioning of muscle-specific genes relative to the periphery of SC-35 domains during skeletal myogenesis. *Mol Biol Cell* **15**, 197-206.
- Mohammad, F., Pandey, R. R., Nagano, T., Chakalova, L., Mondal, T., Fraser, P. and Kanduri, C.** (2008). Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. *Mol Cell Biol* **28**, 3713-28.
- Mohd-Sarip, A., van der Knaap, J. A., Wyman, C., Kanaar, R., Schedl, P. and Verrijzer, C. P.** (2006). Architecture of a polycomb nucleoprotein complex. *Mol Cell* **24**, 91-100.
- Monk, M. and McLaren, A.** (1981). X-chromosome activity in foetal germ cells of the mouse. *J Embryol Exp Morphol* **63**, 75-84.
- Monkhorst, K., de Hoon, B., Jonkers, I., Mulugeta Achame, E., Monkhorst, W., Hoogerbrugge, J., Rentmeester, E., Westerhoff, H. V., Grosveld, F., Grootegoed, J. A. and Gribnau, J.** (2009). The probability to initiate X chromosome inactivation is determined by the X to autosomal ratio and X chromosome specific allelic properties. *PLoS One* **4**, e5616.
- 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**, 410-21.
- Moore, M. J., Query, C. C. and Sharp, P. A.** (1993). Splicing of precursors to messenger RNAs by the spliceosome. In *The RNA World*, pp. 303-357.
- Morey, C., Arnaud, D., Avner, P. and Clerc, P.** (2001). Tsix-mediated repression of Xist accumulation is not sufficient for normal random X inactivation. *Hum Mol Genet* **10**, 1403-11.
- Morey, C. and Avner, P.** (2004). Employment opportunities for non-coding RNAs. *FEBS Lett* **567**, 27-34.
- Morey, C., Navarro, P., Debrand, E., Avner, P., Rougeulle, C. and Clerc, P.** (2004). The region 3' to Xist mediates X chromosome counting and H3 Lys-4 dimethylation within the Xist gene. *Embo J* **23**, 594-604.
- Morison, I. M. and Reeve, A. E.** (1998). A catalogue of imprinted genes and parent-of-origin effects in humans and animals. *Hum Mol Genet* **7**, 1599-609.

- Motamedi, M. R., Verdel, A., Colmenares, S. U., Gerber, S. A., Gygi, S. P. and Moazed, D. (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**, 789-802.
- Mougin, A., Gregoire, A., Banroques, J., Segault, V., Fournier, R., Brule, F., Chevrier-Miller, M. and Branlant, C. (1996). Secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency. *Rna* **2**, 1079-93.
- Muller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B., Miller, E. L., O'Connor, M. B., Kingston, R. E. and Simon, J. A. (2002). Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**, 197-208.
- Nagy, P. L., Griesenbeck, J., Kornberg, R. D. and Cleary, M. L. (2002). A trithorax-group complex purified from *Saccharomyces cerevisiae* is required for methylation of histone H3. *Proc Natl Acad Sci U S A* **99**, 90-4.
- Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. and Grewal, S. I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110-3.
- Nan, X., Campoy, F. J. and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* **88**, 471-81.
- Nasmyth, K. and Haering, C. H. (2005). The structure and function of SMC and kleisin complexes. *Annu Rev Biochem* **74**, 595-648.
- Navarro, P. and Avner, P. (2009). When X-inactivation meets pluripotency: an intimate rendezvous. *FEBS Lett* **583**, 1721-7.
- 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* **321**, 1693-5.
- 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 Dev* **20**, 2787-92.
- 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 Dev* **19**, 1474-84.
- Negishi, M., Saraya, A., Miyagi, S., Nagao, K., Inagaki, Y., Nishikawa, M., Tajima, S., Koseki, H., Tsuda, H., Takasaki, Y., Nakauchi, H. and Iwama, A. (2007). Bmi1 cooperates with Dnmt1-associated protein 1 in gene silencing. *Biochem Biophys Res Commun* **353**, 992-8.
- Nesterova, T. B., Barton, S. C., Surani, M. A. and Brockdorff, N. (2001). Loss of Xist imprinting in diploid parthenogenetic preimplantation embryos. *Dev Biol* **235**, 343-50.
- Nesterova, T. B., Mermoud, J. E., Hilton, K., Pehrson, J., Surani, M. A., McLaren, A. and Brockdorff, N. (2002). Xist expression and macroH2A1.2 localisation in mouse primordial and pluripotent embryonic germ cells. *Differentiation* **69**, 216-25.
- Nesterova, T. B., Popova, B. C., Cobb, B. S., Norton, S., Senner, C. E., Tang, Y. A., Spruce, T., Rodriguez, T. A., Sado, T., Merkschlager, M. and Brockdorff, N. (2008). Dicer regulates Xist promoter methylation in ES cells indirectly through transcriptional control of Dnmt3a. *Epigenetics Chromatin* **1**, 2.
- Ng, H. H. and Bird, A. (1999). DNA methylation and chromatin modification. *Curr Opin Genet Dev* **9**, 158-63.
- Ng, J., Hart, C. M., Morgan, K. and Simon, J. A. (2000). A *Drosophila* ESC-E(Z) protein complex is distinct from other polycomb group complexes and contains covalently modified ESC. *Mol Cell Biol* **20**, 3069-78.
- Nicodemi, M. and Prisco, A. (2007). Symmetry-breaking model for X-chromosome inactivation. *Phys Rev Lett* **98**, 108104.
- Nogami, M., Kohda, A., Taguchi, H., Nakao, M., Ikemura, T. and Okumura, K. (2000). Relative locations of the centromere and imprinted SNRPN gene within chromosome 15 territories during the cell cycle in HL60 cells. *J Cell Sci* **113** ( Pt 12), 2157-65.
- Nolen, L. D., Gao, S., Han, Z., Mann, M. R., Gie Chung, Y., Otte, A. P., Bartolomei, M. S. and Latham, K. E. (2005). X chromosome reactivation and regulation in cloned embryos. *Dev Biol* **279**, 525-40.
- Nusbaum, C. and Meyer, B. J. (1989). The *Caenorhabditis elegans* gene *sdc-2* controls sex determination and dosage compensation in XX animals. *Genetics* **122**, 579-93.
- Nusinow, D. A., Sharp, J. A., Morris, A., Salas, S., Plath, K. and Panning, B. (2007). The histone domain of macroH2A1 contains several dispersed elements that are each sufficient to direct enrichment on the inactive X chromosome. *J Mol Biol* **371**, 11-8.
