

Thèse présentée pour obtenir le grade de
Docteur de l'Université Louis Pasteur
Strasbourg I

Discipline : Sciences du Vivant
Spécialité : Aspects Moléculaires et Cellulaires de la Biologie

Par
Marie-Aline GERARD

**Caractérisation et analyse fonctionnelle
de nouveaux gènes cibles
du facteur de transcription hStaf/ZNF143**

Soutenue publiquement le 19 novembre 2007

Membres du jury

Directeur de thèse : M Philippe CARBON, Professeur à l'Université Louis Pasteur

Rapporteur Interne : M Irwin DAVIDSON, Directeur de Recherche du CNRS, Strasbourg

Rapporteur Externe : M Olivier LEFEBVRE, Chercheur au CEA, Gif-sur-Yvette

Rapporteur Externe : M Martin TEICHMANN, Professeur à l'Université Victor Segalen Bordeaux

Examineur : M Eric WESTHOF, Professeur à l'Université Louis Pasteur

REMERCIEMENTS

Le travail de thèse présenté dans ce mémoire a été effectué au sein du laboratoire Architecture et Réactivité de l'ARN, UPR9002 ULP/CNRS, de l'Université Louis Pasteur de Strasbourg. Je souhaite remercier le Pr. Bernard Ehresmann et son successeur le Pr. Eric Westhof de m'avoir accueillie au sein de leur unité.

J'aimerais également remercier Messieurs Irwin Davidson de l'Institut de Génétique et Biologie Moléculaire et Cellulaire (Strasbourg), Olivier Lefebvre de l'Institut de Biologie et de Technologies de Saclay (Gif-sur-Yvette) et Martin Teichmann de l'Institut Européen de Chimie et Biologie (Pessac) d'avoir accepté d'évaluer mon travail de thèse. Je remercie également Eric Westhof d'avoir bien voulu examiner mon travail.

Mes remerciements se dirigent également vers Alain Krol pour m'avoir permis de réaliser mon DEA et ce doctorat au sein de son équipe.

Un grand MERCI est dédié à mon directeur de thèse, Philippe Carbon, pour sa disponibilité, ses diverses compétences et connaissances scientifiques ainsi que ses nombreux conseils.

TABLE DES MATIÈRES

Table des matières.....	1
Table des illustrations.....	5
Liste des abréviations	7
INTRODUCTION.....	9
1. Diversité des promoteurs des gènes transcrits par l'ARN polymérase II chez les mammifères.	13
1.1. Vision classique des promoteurs.....	13
1.1.1. Les éléments caractéristiques du promoteur de base des gènes de protéines	14
1.1.2. Le promoteur de base des gènes d'ARNsn et d'ARNsno	18
1.1.3. Les promoteurs de base des gènes d'ARNmi	19
1.2. Cartographie des sites d'initiation de la transcription	19
1.2.1. Méthodes reposant sur l'hybridation de produits dérivés de l'ARN	20
1.2.2. Méthodes reposant sur le séquençage de produits dérivés de l'ARN.....	22
1.3. Nouvelle vision du transcriptome issu de l'ARN polymérase II.....	28
1.3.1. Nouvelle anatomie du promoteur de base	29
1.3.2. Transcription permissive	32
1.3.3. Les promoteurs alternatifs	35
1.3.4. Les promoteurs bidirectionnels	38
2. Régulation de l'initiation de la transcription dépendante de l'ARN polymérase II.....	43
2.1. Les activateurs de la transcription.....	43
2.2. Cartographie des sites de fixation de facteurs de transcription	46
2.2.1. Identification des sites de fixation de facteurs de transcription à l'échelle d'un gène	46
2.2.2. Identification des sites de fixation de facteurs de transcription à l'échelle d'un génome.....	47
2.3. Large distribution des sites de fixation des facteurs de transcription dans le génome.....	51
2.3.1. Localisation classique des sites de liaison de facteurs de transcription	51
2.3.2. Répartition inattendue des sites de fixation de facteurs de transcription le long du génome	52
2.3.3. Détermination de réseaux de régulation.....	54

3. Staf, un activateur de la transcription à multiples facettes	57
3.1. Activateur de la transcription du gène de l'ARN de transfert de la sélénocystéine.....	57
3.2. Activateur de la transcription de gènes d'ARNs non codants et d'ARNs messagers	59
3.3. Les protéines homologues de Staf chez l'homme	64
3.4. Mode de liaison de Staf à l'ADN.....	68
 RÉSULTATS ET DISCUSSIONS.....	 75
 CHAPITRE 1 : Identification des sites de liaison de hStaf/ZNF143 dans les promoteurs des gènes de protéines humaines	 79
1. Introduction :	79
2. Résultats :	80
2.1. Résumé des principaux résultats	80
2.2. Publication : « A genome scale location analysis of human Staf/ZNF143-binding sites suggests a widespread role for human hStaf/ZNF143 in mammalian promoters »	83
3. Discussion	95
 CHAPITRE 2 : hStaf/ZNF143 participe à la régulation de l'expression du gène TFAM humain.....	 101
1. Introduction :	101
2. Résultats :	102
2.1. Résumé des principaux résultats	102
2.2. Publication : « Transcription factor hStaf/ZNF143 is required for expression of the human TFAM gene»	106
3. Discussion	117
 CHAPITRE 3 : hStaf/ZNF143 est impliqué dans la transcription du gène BUB1B humain régulé au cours du cycle cellulaire	 122
1. Introduction :	122
2. Résultats :	123
2.1. Résumé des principaux résultats	123
2.2. Publication : « Transcription of the human cell cycle regulated BUB1B gene requires hStaf/ZNF143 »	127
3. Discussion	140

CHAPITRE 4 : hStaf/ZNF143 participe à la transcription indépendante du gène SCARNA2	144
1. Introduction :	144
2. Résultats :	146
2.1. Résumé des principaux résultats	146
2.2. Manuscrit : « The scaRNA2 gene is encoded by an independent unit and its transcription is regulated by hStaf/ZNF143 »	149
3. Discussion	172
 CONCLUSION GÉNÉRALE ET PERSPECTIVES.....	176
 BIBLIOGRAPHIE.....	186

TABLE DES ILLUSTRATIONS

Tableau 1 : Gènes cibles des ARN polymérase ADN dépendantes eucaryotiques	11
Tableau 2 : Résumé du code à une lettre dédié à l'ADN d'après l'IUPAC (International Union of Pure and Applied Chemistry).....	15
Tableau 3 : Exemples de facteurs de transcription dont les sites de liaison ont été identifiés par des analyses à haut débit.	51
Figure 1 : Représentation schématique des éléments du promoteur de base des gènes de protéines transcrits par l'ARN polymérase II	15
Figure 2 : Principales étapes de l'analyse à haut débit du génome sur des puces à ADN (d'après les ressources pédagogiques du site internet de la société Affymetrix (http://www.affymetrix.com)).....	21
Figure 3 : Construction d'une banque de données d'ADN complémentaires complets.	23
Figure 4 : Représentation schématique de la technique Rapid Amplification cDNA End (RACE) (d'après Sandelin et al., 2007).	24
Figure 5 : Utilisation des séquences étiquettes pour la caractérisation des extrémités d'un transcrit (d'après Harbers et Carninci, 2005).....	25
Figure 6 : Principe de la technique de Cap Analysis Gene Expression (CAGE) (d'après Harbers et Carninci, 2005).	26
Figure 7 : Illustration de la technique de PET (Paired-End diTag) (d'après Sandelin et al., 2007).	27
Figure 8 : Deux classes de promoteurs chez les mammifères (d'après Sandelin et al., 2007).....	29
Figure 9 : Organisation chevauchante des unités de transcription (d'après Kapranov et al., 2007b).....	34
Figure 10 : Fonctions et expressions des gènes associés à des promoteurs alternatifs et à des promoteurs uniques (d'après Baek et al., 2007).....	36
Figure 11 : Existence de promoteurs exoniques (d'après Carninci et al., 2006).....	38
Figure 12 : Organisations de deux gènes adjacents sur le génome.	39
Figure 13 : Représentation schématique des modes d'actions possibles d'un activateur de la transcription (d'après Fry et Peterson, 2001).	45
Figure 14 : Différentes stratégies d'identification des sites de fixation d'un facteur de transcription basées sur l'immunoprécipitation de chromatine (adapté d'après Hawkins et Ren, 2006).	48
Figure 15 : Architecture des promoteurs des gènes des ARNs de transfert chez les vertébrés.....	57
Figure 16 : Représentation schématique des domaines structuraux et fonctionnels de l'activateur de la transcription Staf (Schuster et al., 1995; Schuster et al., 1998).	59