- Nye, A., Rajendran, R., Belmont, A., (2005). Chromosomes and Chromatin, *Encyclopedia of life sciences*
- Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D. L. and Allain, F. H. (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* **309**, 2054-7.
- Ogawa, Y. and Lee, J. T. (2003). Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. *Mol Cell* **11**, 731-43.
- Ogawa, Y., Ono, T., Wakata, Y., Okawa, K., Tagami, H. and Shibahara, K. I. (2005). Histone variant macroH2A1.2 is mono-ubiquitinated at its histone domain. *Biochem Biophys Res Commun* **336**, 204-9.
- Ogawa, Y., Sun, B. K. and Lee, J. T. (2008). Intersection of the RNA interference and X-inactivation pathways. *Science* **320**, 1336-41.

- O'Gorman, W., Kwek, K. Y., Thomas, B. and Akoulitchev, A. (2006). Non-coding RNA in transcription initiation. *Biochem Soc Symp*, 131-40.
- Ohhata, T., Hoki, Y., Sasaki, H. and Sado, T. (2008). Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification. *Development* **135**, 227-35.
- 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.
- Ohno, S. and Hauschka, T. S. (1960). Allocyclus of the X-chromosome in tumors and normal tissues. *Cancer Res* **20**, 541-5.
- Ohno, S., (1967) Sex chromosomes and sexlinked genes *Monographs on endocrinology 1*
- Ohno, S., Kaplan, W. D. and Kinoshita, R. (1959). Formation of the sex chromatin by a single X-chromosome in liver cells of *Rattus norvegicus*. *Exp Cell Res* **18**, 415-8.
- Okamoto, I., Arnaud, D., Le Baccon, P., Otte, A. P., Disteché, C. M., Avner, P. and Heard, E. (2005). Evidence for de novo imprinted X-chromosome inactivation independent of meiotic inactivation in mice. *Nature* **438**, 369-73.
- 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* **303**, 644-9.
- Okamoto, I., Tan, S. and Takagi, N. (2000). X-chromosome inactivation in XX androgenetic mouse embryos surviving implantation. *Development* **127**, 4137-45.
- 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**, 247-57.
- Olins, A. L. and Olins, D. E. (1974). Spheroid chromatin units (v bodies). *Science* **183**, 330-2.
- O'Neill, L. P., Keohane, A. M., Lavender, J. S., McCabe, V., Heard, E., Avner, P., Brockdorff, N. and Turner, B. M. (1999). A developmental switch in H4 acetylation upstream of Xist plays a role in X chromosome inactivation. *Embo J* **18**, 2897-907.
- O'Neill, M. J. (2005). The influence of non-coding RNAs on allele-specific gene expression in mammals. *Hum Mol Genet* **14 Spec No 1**, R113-20.
- Ooi, S. K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S. P., Allis, C. D., Cheng, X. and Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714-7.
- Osborne, A. R., Zhang, H., Fejer, G., Palubin, K. M., Niesen, M. I. and Blanck, G. (2004). Oct-1 maintains an intermediate, stable state of HLA-DRA promoter repression in Rb-defective cells: an Oct-1-containing repressosome that prevents NF-Y binding to the HLA-DRA promoter. *J Biol Chem* **279**, 28911-9.
- Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W. and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. *Curr Biol* **10**, 475-8.
- Oudejans, C. B., Westerman, B., Wouters, D., Gooyer, S., Leegwater, P. A., van Wijk, I. J. and Sleutels, F. (2001). Allelic IGF2R repression does not correlate with expression of antisense RNA in human extraembryonic tissues. *Genomics* **73**, 331-7.
- Pageau, G. J., Hall, L. L. and Lawrence, J. B. (2007). BRCA1 does not paint the inactive X to localize XIST RNA but may contribute to broad changes in cancer that impact XIST and Xi heterochromatin. *J Cell Biochem* **100**, 835-50.
- Pageau, G. J. and Lawrence, J. B. (2006). BRCA1 foci in normal S-phase nuclei are linked to interphase centromeres and replication of pericentric heterochromatin. *J Cell Biol* **175**, 693-701.
- Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D. and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* **32**, 232-46.
- Panning, B., Dausman, J. and Jaenisch, R. (1997). X chromosome inactivation is mediated by Xist RNA stabilization. *Cell* **90**, 907-16.
- Parada, L. and Misteli, T. (2002). Chromosome positioning in the interphase nucleus. *Trends Cell Biol* **12**, 425-32.
- Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., Cobb, B. S., Yokomori, K., Dillon, N., Aragon, L., Fisher, A. G. and Merkenschlager, M. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422-33.
- Park, W. W. (1957). The occurrence of sex chromatin in early human and macaque embryos. *J Anat* **91**, 369-73.
- Partridge, J. F., Borgstrom, B. and Allshire, R. C. (2000). Distinct protein interaction domains and protein spreading in a complex centromere. *Genes Dev* **14**, 783-91.
- Patrat, C., Okamoto, I., Diabangouaya, P., Vialon, V., Le Baccon, P., Chow, J. and Heard, E. (2009). Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice. *Proc Natl Acad Sci U S A* **106**, 5198-203.
- Patrinos, G. P., de Krom, M., de Boer, E., Langeveld, A., Imam, A. M., Strouboulis, J., de Laat, W. and Grosveld, F. G. (2004). Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev* **18**, 1495-509.
- Payer, B. and Lee, J. T. (2008). X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet* **42**, 733-72.

- Pehrson, J. R. and Fried, V. A.** (1992). MacroH2A, a core histone containing a large nonhistone region. *Science* **257**, 1398-400.
- Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. and Brockdorff, N.** (1996). Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131-7.
- Perche, P. Y., Vourc'h, C., Konecny, L., Souchier, C., Robert-Nicoud, M., Dimitrov, S. and Khochbin, S.** (2000). Higher concentrations of histone macroH2A in the Barr body are correlated with higher nucleosome density. *Curr Biol* **10**, 1531-4.
- 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**, 11881-90.
- 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. *Nat Genet* **30**, 77-80.
- 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**, 1021-7.
- Platero, J. S., Hartnett, T. and Eissenberg, J. C.** (1995). Functional analysis of the chromo domain of HP1. *Embo J* **14**, 3977-86.
- Plath, K., Fang, J., Mlynarczyk-Evans, S. K., Cao, R., Worringer, K. A., Wang, H., de la Cruz, C. C., Otte, A. P., Panning, B. and Zhang, Y.** (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131-5.
- Plath, K., Talbot, D., Hamer, K. M., Otte, A. P., Yang, T. P., Jaenisch, R. and Panning, B.** (2004). Developmentally regulated alterations in Polycomb repressive complex 1 proteins on the inactive X chromosome. *J Cell Biol* **167**, 1025-35.
- Plenefisch, J. D., DeLong, L. and Meyer, B. J.** (1989). Genes that implement the hermaphrodite mode of dosage compensation in *Caenorhabditis elegans*. *Genetics* **121**, 57-76.
- Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolzheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D. K. and Young, R. A.** (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517-27.
- Polesskaya, A., Cuvellier, S., Naguibneva, I., Duquet, A., Moss, E. G. and Harel-Bellan, A.** (2007). Lin-28 binds IGF-2 mRNA and participates in skeletal myogenesis by increasing translation efficiency. *Genes Dev* **21**, 1125-38.
- Popova, B. C., Tada, T., Takagi, N., Brockdorff, N. and Nesterova, T. B.** (2006). Attenuated spread of X-inactivation in an X;autosome translocation. *Proc Natl Acad Sci U S A* **103**, 7706-11.
- Priest, J. H., Heady, J. E. and Priest, R. E.** (1967). Delayed onset of replication of human X chromosomes. *J Cell Biol* **35**, 483-7.
- Prissette, M., El-Maarri, O., Arnaud, D., Walter, J. and Avner, P.** (2001). Methylation profiles of DXPas34 during the onset of X-inactivation. *Hum Mol Genet* **10**, 31-8.
- Prokhortchouk, A., Hendrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A. and Prokhortchouk, E.** (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev* **15**, 1613-8.
- Rasmussen, T. P., Mastrangelo, M. A., Eden, A., Pehrson, J. R. and Jaenisch, R.** (2000). Dynamic relocalization of histone MacroH2A1 from centrosomes to inactive X chromosomes during X inactivation. *J Cell Biol* **150**, 1189-98.
- Rasmussen, T. P., Wutz, A. P., Pehrson, J. R. and Jaenisch, R. R.** (2001). Expression of Xist RNA is sufficient to initiate macrochromatin body formation. *Chromosoma* **110**, 411-20.
- Rastan, S.** (1983). Non-random X-chromosome inactivation in mouse X-autosome translocation embryos--location of the inactivation centre. *J Embryol Exp Morphol* **78**, 1-22.
- Rastan, S. and Robertson, E. J.** (1985). X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. *J Embryol Exp Morphol* **90**, 379-88.
- Rattner, B. P. and Meller, V. H.** (2004). Drosophila male-specific lethal 2 protein controls sex-specific expression of the roX genes. *Genetics* **166**, 1825-32.
- Redrup, L., Branco, M. R., Perdeaux, E. R., Krueger, C., Lewis, A., Santos, F., Nagano, T., Cobb, B. S., Fraser, P. and Reik, W.** (2009). The long noncoding RNA Kcnqlot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. *Development* **136**, 525-30.
- Reik, W. and Lewis, A.** (2005). Co-evolution of X-chromosome inactivation and imprinting in mammals. *Nat Rev Genet* **6**, 403-10.
- Reik, W. and Walter, J.** (2001). Genomic imprinting: parental influence on the genome. *Nat Rev Genet* **2**, 21-32.
- Reisman, D., Glaros, S. and Thompson, E. A.** (2009). The SWI/SNF complex and cancer. *Oncogene* **28**, 1653-68.
- Rhee, I., Jair, K. W., Yen, R. W., Lengauer, C., Herman, J. G., Kinzler, K. W., Vogelstein, B., Baylin, S. B. and Schuebel, K. E.** (2000). CpG methylation is maintained in human cancer cells lacking DNMT1. *Nature* **404**, 1003-7.
- Richardson, A. L., Wang, Z. C., De Nicolo, A., Lu, X., Brown, M., Miron, A., Liao, X., Iglehart, J. D., Livingston, D. M. and Ganesan, S.** (2006). X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell* **9**, 121-32.



- Riclet, R., Chendeb, M., Vonesch, J. L., Koczan, D., Thiesen, H. J., Losson, R. and Cammas, F. (2009). Disruption of the interaction between transcriptional intermediary factor 1{beta} and heterochromatin protein 1 leads to a switch from DNA hyper- to hypomethylation and H3K9 to H3K27 trimethylation on the MEST promoter correlating with gene reactivation. *Mol Biol Cell* **20**, 296-305.
- Ringrose, L., Ehret, H. and Paro, R. (2004). Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of polycomb complexes. *Mol Cell* **16**, 641-53.
- Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Bruggmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E. and Chang, H. Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-23.
- Rogner, U. C., Spyropoulos, D. D., Le Novere, N., Changeux, J. P. and Avner, P. (2000). Control of neurulation by the nucleosome assembly protein-1-like 2. *Nat Genet* **25**, 431-5.
- Roshon, M. J. and Ruley, H. E. (2005). Hypomorphic mutation in hnRNP U results in post-implantation lethality. *Transgenic Res* **14**, 179-92.