Figure 17 : Architecture des promoteurs des gènes d'ARNsn et de type ARNsn transcrits par la Pol II (A) et la Pol III (B).....	60
Figure 18 : Caractéristiques structurales du domaine d'activation de la transcription de Staf (Schuster et al., 1998).....	63
Figure 19 : Représentation schématique des protéines homologues ZNF143, Staf et ZNF76 (d'après Myslinski <i>et al.</i> , 1998).....	65
Figure 20 : Détermination de la séquence consensus du site de liaison de Staf.	69
Figure 21 : Le domaine de liaison à l'ADN de Staf est constitué de sept doigts de zinc de type C2H2.....	72
Figure 22 : Représentation schématique des résultats obtenus lors de la recherche bioinformatique de SBS dans les promoteurs de gènes de protéines humaines (d'après Myslinski et al., 2006).	81
Figure 23 : Comparaison de 4 motifs conservés et fréquemment retrouvés dans le génome de mammifères avec le motif ACTACAN+SBS.....	96
Figure 24 : Représentation schématique de la région promotrice du gène TFAM humain (d'après Gerard et al., 2007b).....	103
Figure 25 : Représentation schématique de la région promotrice du gène BUB1B humain (d'après Myslinski et al., 2007).....	124
Figure 26 : Représentation schématique de la région promotrice du gène SCARNA2 humain (d'après Gerard et al., 2007a soumis pour publication).....	147
Figure 27 : Représentation schématique des promoteurs des gènes régulés par hStaf/ZNF143 chez les vertébrés caractérisés au laboratoire (Schuster et al., 1995; Schaub et al., 1997; Gerard et al., 2007b; Gerard et al., 2007a soumis pour publication; Myslinski et al., 2007).....	179

LISTE DES ABRÉVIATIONS

AD	activation domain
ADN	acide désoxyribonucléique
ADNc	acide désoxyribonucléique complémentaire
ADNm	acide désoxyribonucléique mitochondrial
ARN	acide ribonucléique
ARNc	acide ribonucléique complémentaire
ARNm	ARN messenger
ARNmi	micro ARN
ARNnc	ARN non codant
ARNr	ARN ribosomique
ARNsca	ARN des corps de Cajal
ARNsn	petit ARN nucléaire
ARNsno	petit ARN nucléolaire
ARNt	ARN de transfert
ATP	adénosine 5'-triphosphate
BRE	TFIIB recognition element
CAGE	cap analysis of gene expression
CDE	cell cycle dependant element
ChIP	chromatin immunoprecipitation
ChIP-seq	chromatin immunoprecipitation-sequencing
CHR	cell cycle genes homology regions
DBD	DNA binding domain
DBTSS	database of transcriptional start sites
DNase	désoxyribonucléase
dNTP	désoxyribonucléoside 5'-triphosphate
DPE	distal promoter element
DSE	distal sequence element
ENCODE	encyclopedia of DNA elements
EST	expressed sequence tag
GIS	gene identification signature
IGBMC	institut de génétique et de biologie moléculaire et cellulaire
Inr	initiateur

IUPAC	international union of pure and applied chemistry
kpb, kDa	kilo paire de bases, kiloDalton
LTR	long terminal repeat
MTE	motif ten element
NTP	ribonucléoside 5'-triphosphate
nts	nucléotides
pb	paire de bases
PCR	polymerase chain reaction
PET	paired-end ditag
PIC	pre-initiation complex
Pol I, II, III	ARN polymérase I, II, III
pré-ARNm	précurseur des ARN messagers
PSE	proximal sequence element
RACE	rapid amplification of cDNA ends
RD	repression domain
RNase	ribonucléase
RNP _{sno}	small nuclear ribonucleoprotein
RT-PCR	reverse transcriptase-polymerase chain reaction
SAGE	serial analysis of gene expression
SBS	staf binding site
Sec	sélénocystéine
SELEX	systematic evolution ligands by exponential enrichment
SNAP _c	snRNA activating protein complex
Staf	selenocysteine tRNA gene transcription activating factor
STAGE	sequence tag analysis of genomic enrichment
TAF	TBP associated factor
TBP	TATA binding protein
TFII	transcription factor II
TSS	site d'initiation de la transcription
5' UTR, 3' UTR	5' untranslated region, 3' untranslated region
Zf	zinc-finger
ZNF143,76	zinc-finger protein 143, 76

INTRODUCTION

Dans la cellule, le processus de transcription est un des principaux mécanismes régulé en réponse aux différents stimuli. La transcription consiste en la copie de l'information génétique contenue sur le brin matrice d'une portion de la molécule d'ADN appelée gène. Cette copie génère une molécule d'ARN qui pourra soit être traduite en protéine dans le cas des ARN messager (ARNm), soit être fonctionnelle en tant que telle pour les molécules appelées, dans ce cas, ARN non codant (ARNnc). Ce dernier type de molécules est impliqué dans des mécanismes cellulaires variés et regroupe, entre autres, les ARN de transfert (ARNt) et les ARN ribosomiques (ARNr) intervenant dans la synthèse protéique ou les petits ARN nucléaires (ARNsn) et nucléolaires (ARNsno) ainsi que les petits ARN des corps de Cajal (ARNsca) participant à la maturation d'ARN cellulaires.

Le mécanisme de transcription est réalisé par des enzymes appelées ARN polymérase ADN dépendantes. Chez les eucaryotes, il existe trois ARN polymérase chacune spécifique d'une certaine classe de gènes (Tableau 1). L'ARN polymérase I (Pol I) transcrit les gènes codant les ARNr tandis que l'ARN polymérase II (Pol II) est responsable de la synthèse des ARNm et de certains petits ARNs nucléaires. Les gènes des ARNt, de l'ARNr 5S ainsi que certains ARNnc tels que l'ARNsn U6 ou l'ARN 7SK sont, quant à eux, transcrits par l'ARN polymérase III (Pol III).

ARN polymérase I	ARNr 28S, 18S, 5.8S (précurseur)
ARN polymérase II	ARNm ARNsn U1, U2, U4, U5, U11, U11 et U4atac ARNsno U3, U8 et U13 ARN de la télomérase ARNmi (précurseurs)
ARN polymérase III	ARNt ARNr 5S ARN 7SL ARNsn U6 ARN de la MRP ARN H1 de la RNase P ARN 7SK ARN Y4 ARNmi (précurseurs)

Tableau 1: ARN synthétisés par les ARN polymérase ADN dépendantes eucaryotiques

La transcription se réalise en trois étapes : l'initiation, l'élongation et la terminaison. L'initiation de la transcription consiste en la reconnaissance du promoteur par la machinerie de transcription, l'ouverture de la double hélice d'ADN et la synthèse d'un hybride ARN/ADN. Si cet hybride est stable, le transcrit naissant sera allongé durant la phase d'élongation. L'étape de terminaison permettra le relargage du transcrit au niveau de signaux de terminaison de la transcription. La maturation du transcrit, comme l'ajout d'une coiffe à l'extrémité 5', est réalisée parallèlement à l'élongation de la transcription.

L'initiation est l'étape la plus régulée du mécanisme de transcription. Cette étape fait intervenir, outre la structure de la chromatine, différents facteurs de transcription qui participent à la reconnaissance du promoteur par la machinerie de transcription et régulent ainsi l'expression des gènes. Au cours de ma thèse, j'ai caractérisé de nouveaux gènes cibles du facteur de transcription Staf. Mes travaux ont consisté en une analyse fine du promoteur de trois gènes particuliers ainsi qu'à une identification des sites de liaison à l'ADN de Staf le long du génome humain. Ces travaux ont porté sur des gènes transcrits par la Pol II. De ce fait, je m'intéresserai dans l'introduction en premier lieu à la diversité des promoteurs reconnus par cette polymérase puis à la régulation de la transcription des gènes transcrits par cette enzyme. Dans une troisième partie, je présenterai plus précisément les caractéristiques et les fonctions du facteur de transcription Staf.

1. Diversité des promoteurs des gènes transcrits par l'ARN polymérase II chez les mammifères.

Dans une cellule, l'expression de chaque gène est sous la dépendance d'une portion d'ADN appelée promoteur. L'étude de ces promoteurs permet de mieux comprendre le mode de régulation de la transcription des gènes. Pendant de nombreuses années, l'identification et la caractérisation des promoteurs ont été effectuées au cas par cas. Ces études individuelles sont actuellement toujours nécessaires pour la caractérisation fine d'un promoteur. Cependant, ces dernières années, l'obtention de séquences presque complètes du génome de nombreux organismes, associées à de nombreuses avancées technologiques ont permis d'étudier les promoteurs à l'échelle d'un génome entier. Ces analyses ont donné des résultats surprenants qui nous poussent à repenser la vision classique de l'initiation de la transcription. Dans celle-ci, l'initiation est dirigée au niveau d'un nucléotide particulier, par un promoteur de base spécifique localisé en amont d'un gène. Cependant, il apparaît maintenant que de nombreux gènes possèdent de multiples sites d'initiation de la transcription ainsi que des promoteurs variés tels que les promoteurs alternatifs ou bidirectionnels. Ceci participe à la production d'un grand nombre d'ARNs à partir d'un nombre limité de gènes et augmente la complexité du transcriptome des mammifères.