- Ross, M. T., Grafham, D. V., Coffey, A. J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G. R., Burrows, C., Bird, C. P., Frankish, A., Lovell, F. L., Howe, K. L., Ashurst, J. L., Fulton, R. S., Sudbrak, R., Wen, G., Jones, M. C., Hurler, M. E., Andrews, T. D., Scott, C. E., Searle, S., Ramser, J., Whittaker, A., Deadman, R., Carter, N. P., Hunt, S. E., Chen, R., Cree, A., Gunaratne, P., Havlak, P., Hodgson, A., Metzker, M. L., Richards, S., Scott, G., Steffen, D., Sodergren, E., Wheeler, D. A., Worley, K. C., Ainscough, R., Ambrose, K. D., Ansari-Lari, M. A., Aradhya, S., Ashwell, R. I., Babbage, A. K., Bagguley, C. L., Ballabio, A., Banerjee, R., Barker, G. E., Barlow, K. F., Barrett, I. P., Bates, K. N., Beare, D. M., Beasley, H., Beasley, O., Beck, A., Bethel, G., Blechschmidt, K., Brady, N., Bray-Allen, S., Bridgeman, A. M., Brown, A. J., Brown, M. J., Bonnini, D., Bruford, E. A., Buhay, C., Burch, P., Burford, D., Burgess, J., Burrill, W., Burton, J., Bye, J. M., Carder, C., Carrel, L., Chako, J., Chapman, J. C., Chavez, D., Chen, E., Chen, G., Chen, Y., Chen, Z., Chinault, C., Ciccodicola, A., Clark, S. Y., Clarke, G., Clee, C. M., Clegg, S., Clerc-Blankenburg, K., Clifford, K., Cobley, V., Cole, C. G., Conquer, J. S., Corby, N., Connor, R. E., David, R., Davies, J., Davis, C., Davis, J., Delgado, O., Deshazo, D., Dhami, P., Ding, Y., Dinh, H., Dodsworth, S., Draper, H., Dugan-Rocha, S., Dunham, A., Dunn, M., Durbin, K. J., Dutta, I., Eades, T., Ellwood, M., Emery-Cohen, A., Errington, H., Evans, K. L., Faulkner, L., Francis, F., Frankland, J., Fraser, A. E., Galgoczy, P., Gilbert, J. Gill, R., Glockner, G., Gregory, S. G., Gribble, S., Griffiths, C., Grocock, R., Gu, Y., Gwilliam, R., Hamilton, C., Hart, E. A., Hawes, A., Heath, P. D., Heitmann, K., Hennig, S., Hernandez, J., Hinemann, B., Ho, S., Hoffs, M., Howden, P. J., Huckle, E. J., Hume, J., Hunt, P. J., Hunt, A. R., Isherwood, J., Jacob, L., Johnson, D., Jones, S., de Jong, P. J., Joseph, S. S., Keenan, S., Kelly, S., Kershaw, J. K., Khan, Z., Kioschis, P., Klages, S., Knights, A. J., Kosiura, A., Kovar-Smith, C., Laird, G. K., Langford, C., Lawlor, S., Leversha, M., Lewis, L., Liu, W., Lloyd, C., Lloyd, D. M., Louseged, H., Loveland, J. E., Lovell, J. D., Lozado, R., Lu, J., Lyne, R., Ma, J., Maheshwari, M., Matthews, L. H., McDowall, J., McLaren, S., McMurray, A., Meidl, P., Meitinger, T., Milne, S., Miner, G., Mistry, S. L., Morgan, M., Morris, S., Muller, I., Mullikin, J. C., Nguyen, N., Nordsiek, G., Nyakatura, G., O'Dell, C. N., Okwuonu, G., Palmer, S., Pandian, R., Parker, D., Parrish, J., Pasternak, S., Patel, D., Pearce, A. V., Pearson, D. M., Pelan, S. E., Perez, L., Porter, K. M., Ramsey, Y., Reichwald, K., Rhodes, S., Ridler, K. A., Schlessinger, D., Schueler, M. G., Sehra, H. K., Shaw-Smith, C., Shen, H., Sheridan, E. M., Shownkeen, R., Skuce, C. D., Smith, M. L., Sotheran, E. C., Steingruber, H. E., Steward, C. A., Storey, R., Swann, R. M., Swarbreck, D., Tabor, P. E., Taudien, S., Taylor, T., Teague, B., Thomas, K., Thorpe, A., Timms, K., Tracey, A., Trevanion, S., Tromans, A. C., d'Urso, M., Verduzco, D., Villasana, D., Waldron, L., Wall, M., Wang, Q., Warren, J., Warry, G. L., Wei, X., West, A., Whitehead, S. L., Whiteley, M. N., Wilkinson, J. E., Willey, D. L., Williams, G., Williams, L., Williamson, A., Williamson, H., Wilming, L., Woodmansey, R. L., Wray, P. W., Yen, J., Zhang, J., Zhou, J., Zoghbi, H., Zorilla, S., Buck, D., Reinhardt, R., Poustka, A., Rosenthal, A., Lehrach, H., Meindl, A., Minx, P. J., Hillier, L. W., Willard, H. F., Wilson, R. K., Waterston, R. H., Rice, C. M., Vaudin, M., Coulson, A., Nelson, D. L., Weinstock, G., Sulston, J. E., Durbin, R., Hubbard, T., Gibbs, R. A., Beck, S., Rogers, J. and Bentley, D. R. (2005). The DNA sequence of the human X chromosome. *Nature* **434**, 325-37.
- Rougeulle, C. and Avner, P. (1996). Cloning and characterization of a murine brain specific gene Bpx and its human homologue lying within the Xic candidate region. *Hum Mol Genet* **5**, 41-9.
- Rougeulle, C. and Avner, P. (2003). Controlling X-inactivation in mammals: what does the centre hold? *Semin Cell Dev Biol* **14**, 331-40.
- Rougeulle, C., Chaumeil, J., Sarma, K., Allis, C. D., Reinberg, D., Avner, P. and Heard, E. (2004). Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome. *Mol Cell Biol* **24**, 5475-84.
- Rougier, N., Bourc'his, D., Gomes, D. M., Niveleau, A., Plachot, M., Paldi, A. and Viegas-Pequignot, E. (1998). Chromosome methylation patterns during mammalian preimplantation development. *Genes Dev* **12**, 2108-13.
- Rouleau, M., McDonald, D., Gagne, P., Ouellet, M. E., Droit, A., Hunter, J. M., Dutertre, S., Prigent, C., Hendzel, M. J. and Poirier, G. G. (2007). PARP-3 associates with polycomb group bodies and with components of the DNA damage repair machinery. *J Cell Biochem* **100**, 385-401.

- Ruffner, H., Joazeiro, C. A., Hemmati, D., Hunter, T. and Verma, I. M.** (2001). Cancer-predisposing mutations within the RING domain of BRCA1: loss of ubiquitin protein ligase activity and protection from radiation hypersensitivity. *Proc Natl Acad Sci U S A* **98**, 5134-9.
- 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. *Dev Biol* **225**, 294-303.
- Sado, T., Hoki, Y. and Sasaki, H.** (2006). Tsix defective in splicing is competent to establish Xist silencing. *Development* **133**, 4925-31.
- Sado, T., Okano, M., Li, E. and Sasaki, H.** (2004). De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation. *Development* **131**, 975-82.
- Sado, T., Wang, Z., Sasaki, H. and Li, E.** (2001). Regulation of imprinted X-chromosome inactivation in mice by Tsix. *Development* **128**, 1275-86.
- Sadoni, N. and Zink, D.** (2004). Nascent RNA synthesis in the context of chromatin architecture. *Chromosome Res* **12**, 439-51.
- Sambrook, J., Fritsch, E. F. and Maniatis, T.** (1989). Molecular cloning. A laboratory manual. New York: Cold Spring Harbor Laboratory Press.
- Sanchez, C., Sanchez, I., Demmers, J. A., Rodriguez, P., Strouboulis, J. and Vidal, M.** (2007). Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbx110/JhdmlB histone demethylase and the Bcl6 interacting corepressor. *Mol Cell Proteomics* **6**, 820-34.
- Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J. and Kouzarides, T.** (2002). Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407-11.
- Sarraf, S. A. and Stancheva, I.** (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol Cell* **15**, 595-605.
- Savarese, F., Flahndorfer, K., Jaenisch, R., Busslinger, M. and Wutz, A.** (2006). Hematopoietic precursor cells transiently reestablish permissiveness for X inactivation. *Mol Cell Biol* **26**, 7167-77.
- Sawicka, K., Bushell, M., Spriggs, K. A. and Willis, A. E.** (2008). Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. *Biochem Soc Trans* **36**, 641-7.
- 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**, 787-94.
- Scherer, W. F., Syverton, J. T. and Gey, G. O.** (1953). Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix. *J Exp Med* **97**, 695-710.
- Scheuermann, M. O., Tajbakhsh, J., Kurz, A., Saracoglu, K., Eils, R. and Lichter, P.** (2004). Topology of genes and nontranscribed sequences in human interphase nuclei. *Exp Cell Res* **301**, 266-79.
- Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C. and Kouzarides, T.** (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol* **6**, 73-7.
- Schoeftner, S., Sengupta, A. K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T. and Wutz, A.** (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *Embo J* **25**, 3110-22.
- Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D. and Jenuwein, T.** (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev* **18**, 1251-62.
- Schramke, V., Sheedy, D. M., Denli, A. M., Bonila, C., Ekwall, K., Hannon, G. J. and Allshire, R. C.** (2005). RNA-interference-directed chromatin modification coupled to RNA polymerase II transcription. *Nature* **435**, 1275-9.
- Schubeler, D., MacAlpine, D. M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D. E., O'Neill, L. P., Turner, B. M., Delrow, J., Bell, S. P. and Groudine, M.** (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev* **18**, 1263-71.
- Scott, M. J., Pan, L. L., Cleland, S. B., Knox, A. L. and Heinrich, J.** (2000). MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in Drosophila. *Embo J* **19**, 144-55.
- Senner, C. E. and Brockdorff, N.** (2009). Xist gene regulation at the onset of X inactivation. *Curr Opin Genet Dev* **19**, 122-6.
- Shankaranarayana, G. D., Motamedi, M. R., Moazed, D. and Grewal, S. I.** (2003). Sir2 regulates histone H3 lysine 9 methylation and heterochromatin assembly in fission yeast. *Curr Biol* **13**, 1240-6.
- Shao, C. and Takagi, N.** (1990). An extra maternally derived X chromosome is deleterious to early mouse development. *Development* **110**, 969-75.
- Shao, C. S. and Takagi, N.** (1991). Karyotypes and X chromosome inactivation in segregants of a murine X-autosome translocation, T(X;4)37H. *Jpn J Genet* **66**, 433-47.
- Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J. R., Wu, C. T., Bender, W. and Kingston, R. E.** (1999). Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**, 37-46.
- Sharman, G. B.** (1971). Late DNA replication in the paternally derived X chromosome of female kangaroos. *Nature* **230**, 231-2.



- Sheardown, S. A., Duthie, S. M., Johnston, C. M., Newall, A. E., Formstone, E. J., Arkell, R. M., Nesterova, T. B., Alghisi, G. C., Rastan, S. and Brockdorff, N. (1997a). Stabilization of Xist RNA mediates initiation of X chromosome inactivation. *Cell* **91**, 99-107.
- Sheardown, S. A., Newall, A. E., Norris, D. P., Rastan, S. and Brockdorff, N. (1997b). Regulatory elements in the minimal promoter region of the mouse Xist gene. *Gene* **203**, 159-68.
- Shen, Y., Matsuno, Y., Fouse, S. D., Rao, N., Root, S., Xu, R., Pellegrini, M., Riggs, A. D. and Fan, G. (2008). X-inactivation in female human embryonic stem cells is in a nonrandom pattern and prone to epigenetic alterations. *Proc Natl Acad Sci U S A* **105**, 4709-14.
- Shevchenko, A. I., Zakharova, I. S., Elisaphenko, E. A., Kolesnikov, N. N., Whitehead, S., Bird, C., Ross, M., Weidman, J. R., Jirtle, R. L., Karamysheva, T. V., Rubtsov, N. B., VandeBerg, J. L., Mazurok, N. A., Nesterova, T. B., Brockdorff, N. and Zakian, S. M. (2007). Genes flanking Xist in mouse and human are separated on the X chromosome in American marsupials. *Chromosome Res* **15**, 127-36.
- Shibata, S. and Lee, J. T. (2003). Characterization and quantitation of differential Tsix transcripts: implications for Tsix function. *Hum Mol Genet* **12**, 125-36.
- Shibata, S. and Lee, J. T. (2004). Tsix transcription- versus RNA-based mechanisms in Xist repression and epigenetic choice. *Curr Biol* **14**, 1747-54.
- Shibata, S. and Yokota, T. (2008). Alteration of histone tail modifications in the Xist locus in wild-type and Tsix-mutant male embryonic stem cells during differentiation. *Exp Anim* **57**, 153-7.
- Shibata, S., Yokota, T. and Wutz, A. (2008). Synergy of Eed and Tsix in the repression of Xist gene and X-chromosome inactivation. *Embo J* **27**, 1816-26.
- Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T. B., Webster, Z., Peters, A. H., Jenuwein, T., Otte, A. P. and Brockdorff, N. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell* **4**, 481-95.
- Silva, S. S., Rowntree, R. K., Mekhoubad, S. and Lee, J. T. (2008). X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. *Proc Natl Acad Sci U S A* **105**, 4820-5.
- Silver, D. P., Dimitrov, S. D., Feunteun, J., Gelman, R., Drapkin, R., Lu, S. D., Shestakova, E., Velmurugan, S., Denunzio, N., Dragomir, S., Mar, J., Liu, X., Rottenberg, S., Jonkers, J., Ganesan, S. and Livingston, D. M. (2007). Further evidence for BRCA1 communication with the inactive X chromosome. *Cell* **128**, 991-1002.
- Silverstein, R. A. and Ekwall, K. (2005). Sin3: a flexible regulator of global gene expression and genome stability. *Curr Genet* **47**, 1-17.