1.1. Vision classique des promoteurs

Le promoteur est constitué de deux entités : le promoteur de base et le promoteur étendu. Le promoteur de base est défini comme la séquence d'ADN minimale permettant l'initiation de la transcription *in vitro* (pour une revue voir Roeder, 1996). Cette séquence, d'environ 70 nucléotides (nts), est retrouvée de part et d'autre du site d'initiation de la transcription (TSS) annoté +1 au niveau de la molécule d'ADN. Cette région contient plusieurs motifs qui, seuls ou associés, permettent le recrutement de la Pol II via les facteurs généraux de la transcription (pour des revues voir Smale et Kadonaga, 2003; Juven-Gershon et al., 2006; Muller et al., 2007). Le promoteur étendu, quant à lui, permet de moduler l'expression des gènes, en réponse à des signaux temporels ou spatiaux, en contrôlant l'activité du promoteur de base. Le promoteur étendu contient des séquences régulatrices spécifiques, plus ou moins éloignées du promoteur de base qui agissent en *cis* en recrutant des régulateurs transcriptionnels. On parle

de promoteur proximal lorsque ces éléments sont localisés à proximité du promoteur de base (+/- 250 nts) et de promoteur distal lorsqu'ils se situent à plusieurs milliers de paires de bases (pb). Pour ces derniers, les éléments activateurs sont appelés « enhancers » tandis que les « silencers » correspondent aux éléments répresseurs (pour une revue voir Blackwood et Kadonaga, 1998). Afin de limiter leur action longue distance, il existe, dans le génome, des éléments appelés « insulators » (pour une revue voir Gaszner et Felsenfeld, 2006). Seule la composition des promoteurs de base sera détaillée ci-dessous.

1.1.1. Les éléments caractéristiques du promoteur de base des gènes de protéines

Sept éléments du promoteur de base des gènes codant des protéines ont été identifiés à ce jour : la boîte TATA, l'initiateur Inr, l'élément promoteur aval ou DPE (Downstream Promoter Element), les deux éléments reconnus par TFIIB ou BREs (TFIIB Recognition Elements), le motif 10 ou MTE (Motif Ten Element) et les îlots CpG.

- La boîte TATA

Ce motif est le premier élément identifié dans le promoteur de base des gènes codant des protéines chez les eucaryotes (Goldberg, 1979; Breathnach et Chambon, 1981). Chez les métazoaires, il est localisé 25 à 30 pb en amont du TSS (Figure 1). Chez *Saccharomyces cerevisiae*, cette séquence est située plus en amont, à une distance comprise entre 40 et 120 pb. La séquence consensus pour la boîte TATA est TATAWAAR d'après le code IUPAC présenté dans le Tableau 2 (Smale et Kadonaga, 2003). Cependant, plusieurs études ont montré qu'une grande variété de séquences riches en nucléotides A et T pouvait fonctionner *in vivo* (Singer et al., 1990; Patikoglou et al., 1999). Des mutations dans cet élément entraînent une réduction ou l'abolition de la transcription (Grosschedl et Birnstiel, 1980; Wasylyk et al., 1980; Breathnach et Chambon, 1981; Grosveld et al., 1981; Hu et Manley, 1981). Il a été montré que le déplacement de la boîte TATA affecte la position du TSS et que sa localisation optimale est -31/-30 (Ponjavic et al., 2006). Il a été longtemps admis que cet élément était caractéristique des promoteurs de gènes de protéines. Cependant, ces dernières années de nombreuses études ont montré que moins de 10 % des gènes humains transcrits par la Pol II contiennent une boîte TATA dans leur promoteur (FitzGerald et al., 2004; Bajic et al., 2006). L'orientation de cet élément n'est pas essentielle à la direction de la transcription.

On observe uniquement une réduction de la force du promoteur si cette boîte est en orientation inverse (Xu et al., 1991). De nombreuses études ont mis en évidence qu'un complexe protéique appelé TFIID (Transcription Factor II D) se lie à cet élément via le facteur TBP (TATA Binding Protein) (pour une revue voir Burley et Roeder, 1996). Cette séquence peut, à elle seule, diriger l'initiation de la transcription mais est souvent retrouvée associée à un autre élément du promoteur de base : l'Initiateur (Inr).

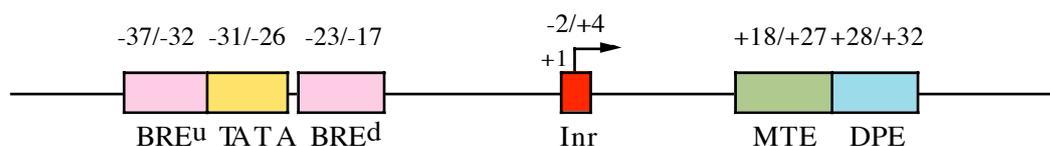


Figure 1 : Représentation schématique des éléments du promoteur de base des gènes de protéines transcrits par l'ARN polymérase II

Symboles	Bases concernées	Origine de la désignation
A	A	Adénine
C	C	Cytosine
G	G	Guanine
T	T	Thymine
R	A ou G	puRine
Y	C ou T	pYrimidine
M	A ou C	aMine
K	G ou T	cétone (Keto)
S	G ou C	interactions fortes (Strong) (trois liaisons hydrogènes)
W	A ou T	interactions faibles (Weak) (deux liaisons hydrogènes)
H	A ou T ou C	tous sauf G, H précède G dans l'alphabet
B	G ou T ou C	tous sauf A, B suit A dans l'alphabet
V	G ou C ou A	tous sauf T, V précède U (Uracile) qui précède T dans l'alphabet
D	G ou T ou A	tous sauf C, D suit C dans l'alphabet
N	A ou C ou G ou T	n'importe quelle base (aNy)

Tableau 2 : Résumé du code à une lettre dédié à l'ADN d'après l'IUPAC (International Union of Pure and Applied Chemistry)

- L'élément Initiateur

Cet élément chevauche le TSS et peut diriger *per se* l'initiation de la transcription. Cependant, il peut être associé à une boîte TATA et fonctionner en synergie avec celle-ci lorsque la distance entre ces deux éléments est comprise entre 25 et 30 pb chez les métazoaires (Figure 1) (O'Shea-Greenfield et Smale, 1992). La séquence consensus de l'Inr chez les mammifères est YYA₊₁NWYY (Javahery et al., 1994; Lo et Smale, 1996). Le nucléotide A₊₁ de cette séquence correspond au TSS le plus communément utilisé. Cependant, il a été montré que la transcription pouvait être initiée au niveau d'un autre nucléotide. Comme dans le cas de la boîte TATA, l'Inr n'est pas essentiel à la direction de la transcription et seule la force du promoteur est affectée lorsque l'orientation de l'Inr est inversée (O'Shea-Greenfield et Smale, 1992). Cet élément est reconnu par le complexe TFIID par l'intermédiaire de TAF1 et TAF2 (TBP Associated Factor 1 et 2) (Chalkley et Verrijzer, 1999).

- L'élément promoteur aval ou DPE (Downstream Promoter Element)

Cet élément a été identifié chez la drosophile comme liant TFIID dans les promoteurs dépourvus de boîte TATA (Burke et Kadonaga, 1996). Contrairement à la boîte TATA et à l'Inr, reconnus eux aussi par TFIID, l'élément DPE ne fonctionne pas indépendamment (Burke et Kadonaga, 1996; Burke et Kadonaga, 1997; Kutach et Kadonaga, 2000). Il est associé à un initiateur et ces deux éléments lient coopérativement le facteur général de transcription TFIID (Burke et Kadonaga, 1996). Des expériences de pontage ont révélé que les protéines TAF6 et TAF9 sont localisées à proximité de l'élément DPE et pourraient donc participer à la fixation de TFIID (Burke et Kadonaga, 1997). La DPE est localisée en aval du TSS aux positions +28/+32 par rapport à l'A₊₁ et sa séquence consensus est RGWYV (Figure 1) (Kutach et Kadonaga, 2000). Les promoteurs pourvus d'un élément DPE présentent une distance Inr-DPE identique. L'addition ou la suppression d'un seul nucléotide perturbe la liaison de TFIID et fait chuter le taux de transcription de base (Burke et Kadonaga, 1997; Kutach et Kadonaga, 2000).

- Les éléments reconnus par TFIIB ou BREs (TFIIB Recognition Elements)

Les structures cristallographiques du complexe ternaire TFIIB-TBP-TATA montrent que TFIIB établit des contacts avec l'ADN en amont et en aval de la boîte TATA (Nikolov et al.,

1995; Littlefield et al., 1999; Tsai et Sigler, 2000). Lagrange *et al* ont identifié l'élément amont appelé BRE^u (upstream BRE), localisé aux positions -37/-32, dont la séquence consensus est SSRCGCC (Lagrange et al., 1998). Le deuxième élément BRE^d (downstream BRE), localisé en aval de la boîte TATA, a été caractérisé plus tardivement et sa séquence consensus est RTDKKKK (Figure 1) (Deng et Roberts, 2005). Ces éléments sont, à ce jour, les seuls éléments du promoteur de base qui ne lient pas le facteur TFIID. Il a été montré que ces éléments possèdent un effet négatif ou positif sur la transcription de base selon le promoteur étudié (Lagrange et al., 1998; Evans et al., 2001).

- Le motif 10 ou MTE (Motif Ten Element)

Un autre élément du promoteur de base a été identifié récemment à partir d'une analyse comparative des promoteurs de base chez la drosophile (Ohler et al., 2002). Il a été montré que le motif 10, identifié dans cette analyse, renferme un élément intervenant dans le niveau basal de la transcription (Lim et al., 2004). Cet élément, appelé MTE, est localisé juste en amont de l'élément DPE aux positions +18/+27 par rapport à l'A₊₁ de l'Inr et sa séquence consensus est CSARCSSAAC (Figure 1) (Lim et al., 2004). Des similitudes avec l'élément DPE ont été observées. En effet, l'activité de cet élément nécessite la présence d'un Initiateur et la distance séparant ces deux éléments est essentielle (Lim et al., 2004). La plupart des promoteurs contenant un MTE ne possède pas de boîte TATA (Lim et al., 2004). L'élément MTE agit en synergie avec l'élément DPE et avec la boîte TATA lorsque celle-ci est présente (Lim et al., 2004).