- Simmler, M. C., Cattanach, B. M., Raspberry, C., Rougeulle, C. and Avner, P. (1993). Mapping the murine Xce locus with (CA)<sub>n</sub> repeats. *Mamm Genome* **4**, 523-30.
- Simmler, M. C., Cunningham, D. B., Clerc, P., Vermat, T., Caudron, B., Cruaud, C., Pawlak, A., Szpirer, C., Weissenbach, J., Claverie, J. M. and 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. *Hum Mol Genet* **5**, 1713-26.
- Simmler, M. C., Heard, E., Rougeulle, C., Cruaud, C., Weissenbach, J. and Avner, P. (1997). Localization and expression analysis of a novel conserved brain expressed transcript, Brx/BRX, lying within the Xic/XIC candidate region. *Mamm Genome* **8**, 760-6.
- Simon, J. A. and Tamkun, J. W. (2002). Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes. *Curr Opin Genet Dev* **12**, 210-8.
- Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B. and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**, 1348-54.
- Simpson, P. J., Monie, T. P., Szendroi, A., Davydova, N., Tyzack, J. K., Conte, M. R., Read, C. M., Cary, P. D., Svergun, D. I., Konarev, P. V., Curry, S. and Matthews, S. (2004). Structure and RNA interactions of the N-terminal RRM domains of PTB. *Structure* **12**, 1631-43.
- Sims, R. J., 3rd, Nishioka, K. and Reinberg, D. (2003). Histone lysine methylation: a signature for chromatin function. *Trends Genet* **19**, 629-39.
- Sirchia, S. M., Ramoscelli, L., Grati, F. R., Barbera, F., Coradini, D., Rossella, F., Porta, G., Lesma, E., Ruggeri, A., Radice, P., Simoni, G. and Miozzo, M. (2005). Loss of the inactive X chromosome and replication of the active X in BRCA1-defective and wild-type breast cancer cells. *Cancer Res* **65**, 2139-46.
- Skibbens, R. V., Corson, L. B., Koshland, D. and Hieter, P. (1999). Ctf7p is essential for sister chromatid cohesion and links mitotic chromosome structure to the DNA replication machinery. *Genes Dev* **13**, 307-19.
- Sleutels, F., Zwart, R. and Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810-3.
- Smilnich, N. J., Day, C. D., Fitzpatrick, G. V., Caldwell, G. M., Lossie, A. C., Cooper, P. R., Smallwood, A. C., Joyce, J. A., Schofield, P. N., Reik, W., Nicholls, R. D., Weksberg, R., Driscoll, D. J., Maher, E. R., Shows, T. B. and Higgins, M. J. (1999). A maternally methylated CpG island in KvLQT1 is

- associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proc Natl Acad Sci U S A* **96**, 8064-9.
- Smith, E. R., Pannuti, A., Gu, W., Steurnagel, A., Cook, R. G., Allis, C. D. and Lucchesi, J. C.** (2000). The drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Mol Cell Biol* **20**, 312-8.
- Smith, K. P., Byron, M., Clemson, C. M. and Lawrence, J. B.** (2004). Ubiquitinated proteins including uH2A on the human and mouse inactive X chromosome: enrichment in gene rich bands. *Chromosoma* **113**, 324-35.
- Starmer, J. and Magnuson, T.** (2009). A new model for random X chromosome inactivation. *Development* **136**, 1-10.
- Stavropoulos, N., Lu, N. and Lee, J. T.** (2001). A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. *Proc Natl Acad Sci U S A* **98**, 10232-7.
- Stavropoulos, N., Rowntree, R. K. and Lee, J. T.** (2005). Identification of developmentally specific enhancers for Tsix in the regulation of X chromosome inactivation. *Mol Cell Biol* **25**, 2757-69.
- Stoger, R., Kubicka, P., Liu, C. G., Kafri, T., Razin, A., Cedar, H. and Barlow, D. P.** (1993). Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. *Cell* **73**, 61-71.
- Strahl, B. D. and Allis, C. D.** (2000). The language of covalent histone modifications. *Nature* **403**, 41-5.
- Straub, T., Neumann, M. F., Prestel, M., Kremmer, E., Kaether, C., Haass, C. and Becker, P. B.** (2005). Stable chromosomal association of MSL2 defines a dosage-compensated nuclear compartment. *Chromosoma* **114**, 352-64.
- Struhl, G.** (1981). A gene product required for correct initiation of segmental determination in Drosophila. *Nature* **293**, 36-41.
- Stryer, L. and Haugland, R. P.** (1967). Energy transfer: a spectroscopic ruler. *Proc Natl Acad Sci U S A* **58**, 719-26.
- Studier, F. W.** (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207-34.
- 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**, 113-30.
- Sudbrak, R., Wieczorek, G., Nuber, U. A., Mann, W., Kirchner, R., Erdogan, F., Brown, C. J., Wohrle, D., Sterk, P., Kalscheuer, V. M., Berger, W., Lehrach, H. and Ropers, H. H.** (2001). X chromosome-specific cDNA arrays: identification of genes that escape from X-inactivation and other applications. *Hum Mol Genet* **10**, 77-83.
- 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**, 133-8.
- Sugimoto, M. and Abe, K.** (2007). X chromosome reactivation initiates in nascent primordial germ cells in mice. *PLoS Genet* **3**, e116.
- Sugimoto, M., Tan, S. S. and Takagi, N.** (2000). X chromosome inactivation revealed by the X-linked lacZ transgene activity in periimplantation mouse embryos. *Int J Dev Biol* **44**, 177-82.
- Sugiyama, T., Cam, H., Verdel, A., Moazed, D. and Grewal, S. I.** (2005). RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc Natl Acad Sci U S A* **102**, 152-7.
- Sun, B. K., Deaton, A. M. and Lee, J. T.** (2006). A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol Cell* **21**, 617-28.
- Surani, M. A., Barton, S. C. and Norris, M. L.** (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308**, 548-50.
- Taddei, A.** (2007). Active genes at the nuclear pore complex. *Curr Opin Cell Biol* **19**, 305-10.
- Takagi, N.** (1983). De novo X-chromosome inactivation in somatic hybrid cells between the XO mouse embryonal carcinoma cell and XY rat lymphocyte. *Exp Cell Res* **145**, 397-404.
- Takagi, N. and Sasaki, M.** (1975). Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature* **256**, 640-2.
- Takahashi, I., Kameoka, Y. and Hashimoto, K.** (2002). MacroH2A1.2 binds the nuclear protein Spop. *Biochim Biophys Acta* **1591**, 63-8.