- Les îlots CpG

Il existe dans le génome des régions enrichies en dinucléotides CG. Ces régions sont appelées îlots CpG lorsqu'elles contiennent au minimum 10 % de ces dinucléotides sur une longueur de 0,5 à 2 kpb (pour une revue voir Antequera, 2003). Pour comparaison, les régions du génome en dehors des îlots CpG contiennent en moyenne dix fois moins de dinucléotides CG. Une récente étude a montré que 72% des promoteurs des gènes de protéines sont associés à ce type de régions chez l'homme (Saxonov et al., 2006). Les îlots CpG sont majoritairement associés aux gènes constitutifs exprimés de façon ubiquitaire. Ces régions ne présentent pas de boîte TATA, de DPE et d'Inr et chevauchent l'extrémité 5' des gènes (Smale, 1994). Elles contiennent des sites de liaison pour des facteurs de transcription tels que le facteur Sp1 et

présentent une structure chromatinienne relâchée (Antequera et al., 1989; Blake et al., 1990; Somma et al., 1991; Brandeis et al., 1994; Macleod et al., 1994; Tommasi et Pfeifer, 1995; Litt et al., 1996; Gilbert et Sharp, 1999; Tommasi et Pfeifer, 1999). Le mécanisme d'initiation de la transcription au niveau de ces régions est peu connu à ce jour. Le recrutement de la machinerie de transcription est sûrement dépendant des facteurs de transcription se liant spécifiquement au promoteur distal (Smale et Kadonaga, 2003).

Le promoteur de base des gènes de protéines est composé d'une combinaison variable de tous ces éléments promoteurs (pour des revues voir Gross et Oelgeschlager, 2006; Juven-Gershon et al., 2006). Il est intéressant de noter que d'autres gènes sont dépendants de ces types de promoteurs. En effet, chez les vertébrés, la majorité des gènes d'ARNsno et d'ARNsca sont localisés dans des introns de gènes de protéines et le produit de ces gènes résulte de la maturation des introns des pré-ARNm (Bachelier et al., 2002).

1.1.2. Le promoteur de base des gènes d'ARNsn et d'ARNsno

La Pol II est également responsable de la synthèse de nombreux ARNnc à partir d'unités de transcription indépendantes. En effet, à l'exception de l'ARNsn U6, tous les gènes des petits ARNsn et les ARNsno U3, U8 et U13 sont transcrits par cette enzyme. Le promoteur de base de ces gènes renferme un élément en amont du TSS appelé PSE (Proximal Sequence Element) chez les vertébrés (Skuzeski et al., 1984). Cet élément a été initialement identifié dans le promoteur du gène de l'ARNsn U1 humain et constitue le promoteur de base des gènes d'ARNsn et de type ARNsn (Ares et al., 1985; Ciliberto et al., 1985; Mattaj et al., 1985; Murphy et al., 1987; Dahlberg et Schenborn, 1988; Hernandez et Lucito, 1988; McNamara et Stumph, 1990; Lobo et Hernandez, 1994). La PSE est centrée sur les positions -59/-56 et est requise pour la détermination de la position du TSS (Skuzeski et al., 1984; Lescure et al., 1992). La séquence de la PSE est très variable d'une espèce à l'autre ainsi que parmi les gènes d'une même espèce (Lund et Dahlberg, 1988). De plus, la longueur de cet élément peut varier de 10 à 18 pb dans le cas des gènes d'ARNsn chez le xénope ou l'homme respectivement (Hernandez, 1992). Cependant, la séquence TNNCCNT est relativement bien conservée d'un organisme à l'autre et est essentielle à la fonction de la PSE (Parry et al., 1989; McNamara et Stumph, 1990). La PSE est reconnue par un unique complexe multiprotéique appelé SNAPc via les sous-unités SNAP190 et SNAP50 (pour une revue voir Hernandez, 2001). On peut

noter que cet élément est associé à une boîte TATA dans le promoteur de base des gènes d'ARNsn transcrits par la Pol III (Carbon et al., 1987; Krol et al., 1987; Mattaj et al., 1988; Lobo et Hernandez, 1989).

1.1.3. Les promoteurs de base des gènes d'ARNmi

Les micro ARNs (ARNmi) sont de très courts ARNnc qui interviennent au niveau de la régulation post-transcriptionnelle de l'expression des gènes. De nombreux ARNmi ont été identifiés ces dernières années mais le mode de transcription de leurs gènes est encore peu connu à l'heure actuelle. Certains sont le produit de la transcription de leur propre gène par la Pol II (Lee et al., 2004b). Un long précurseur, appelé pri-ARNmi, coiffé et polyadénylé, est synthétisé puis il sera maturé par les RNases de type III, Drosha et Dicer, afin de libérer des duplexes d'ARNmi (Cullen, 2004; Lee et al., 2004b). Un seul brin de ce duplexe sera utilisé lors de la régulation. Récemment, il a été montré que certains gènes d'ARNmi sont transcrits par la Pol III (Borchert et al., 2006). Cependant, les promoteurs de ces gènes, transcrits par la Pol II ou la Pol III, n'ont pas encore été caractérisés.

Nous avons vu que les gènes transcrits par la Pol II sont sous le contrôle de promoteurs variés. Cependant, cette connaissance est la résultante de travaux réalisés sur des gènes au cas par cas. Ces dernières années, avec l'obtention de la séquence de génomes presque complets, une partie de la communauté scientifique s'est focalisée sur l'analyse de la régulation de l'expression des gènes à l'échelle des génomes. Pour mener à bien un tel projet, il est nécessaire de disposer d'approches globales qui permettent de cartographier les régions promotrices qui gouvernent l'expression des gènes.

1.2. Cartographie des sites d'initiation de la transcription

La cartographie des régions promotrices se réalise en caractérisant les sites d'initiation de la transcription le long du génome. Cette identification utilise des techniques basées soit sur l'hybridation, soit sur le séquençage, de produits dérivés de l'ARN issu de la transcription des gènes. On parlera d'analyse à haut débit lorsque les approches utilisées permettent d'obtenir un grand nombre de résultats à partir d'un grand échantillon dans un temps relativement court et suite à un nombre limité de manipulations.

1.2.1. Méthodes reposant sur l'hybridation de produits dérivés de l'ARN

Il y a quelques années, on utilisait encore la technique de protection à la nucléase S1 pour déterminer le TSS d'un gène. De nos jours, cette technique a été supplantée par des techniques d'extension d'amorces. Dans un premier temps, une réaction de transcription inverse est réalisée avec un oligonucléotide complémentaire à l'ARN d'intérêt. Le TSS est déterminé par comparaison de la migration sur gel d'électrophorèse du produit obtenu avec une échelle de séquençage d'ADN réalisée avec le même oligonucléotide.

Des techniques d'analyse à l'échelle d'un génome entier ont été développées ces dernières années. Celles-ci utilisent des puces contenant des milliers de fragments d'ADN. Ces fragments peuvent correspondre à des régions spécifiques du génome comme des îlots CpG ou recouvrir la séquence de chromosomes entiers ou d'un génome à l'exception des séquences répétées. Ces puces sont de deux types selon la nature des fragments les composant. Le premier type de puce est constitué de fragments de PCR ou d'ADNc de 200 à 2000 pb fixés à hauteur de 10 000 à 40 000 fragments par puce. Ces puces, appelées puces à ADN, peuvent être synthétisées en laboratoire. Le deuxième type de puce est composé d'oligonucléotides de 25 mers synthétisés directement sur le support à hauteur de 6 millions de fragments différents par puce. Ces puces, appelées puces à oligonucléotides, sont commercialisées par des sociétés comme Affymetrix, NimbleGen et Agilent. Dans le cas de ces puces, la résolution, définie comme la distance entre le centre de deux oligonucléotides adjacents, est généralement de 35 pb. Pour l'utilisation de ces deux types de puces, les ADNc ou des ARNc marqués à la biotine sont obtenus par une réaction de transcription inverse sur une population globale d'ARN (Figure 2). Ces ADNc et ARNc sont fragmentés puis hybridés sur les puces et après de nombreux lavages, les fragments hybridés sont révélés par un fluorochrome. La quantification des signaux obtenus, leur normalisation et l'identification des fragments reconnus se fait à l'aide de logiciels informatiques spécifiques. On pourra ainsi visualiser les loci génomiques responsables de la synthèse des transcrits analysés et déduire la région promotrice localisé en 5' (Figure 2). La validation des résultats obtenus ainsi que la détermination précise de l'extrémité 5' des ARNs seront réalisées en utilisant des techniques à haut débit basées sur le séquençage (SAGE, PET) présentées dans la partie suivante de mon manuscrit.