- Tam, P. P., Zhou, S. X. and Tan, S. S.** (1994). X-chromosome activity of the mouse primordial germ cells revealed by the expression of an X-linked lacZ transgene. *Development* **120**, 2925-32.
- Tate, P. H. and Bird, A. P.** (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr Opin Genet Dev* **3**, 226-31.
- Taylor, J. H. and Miner, P.** (1968). Units of DNA replication in mammalian chromosomes. *Cancer Res* **28**, 1810-4.
- Terranova, R., Yokobayashi, S., Stadler, M. B., Otte, A. P., van Lohuizen, M., Orkin, S. H. and Peters, A. H.** (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev Cell* **15**, 668-79.
- Thakur, N., Tiwari, V. K., Thomassin, H., Pandey, R. R., Kanduri, M., Gondor, A., Grange, T., Ohlsson, R. and Kanduri, C.** (2004). An antisense RNA regulates the bidirectional silencing property of the Kcnq1 imprinting control region. *Mol Cell Biol* **24**, 7855-62.

- Thomas, J. O. (1999). Histone H1: location and role. *Curr Opin Cell Biol* **11**, 312-7.
- Thornhill, A. R. and Burgoyne, P. S. (1993). A paternally imprinted X chromosome retards the development of the early mouse embryo. *Development* **118**, 171-4.
- Thorvaldsen, J. L., Duran, K. L. and Bartolomei, M. S. (1998). Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes Dev* **12**, 3693-702.
- Tijsterman, M. and Plasterk, R. H. (2004). Dicers at RISC; the mechanism of RNAi. *Cell* **117**, 1-3.
- Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**, 1453-65.
- Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A. D., Filipowicz, W., Ramos, A., Gherzi, R. and Rosenfeld, M. G. (2009). The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* **459**, 1010-4.
- Tremblay, G. A. and Richard, S. (2006). mRNAs associated with the Sam68 RNA binding protein. *RNA Biol* **3**, 90-3.
- Trent, C., Tsuing, N. and Horvitz, H. R. (1983). Egg-laying defective mutants of the nematode *Caenorhabditis elegans*. *Genetics* **104**, 619-47.
- 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. *Dev Biol* **319**, 416-25.
- Tsuchiya, K. D., Greally, J. M., Yi, Y., Noel, K. P., Truong, J. P. and Distèche, C. M. (2004). Comparative sequence and x-inactivation analyses of a domain of escape in human xp11.2 and the conserved segment in mouse. *Genome Res* **14**, 1275-84.
- Turner, B. M. (1993). Decoding the nucleosome. *Cell* **75**, 5-8.
- Turner, B. M. (2000). Histone acetylation and an epigenetic code. *Bioessays* **22**, 836-45.
- Turner, J. M., Aprelikova, O., Xu, X., Wang, R., Kim, S., Chandramouli, G. V., Barrett, J. C., Burgoyne, P. S. and Deng, C. X. (2004). BRCA1, histone H2AX phosphorylation, and male meiotic sex chromosome inactivation. *Curr Biol* **14**, 2135-42.
- Venkitaraman, A. R. (2002). Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* **108**, 171-82.
- Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S. I. and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672-6.
- Verona, R. I., Mann, M. R. and Bartolomei, M. S. (2003). Genomic imprinting: intricacies of epigenetic regulation in clusters. *Annu Rev Cell Dev Biol* **19**, 237-59.
- Verschure, P. J., van Der Kraan, I., Manders, E. M. and van Driel, R. (1999). Spatial relationship between transcription sites and chromosome territories. *J Cell Biol* **147**, 13-24.
- Vigneau, S., Augui, S., Navarro, P., Avner, P. and Clerc, P. (2006). An essential role for the DXPas34 tandem repeat and Tsix transcription in the counting process of X chromosome inactivation. *Proc Natl Acad Sci U S A* **103**, 7390-5.
- Villeneuve, A. M. and Meyer, B. J. (1987). *sdc-1*: a link between sex determination and dosage compensation in *C. elegans*. *Cell* **48**, 25-37.
- Vincent-Salomon, A., Ganem-Elbaz, C., Manie, E., Raynal, V., Sastre-Garau, X., Stoppa-Lyonnet, D., Stern, M. H. and Heard, E. (2007). X inactive-specific transcript RNA coating and genetic instability of the X chromosome in BRCA1 breast tumors. *Cancer Res* **67**, 5134-40.
- Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J. M., Bollen, M., Esteller, M., Di Croce, L., de Launoit, Y. and Fuks, F. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871-4.
- Volpi, E. V., Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., Campbell, R. D., Goldsworthy, M., Powis, S. H., Ragoussis, J., Trowsdale, J. and Sheer, D. (2000). Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J Cell Sci* **113** (Pt 9), 1565-76.
- Vu, T. H., Jirtle, R. L. and Hoffman, A. R. (2006). Cross-species clues of an epigenetic imprinting regulatory code for the IGF2R gene. *Cytogenet Genome Res* **113**, 202-8.
- Vu, T. H., Li, T. and Hoffman, A. R. (2004). Promoter-restricted histone code, not the differentially methylated DNA regions or antisense transcripts, marks the imprinting status of IGF2R in human and mouse. *Hum Mol Genet* **13**, 2233-45.
- 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. *Proc Natl Acad Sci U S A* **94**, 9665-8.
- Walsh, C. P., Chaillet, J. R. and Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* **20**, 116-7.
- Wang, J., Mager, J., Chen, Y., Schneider, E., Cross, J. C., Nagy, A. and Magnuson, T. (2001). Imprinted X inactivation maintained by a mouse Polycomb group gene. *Nat Genet* **28**, 371-5.
- Wang, L., Brown, J. L., Cao, R., Zhang, Y., Kassis, J. A. and Jones, R. S. (2004). Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell* **14**, 637-46.
- Wang, Q., Zhang, Z., Blackwell, K. and Carmichael, G. G. (2005). Vigilins bind to promiscuously A-to-I-edited RNAs and are involved in the formation of heterochromatin. *Curr Biol* **15**, 384-91.
- Wang, Z. Q., Fung, M. R., Barlow, D. P. and Wagner, E. F. (1994). Regulation of embryonic growth and lysosomal targeting by the imprinted Igf2/Mpr gene. *Nature* **372**, 464-7.
- Watson, J. D. and Crick, F. H. (1953). The structure of DNA. *Cold Spring Harb Symp Quant Biol* **18**, 123-31.