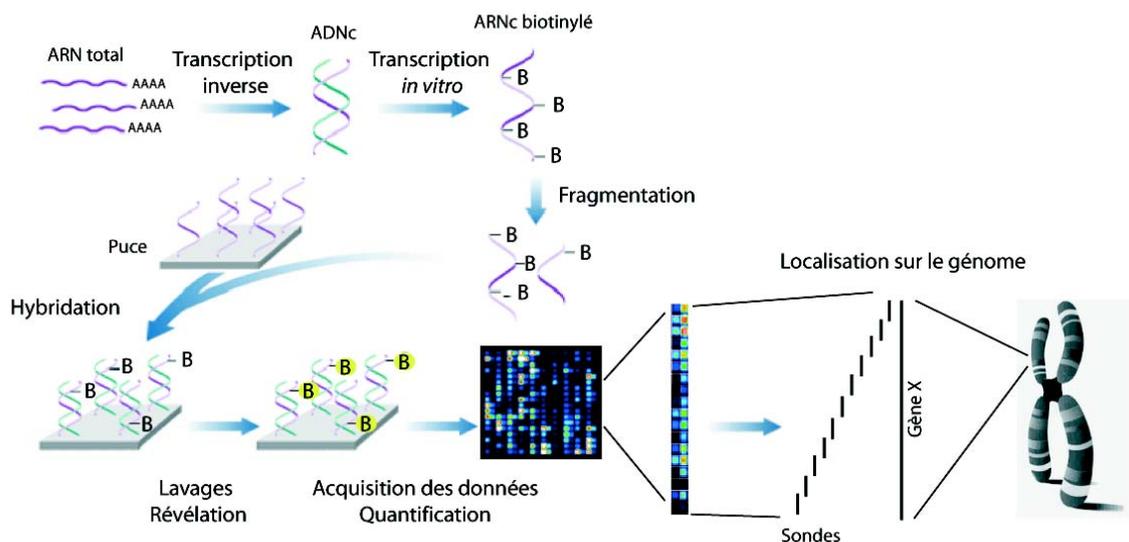


Figure 2 : Principales étapes de l'analyse à haut débit du génome sur des puces à ADN (d'après les ressources pédagogiques du site internet de la société Affymetrix (<http://www.affymetrix.com>)).

Les ARNs sont purifiés et marqués à la biotine par une réaction de transcription inverse suivie d'une transcription in vitro réalisée en présence de NTPs biotinylés. Les ARNc obtenus sont clivés puis hybridés sur les puces. Après des lavages successifs, les hybrides sont révélés à l'aide de streptavidine ou d'anticorps anti-biotine couplés à un fluorochrome. La puce est ensuite analysée au microscope confocal. La quantification des ARNs et leur localisation sur le génome sont obtenues après analyse des données brutes à l'aide de logiciels informatiques.

Les sites d'initiation de la transcription peuvent être également déterminés en localisant les complexes d'initiation de la transcription sur le génome. Ceci est réalisé par des expériences d'immunoprécipitation de chromatine couplées à l'analyse systématique des fragments immunoprécipités sur des puces (ChIP on chip). L'immunoprécipitation est réalisée avec des anticorps spécifiques de la Pol II hypo-phosphorylée ou dirigés contre des facteurs spécifiques de l'initiation de la transcription comme la protéine TAF1, un des composants du facteur général de la transcription TFIID. Les fragments d'ADN sélectionnés seront hybridés sur des puces à oligonucléotides afin de déterminer à l'échelle d'un génome entier les régions promotrices correspondantes. De manière alternative, les fragments immunoprécipités pourront être clonés et séquencés à haut débit selon la technique de ChIP-PET qui sera développée dans la partie 2.2.2 de mon introduction.

1.2.2. Méthodes reposant sur le séquençage de produits dérivés de l'ARN

Cette approche est basée sur l'isolement des ARNm cellulaires, la synthèse des ADNc complets, le séquençage et la localisation de ces ADNc sur le génome. La position de l'extrémité 5' de l'ARN sur le génome identifie le site d'initiation de la transcription. Cette détermination peut être également réalisée en utilisant seulement un petit fragment de l'ADNc, appelé étiquette, provenant de la partie 5' de l'ARN. Cette approche permet d'annoter très précisément le TSS et d'évaluer le niveau d'expression des gènes.

La synthèse d'ADNc par le mécanisme de transcription inverse a permis de constituer des banques de données d'ADNc telles que celles trouvées dans la NCBI Reference Sequence (RefSeq). Cependant, ces ADNc peuvent ne pas correspondre à une transcription inverse complète de l'ARN en particulier dans sa partie 5'. Ceci peut être dû à l'arrêt prématuré de la transcriptase inverse ou à la présence de molécules d'ARN tronquées. Afin de résoudre ce problème, différentes techniques ont été élaborées pour obtenir des ADNc complets portant la totalité de l'information de l'ARNm. Ces techniques reposent sur la présence de la coiffe qui se trouve uniquement à l'extrémité 5' des ARNm intègres. Deux approches sont principalement employées : le « cap-trapper » (Carninci et al., 1996; Carninci et al., 1997) et l'« oligo-capping » (Maruyama et Sugano, 1994; Suzuki et al., 1997). Dans la méthode de cap-trapper, la coiffe des ARNm est biotinylée afin de purifier, après la synthèse du premier brin d'ADNc, les hybrides ADNc/ARNm renfermant les copies complètes de l'ARNm sur des billes couplées à la streptavidine. Dans la deuxième approche, la coiffe est remplacée par un oligonucléotide de synthèse avant la synthèse du premier brin d'ADNc (Figure 3).

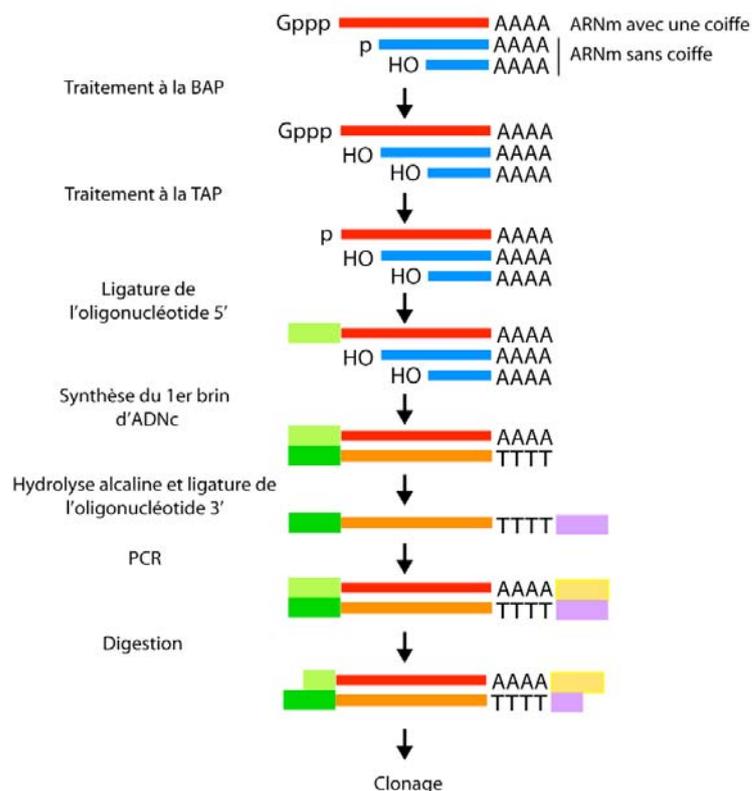


Figure 3 : Construction d'une banque de données d'ADN complémentaires complets.

La coiffe est remplacée par un oligonucléotide par trois réactions enzymatiques. Dans un premier temps, la phosphatase alcaline bactérienne (BAP) élimine le groupement phosphate présent à l'extrémité 5' des ARNs dépourvus de coiffe. Dans un deuxième temps, la pyrophosphatase acide du tabac (TAP) élimine la coiffe et libère un groupement phosphate en 5'. Un oligonucléotide sera lié à ces extrémités 5' par la T4 ARN ligase qui catalyse la formation d'une liaison phosphodiester entre une molécule portant un groupement phosphate en 5' et une molécule portant un groupement hydroxyl en 3'. Le premier brin d'ADNc est ensuite synthétisé en utilisant un oligonucléotide polydT qui s'hybride à la queue polyA des ARNm. Après dégradation de l'ARN et addition d'un autre oligonucléotide en 3' du brin d'ADNc, l'ADNc sera amplifié par PCR puis cloné.

Cette dernière méthode a été utilisée par Suzuki *et al* pour construire, après séquençage d'un grand nombre d'ADNc complets humains, la banque de données DBTSS (DataBase of Transcriptional Start Sites, <http://dbtss.hgc.jp/>) (Suzuki et al., 2002). Cette banque permet d'accéder avec précision aux sites d'initiation de la transcription de plusieurs milliers de gènes sur le génome puisque seuls les ADNc issus d'ARNm coiffés ont été utilisés pour sa construction. Ayant été amenée à utiliser cette banque de données durant mes travaux de thèse, je souhaite donner ici quelques informations supplémentaires sur celle-ci. La version actuelle 5.2.0 de la banque DBTSS repose respectivement sur les versions hg17 et mm5 des séquences du génome humain et murin (University of California, Santa Cruz, UCSC,

<http://genome.ucsc.edu/index.html>). De nombreuses améliorations ont été apportées depuis la création de DBTSS en 2002 (Yamashita et al., 2006). La quantité des données a particulièrement augmenté puisque l'on trouve désormais 1 359 000 ADNc humains dans cette banque en 2006 alors qu'elle n'en contenait qu'environ 200 000 en 2002. De plus, elle a été améliorée en 2004 par l'ajout des séquences d'ADNc murins. L'augmentation des données a par exemple permis de mettre en évidence la présence de promoteurs alternatifs pour certains gènes.

Parallèlement à la constitution de banque d'ADNc complets, le développement de la technique de RACE (Rapid Amplification cDNA End) a permis de caractériser la partie 5' des ARNm cappés (Figure 4). Cette technique est utilisée généralement pour identifier le TSS de gènes particuliers ou pour vérifier des TSS identifiés par des analyses à haut débit. Cette approche permet de séquencer la région 5' d'un ARN en utilisant la technique d'oligo-capping pour sélectionner les ARNs coiffés.

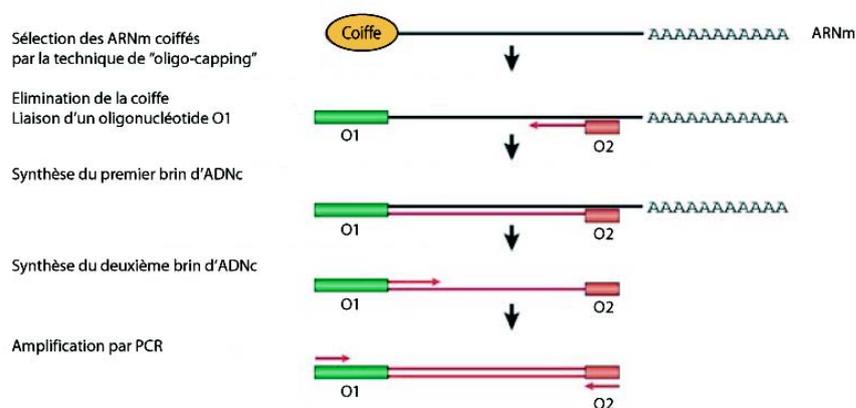


Figure 4 : Représentation schématique de la technique Rapid Amplification cDNA End (RACE) (d'après Sandelin et al., 2007).

La coiffe des ARNm sélectionnés par la technique d'oligo-capping est éliminée par un traitement à pyrophosphatase acide de tabac (TAP). Un oligonucléotide O1 est lié à l'extrémité 5' de l'ARNm. Le premier brin d'ADNc est alors synthétisé en utilisant un oligonucléotide O2 s'hybridant dans une région interne de l'ARNm. La région 5' des ARNm sélectionnés est ensuite amplifiée par une réaction de PCR en présence des oligonucléotides O1 et O2 puis clonée et séquencée.

Afin d'analyser le transcriptome à encore plus grande échelle, des techniques basées sur le séquençage à haut débit de très courtes régions des transcrits ont été développées ces dernières années. La Figure 5 représente schématiquement ces différents types d'étiquettes.

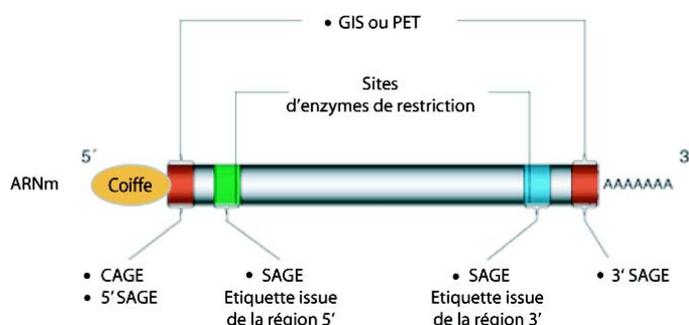


Figure 5 : Utilisation des séquences étiquettes pour la caractérisation des extrémités d'un transcrit (d'après Harbers et Carninci, 2005).

Chaque transcrit peut être caractérisé par une courte séquence qui, selon la technique utilisée, correspond à différentes régions. La technique de SAGE permet l'identification de régions proches de l'extrémité 3' ou 5' dont la localisation est déterminée par un site de restriction pour une enzyme de restriction à 4 pb (bleu et vert respectivement). Des adaptations de cette technique permettent l'isolement d'étiquettes correspondant précisément à l'extrémité 5' (CAGE, 5'SAGE) ou 3' du transcrit (3'SAGE) ou aux deux à la fois (GIS) (rouge).

La technique de SAGE (Serial Analysis of Gene Expression) fut la première méthode à utiliser de courtes séquences d'ADN comme une étiquette caractéristique d'un transcrit. Initialement, cette technique permettait de générer une étiquette correspondant à une région proche de l'extrémité 3' d'un transcrit (Figure 5) (Velculescu et al., 1995). Des adaptations permettent, à présent, de déterminer précisément les extrémités 5' et 3' d'un transcrit (5'SAGE et 3'SAGE) (Figure 5) (Hashimoto et al., 2004) (pour des revues voir Powell, 2000; Wang, 2005). De nouvelles techniques comme celles de CAGE (Cap Analysis of Gene Expression) (Shiraki et al., 2003) et de GIS (Gene Identification Signature) (Ng et al., 2005), permettant elles aussi d'identifier précisément l'extrémité 5' d'un transcrit, dérivent de la technique du SAGE (Figure 5). La technique de CAGE utilise la méthode de « cap-trapper » vue précédemment afin de sélectionner les ARNm coiffés (Figure 6). Les étiquettes obtenues ont une taille de 20 à 21 nts. Cette petite taille peut parfois poser des problèmes lors de l'étape de localisation de l'étiquette sur le génome. En effet, une même étiquette peut parfois

correspondre à plusieurs régions du génome. Pour limiter ce problème, la technique de SuperSAGE a été développée afin d'obtenir des étiquettes de 26 nts. Malgré cet inconvénient, la technique de CAGE est beaucoup utilisée pour l'identification à haut débit des sites d'initiation de la transcription car elle est plus facile à mettre en œuvre que le SuperSAGE.

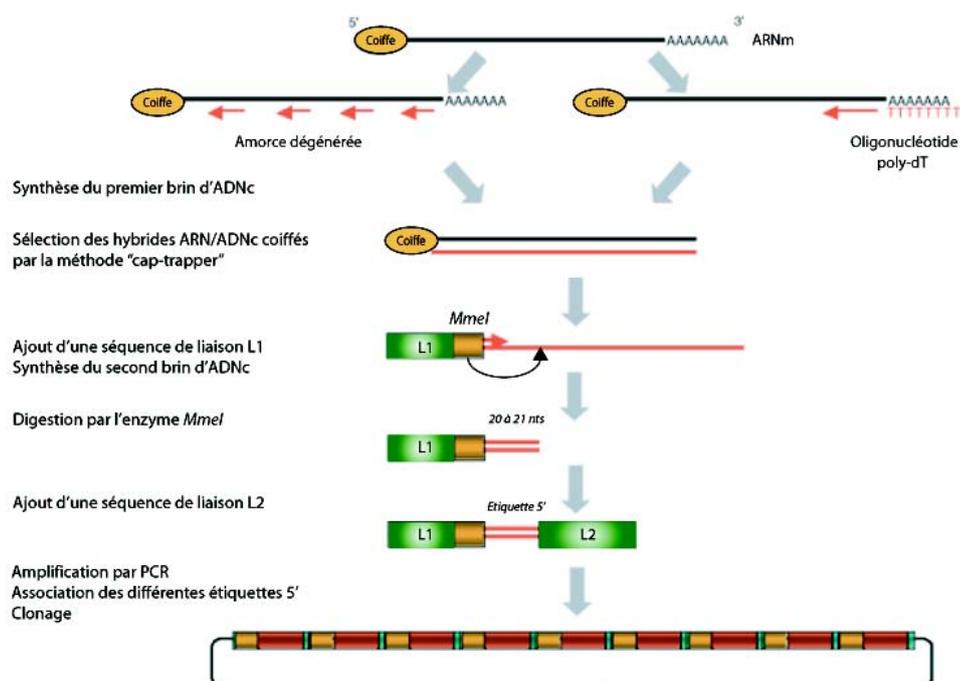


Figure 6 : Principe de la technique de Cap Analysis Gene Expression (CAGE) (d'après Harbers et Carninci, 2005).

Le premier brin d'ADNc est synthétisé en utilisant un oligonucléotide poly-dT ou dégénéré. Les hybrides ARN/ADNc coiffés sont sélectionnés par la méthode de « cap-trapping ». La coiffe est ensuite éliminée, l'ARN est hydrolysé et une séquence de liaison (L1), portant un site pour l'enzyme de restriction de type II *MmeI*, est ajoutée à l'extrémité 3' du brin d'ADNc. Le deuxième brin d'ADNc est alors synthétisé puis clivé par l'enzyme de restriction *MmeI* dont le site de coupure se localise 20 à 21 nucléotides en aval de son site de reconnaissance. Une séquence de liaison (L2) est liée à l'extrémité 3' du fragment obtenu. Les étiquettes caractéristiques de l'extrémité 5' des ARNm initiaux sont amplifiées par PCR, assemblées puis clonées.

Cependant, cette technique de CAGE ne permet pas d'avoir une idée de l'étendue du gène et de son produit de transcription. Cette information peut être déterminée en utilisant la technique de GIS (ou PETs pour Paired-End diTags) qui génère un court fragment contenant à

la fois l'extrémité 5' et 3' du transcrit sous forme de deux étiquettes de 18 nts chacune (Figure 7).