- Watt, F. and Molloy, P. L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev* **2**, 1136-43.
- Welshons, W. J. and Russell, L. B. (1959). The Y-Chromosome as the Bearer of Male Determining Factors in the Mouse. *Proc Natl Acad Sci U S A* **45**, 560-6.
- West, J. D., Frels, W. I., Papaioannou, V. E., Karr, J. P. and Chapman, V. M. (1977). Development of interspecific hybrids of Mus. *J Embryol Exp Morphol* **41**, 233-43.
- Wilkinson, F. H., Park, K. and Atchison, M. L. (2006). Polycomb recruitment to DNA in vivo by the YY1 REPO domain. *Proc Natl Acad Sci U S A* **103**, 19296-301.
- Williams, R. R. (2003). Transcription and the territory: the ins and outs of gene positioning. *Trends Genet* **19**, 298-302.
- Wolf, S. F., Dintzis, S., Toniolo, D., Persico, G., Lunnen, K. D., Axelman, J. and Migeon, B. R. (1984). Complete concordance between glucose-6-phosphate dehydrogenase activity and hypomethylation of 3' CpG clusters: implications for X chromosome dosage compensation. *Nucleic Acids Res* **12**, 9333-48.
- 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. *Mol Cell* **13**, 91-100.
- Wright, J. H. and Zakian, V. A. (1995). Protein-DNA interactions in soluble telosomes from *Saccharomyces cerevisiae*. *Nucleic Acids Res* **23**, 1454-60.
- Wu, H. A. and Bernstein, E. (2008). Partners in imprinting: noncoding RNA and polycomb group proteins. *Dev Cell* **15**, 637-8.
- Wutz, A. and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol Cell* **5**, 695-705.
- Wutz, A., Rasmussen, T. P. and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet* **30**, 167-74.
- Wutz, A., Smrzka, O. W., Schweifer, N., Schellander, K., Wagner, E. F. and Barlow, D. P. (1997). Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature* **389**, 745-9.
- Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., Kauer, M., Tackett, A. J., Chait, B. T., Badenhorst, P., Wu, C. and Allis, C. D. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86-90.
- Xiao, C., Sharp, J. A., Kawahara, M., Davalos, A. R., Difilippantonio, M. J., Hu, Y., Li, W., Cao, L., Buetow, K., Ried, T., Chadwick, B. P., Deng, C. X. and Panning, B. (2007). The XIST noncoding RNA functions independently of BRCA1 in X inactivation. *Cell* **128**, 977-89.
- Xu, N., Donohoe, M. E., Silva, S. S. and Lee, J. T. (2007). Evidence that homologous X-chromosome pairing requires transcription and Ctf protein. *Nat Genet* **39**, 1390-6.
- Xu, N., Tsai, C. L. and Lee, J. T. (2006). Transient homologous chromosome pairing marks the onset of X inactivation. *Science* **311**, 1149-52.
- Xu, Y., Goodyer, C. G., Deal, C. and Polychronakos, C. (1993). Functional polymorphism in the parental imprinting of the human IGF2R gene. *Biochem Biophys Res Commun* **197**, 747-54.
- Xue, F., Tian, X. C., Du, F., Kubota, C., Taneja, M., Dinnyes, A., Dai, Y., Levine, H., Pereira, L. V. and Yang, X. (2002). Aberrant patterns of X chromosome inactivation in bovine clones. *Nat Genet* **31**, 216-20.
- Yamaguchi, S., Kimura, H., Tada, M., Nakatsuji, N. and Tada, T. (2005). Nanog expression in mouse germ cell development. *Gene Expr Patterns* **5**, 639-46.
- Yamamoto, H., Tsukahara, K., Kanaoka, Y., Jinno, S. and Okayama, H. (1999). Isolation of a mammalian homologue of a fission yeast differentiation regulator. *Mol Cell Biol* **19**, 3829-41.
- Yang, T., Adamson, T. E., Resnick, J. L., Leff, S., Wevrick, R., Francke, U., Jenkins, N. A., Copeland, N. G. and Brannan, C. I. (1998). A mouse model for Prader-Willi syndrome imprinting-centre mutations. *Nat Genet* **19**, 25-31.
- 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**, 453-63.
- Yoder, J. A., Soman, N. S., Verdine, G. L. and Bestor, T. H. (1997). DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. *J Mol Biol* **270**, 385-95.
- Zamore, P. D. and Green, M. R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc. Natl. Acad. Sci.* **86**, 9243-7.
- Zamore, P. D. and Green, M. R. (1991). Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *EMBO J.* **10**, 207-14.
- Zamore, P. D. and Haley, B. (2005). Ribo-gnome: the big world of small RNAs. *Science* **309**, 1519-24.
- Zhang, H., Christoforou, A., Aravind, L., Emmons, S. W., van den Heuvel, S. and Haber, D. A. (2004). The *C. elegans* Polycomb gene SOP-2 encodes an RNA binding protein. *Mol Cell* **14**, 841-7.
- 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**, 693-706.
- Zhang, M., Zamore, P. D., Carmo-Fonseca, M., Lamond, A. I. and Green, M. R. (1992). Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc. Natl. Acad. Sci. USA* **89**, 8769-73.

- 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* **322**, 750-6.
- Zhao, Q., Rank, G., Tan, Y. T., Li, H., Moritz, R. L., Simpson, R. J., Cerruti, L., Curtis, D. J., Patel, D. J., Allis, C. D., Cunningham, J. M. and Jane, S. M.** (2009). PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat Struct Mol Biol* **16**, 304-11.
- Zhou, G. L., Xin, L., Song, W., Di, L. J., Liu, G., Wu, X. S., Liu, D. P. and Liang, C. C.** (2006). Active chromatin hub of the mouse alpha-globin locus forms in a transcription factory of clustered housekeeping genes. *Mol Cell Biol* **26**, 5096-105.
- Zirbel, R. M., Mathieu, U. R., Kurz, A., Cremer, T. and Lichter, P.** (1993). Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries. *Chromosome Res* **1**, 93-106.
- Zlatanova, J. and van Holde, K.** (1998). Linker histones versus HMG1/2: a struggle for dominance? *Bioessays* **20**, 584-8.
- Zuccotti, M., Boiani, M., Ponce, R., Guizzardi, S., Scandroglio, R., Garagna, S. and Redi, C. A.** (2002). Mouse Xist expression begins at zygotic genome activation and is timed by a zygotic clock. *Mol Reprod Dev* **61**, 14-20.