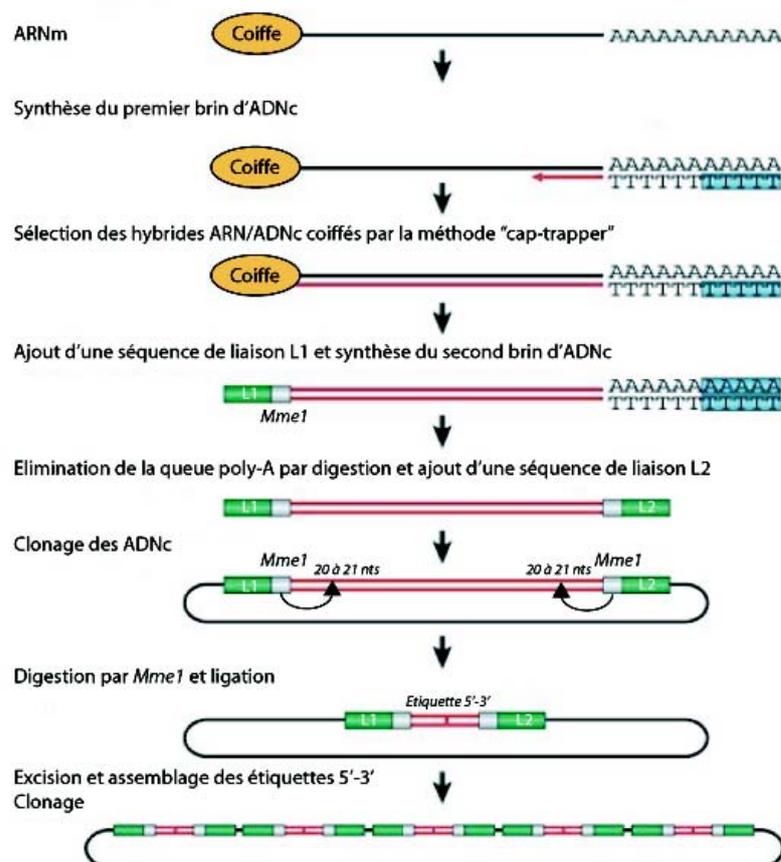


Figure 7 : Illustration de la technique de PET (Paired-End diTag) (d'après Sandelin et al., 2007).

Le premier brin d'ADNc est synthétisé en utilisant un oligonucléotide poly-dT renfermant un site pour l'enzyme de restriction *GsuI*. Les hybrides ARN/ADNc coiffés sont sélectionnés par la méthode de « cap-trapping ». La coiffe est ensuite éliminée, l'ARN est hydrolysé et une séquence de liaison portant un site pour l'enzyme de restriction de type II *MmeI* est ajoutée à l'extrémité 3' du brin d'ADNc. Le deuxième brin d'ADNc est alors synthétisé. Une séquence de liaison portant également un site *MmeI* est liée à l'extrémité 3' du double brin d'ADNc après élimination de la queue polyA par digestion avec l'enzyme de restriction *GsuI*. Après clonage, la partie interne des ADNc est éliminée, à l'exception des extrémités 5' et 3', par digestion avec l'enzyme *MmeI* dont le site de coupure se localise 20 à 21 nts en aval de son site de reconnaissance. Les différentes étiquettes 5'-3' sont excisées puis assemblées et clonées.

Les résultats avec cette approche sont obtenus plus rapidement qu'en appliquant les techniques de 5'SAGE et 3'SAGE sur un même échantillon. De plus, malgré la courte taille des étiquettes, l'assignation du site d'initiation de la transcription est facilitée par la vision

globale du transcrit. L'efficacité de cette technique repose essentiellement sur l'obtention d'ADNc complets. Ceci entraîne un biais puisque les longs transcrits seront sous-représentés car il est difficile de les cloner. On peut noter que la technique de CAGE peut éviter ce problème en utilisant des oligonucléotides dégénérés pour synthétiser le premier brin d'ADNc (Figure 6). Cette stratégie permet aussi d'étudier les ARNs dépourvus de queue polyadénylée. La technique de GIS/PETs est une approche plus fastidieuse que le CAGE puisqu'elle nécessite une étape de clonage supplémentaire. Cette technique est plus utilisée dans le cadre de projets d'annotation de gène que pour l'identification des TSS et donc la cartographie de promoteurs. Cependant, une extension de cette technique permet cette étude. En effet, des promoteurs peuvent être identifiés via des facteurs de transcription en couplant la technique d'immunoprécipitation de chromatine à la technique de PETs (ChIP-PET). Cette approche sera abordée plus précisément dans la partie 2.2.2 de mon manuscrit.

Ces techniques sont beaucoup plus productives et financièrement d'un moindre coût que les techniques de RACE et les banques de données d'ADNc complets. En effet, la multimérisation des étiquettes et le développement de techniques de séquençage à haut débit permettent d'analyser plusieurs étiquettes en une seule réaction de séquençage et donc d'accumuler plus d'informations. Par exemple dans le cas de la souris, la communauté scientifique dispose d'environ 100 000 ADNc complets et de plus de 10 millions d'étiquettes définissant l'extrémité 5' d'un transcrit (CAGE et autres).

A l'heure actuelle, de nombreuses études ont utilisées ces différentes techniques basées sur l'hybridation ou le séquençage pour analyser de manière systématique l'information contenue dans le génome. Une énorme quantité de données est sortie de ces analyses, nous allons nous intéresser particulièrement aux résultats obtenus concernant le transcriptome et les régions promotrices régulant l'expression des gènes.

1.3. Nouvelle vision du transcriptome issu de l'ARN polymérase II

Les analyses à haut débit du transcriptome remettent en question la vision conventionnelle du génome composé des gènes distincts possédant un promoteur spécifique qui dirige l'initiation de la transcription au niveau d'un site unique et montrent que le transcriptome est plus important qu'attendu.

1.3.1. Nouvelle anatomie du promoteur de base

Une cartographie fine des sites d'initiation de la transcription réalisée sur environ 6000 ARNm humains a montré l'existence d'une diversité de distribution des TSS pour un même gène (Suzuki et al., 2001). Cette observation a été confirmée par une analyse systématique des extrémités 5' du transcriptome humain et murin, par la technique de CAGE. Dans cette dernière étude, les auteurs définissent deux types de promoteurs (Carninci et al., 2006). Le premier type de promoteur dirige l'initiation de la transcription au niveau d'un nucléotide unique, tandis que pour le second type, la transcription débute au niveau de nombreux nucléotides dans une fenêtre de 50 à 100 pb (Figure 8A).

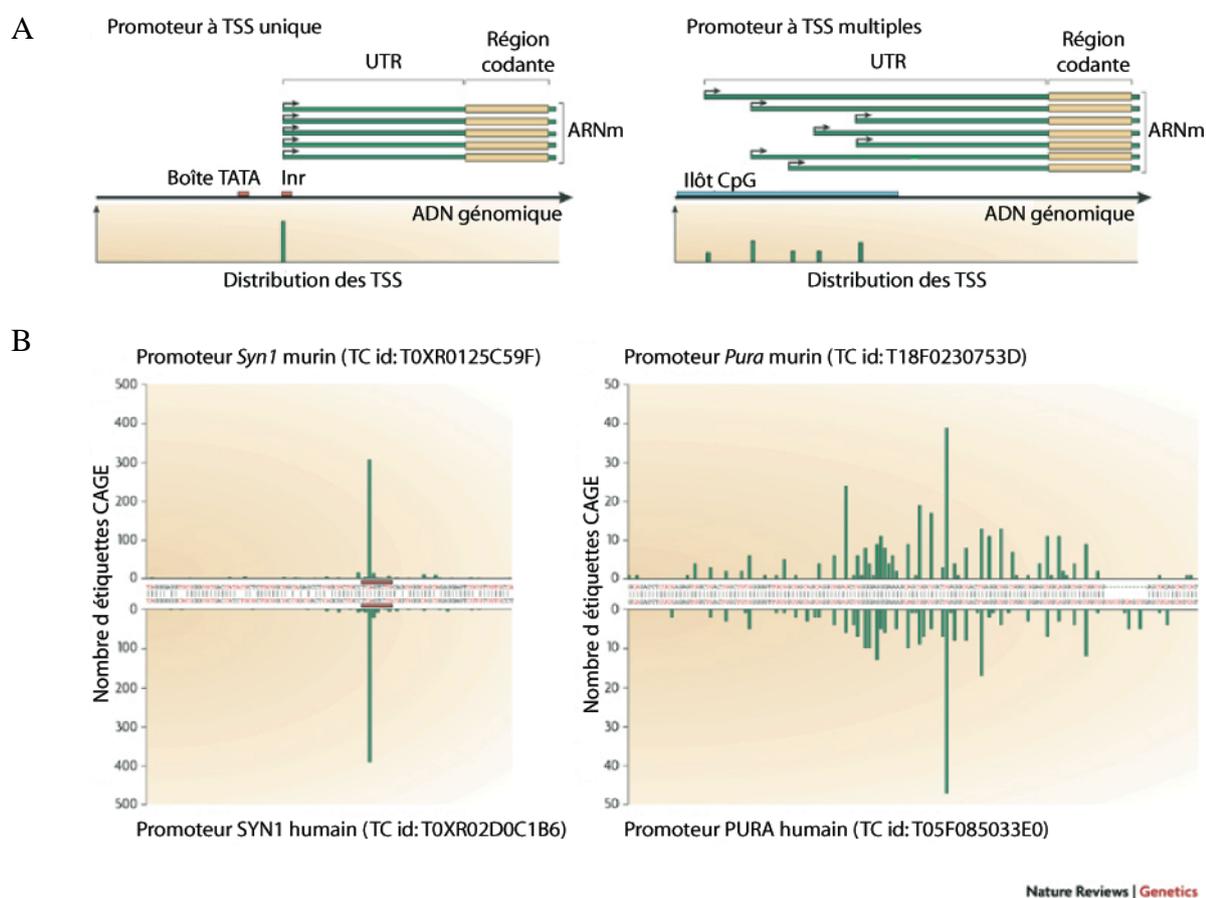


Figure 8 : Deux classes de promoteurs chez les mammifères (d'après Sandelin et al., 2007).

Les promoteurs peuvent être classés en deux catégories selon la distribution des sites d'initiation de la transcription. **A.** Représentation schématique des caractéristiques de ces deux classes. La transcription initie au niveau d'un seul site pour la classe des promoteurs à TSS unique et produit une population homogène d'ARNm. Ces promoteurs sont généralement associés à une boîte TATA et à un Inr. Les promoteurs à TSS multiples sont caractérisés par une population d'ARNm dont l'extrémité 5' est hétérogène mais codant la même protéine. Ils

sont généralement localisés dans des îlots CpG. **B.** Deux exemples illustrent les deux classes de promoteurs en se basant sur la fréquences des étiquettes obtenues avec la technique de CAGE. Le gène *Syn1* présente un promoteur avec une boîte TATA où l'initiation de la transcription se réalise principalement au niveau d'un site correspondant à un Inr. Le promoteur du gène *Pura*, dépourvu de boîte TATA, présente, quant à lui, une fenêtre d'initiation de la transcription. La distribution des TSS est similaire entre deux gènes orthologues humains et murins.

Les promoteurs à site d'initiation unique sont majoritairement associés à une boîte TATA et sont souvent impliqués dans la régulation de l'expression des gènes tissu-spécifiques. Les promoteurs à sites d'initiation multiples sont, quant à eux, majoritairement retrouvés dans des îlots CpG et associés à l'expression ubiquitaire des gènes constitutifs (Schug et al., 2005). Cependant ces observations ne sont pas toujours vraies puisqu'il existe des promoteurs dépourvus de boîte TATA régulant des gènes tissu-spécifiques et des gènes constitutifs sous le contrôle de promoteurs associés à une boîte TATA. De même, les promoteurs spécifiques du système nerveux central sont particulièrement riches en CpG. Ces deux types d'organisation des promoteurs sont conservés entre l'homme et la souris. En effet, des gènes orthologues humains et murins possèdent très fréquemment un promoteur présentant le même profil d'initiation de la transcription (Figure 8B).

Dans les deux types d'organisation, le point d'initiation de la transcription est préférentiellement une purine en position +1 (R_{+1}) avec une préférence pour une pyrimidine en position -1 (Y_{-1}). Bien que ce dinucléotide $YR_{(-1/+1)}$ corresponde à une partie de l'élément Inr, l'ensemble de la séquence consensus de cet élément est peu observée dans ces deux types de promoteurs. L'analyse à haut débit réalisée par Carninci et ses collaborateurs met ainsi en évidence que la position +1 n'est pas préférentiellement occupée par une adénine. Les dinucléotides (-1/+1) les plus employés lors de l'initiation sont CG, CA et TG (Carninci et al., 2006). Ces dinucléotides sont déterminants pour l'initiation de la transcription puisque la comparaison de TSS entre des promoteurs orthologues chez l'homme et la souris montre que le gain ou la perte de TSS sont respectivement corrélés à la création ou à l'élimination de dinucléotides YR (Gallagher et al., 2005; Carninci et al., 2006). De plus, il a été montré récemment que la mutation du dinucléotide TG, utilisé comme TSS dans le promoteur du gène de l'ankyrine, est responsable d'une diminution de l'activité de ce promoteur et empêche la liaison du facteur TFIID sur le promoteur de base (Gallagher et al., 2005).

Puisque le ribosome initie la traduction au niveau du premier ATG rencontré sur la molécule d'ARN, on peut se demander comment l'utilisation de multiples sites d'initiation de la transcription, dispersés sur une région 100 pb, ne perturbent pas la nature de la protéine codée par les nombreux ARNm ainsi produits. Une étude récente du promoteur à TSS multiples du gène PD1 a montré l'absence de codons d'initiation potentiels de la traduction dans une région de 430 pb en amont du site de transcription majeur (Lee et al., 2005). Cette analyse étendue au génome humain de souris et de rat a permis de mettre en évidence l'existence d'une zone, d'environ 1 kpb de part et d'autre de la région principale d'initiation, totalement dépourvue de triplets ATG. Ces zones ont été appelées « déserts d'ATG ». Ces déserts sont associés à des promoteurs dépourvus de boîte TATA, présentant des sites multiples d'initiation de la transcription mais localisés ou non dans des îlots CpG. Les auteurs estiment qu'environ 82% des gènes humains possèdent des déserts d'ATG dans leur promoteur (Lee et al., 2005).

La structure de la chromatine est importante pour la régulation de l'expression des gènes (pour une revue voir Li et al., 2007). L'analyse de la sensibilité du génome à la DNase I montre qu'il y a des régions où la chromatine est dans un état plus relâché que dans d'autres régions. Ces régions dites « actives », car associées à l'expression des gènes situés juste en aval, sont enrichies en histones H3 di et tri-méthylées sur la lysine 4 et en histones H3 et H4 acétylées (Pokholok et al., 2005; Birney et al., 2007). En parallèle, de récentes études montrent que la densité en nucléosomes dans les régions promotrices est moins élevée que dans les régions codantes (Bernstein et al., 2004; Lee et al., 2004a; Sekinger et al., 2005; Nishida et al., 2006). De plus, il a été observé, chez la levure, que les nucléosomes sont exclus des régions promotrices sur une région d'environ 200 pb (Yuan et al., 2005; Segal et al., 2006).

Toutes ces données tendent à montrer que la structure chromatiniennne particulière, retrouvée au niveau des promoteurs, intervient dans le recrutement de la machinerie de transcription. Dans le cas des promoteurs dépourvus de boîte TATA et d'Inr, la fixation de TFIID serait non spécifique. Le complexe d'initiation de la transcription se formerait au niveau de la chromatine active, parcourrait l'ADN jusqu'à un dinucléotide YR où l'initiation de la transcription pourrait débiter. Ce qui expliquerait la présence de multiples sites d'initiation au niveau de ces types de promoteurs (Carcamo et al., 1991; Sandelin et al., 2007).

1.3.2. Transcription permissive

Ces dernières années, de nombreuses études à haut débit sur l'expression du génome humain et murin employant des techniques basées sur l'hybridation ou le séquençage ont montré que la majorité du génome est transcrit (Kapranov et al., 2002; Okazaki et al., 2002; Bertone et al., 2004; Carninci et al., 2005; Cheng et al., 2005). Récemment, le projet ENCODE a rassemblé les données obtenues sur 1% du génome humain, provenant d'expériences de RACE, de PET ainsi que les annotations contenues dans la banque de données GENCODE. Il est apparu que 93% des bases analysées sont observées au moins deux fois comme faisant partie d'un transcrit primaire (Birney et al., 2007). Ainsi le génome est largement plus transcrit que ce que pourrait laisser penser le nombre de gènes identifiés dans le génome (23 341 gènes ; NCBI version 36). Des analyses ciblées sur les chromosomes 21 et 22 ont montré qu'environ la moitié des transcrits détectés ne colocalise pas avec une région annotée du génome (Kapranov et al., 2002; Rinn et al., 2003; Kampa et al., 2004). Ces nouveaux transcrits colocalisent soit avec des introns de gènes connus soit avec des régions intergéniques (Rinn et al., 2003). Cheng *et al* ont analysé 10 chromosomes de 8 lignées cellulaires différentes et ont montré que 31% et 26% des ARNs cytoplasmiques polyadénylés sont issus respectivement de régions intergéniques et introniques (Cheng et al., 2005). De plus, l'analyse globale de la répartition génomique des complexes d'initiation de la transcription dans des fibroblastes humains a permis d'identifier environ 10 000 promoteurs actifs dont 13% se localisent dans des régions non annotées du génome (Brodsky et al., 2005; Kim et al., 2005b).

Récemment, l'existence de petits ARNs probablement issus de la maturation de long transcrits primaires codés par des régions non annotées a été observée chez l'homme (Kapranov et al., 2007a). Une large proportion des régions non annotées transcrites sont conservées entre l'homme et la souris (Rinn et al., 2003; Bertone et al., 2004; Kapranov et al., 2007a). Ceci laisse à penser que les ARNs issus de ces régions sont potentiellement fonctionnels. Ils peuvent être à l'origine de protéines peu abondantes codées par de nouveaux gènes ou posséder une fonction propre en tant qu'ARN. Ces nouveaux ARNs non codants pourraient assurer une fonction catalytique, structurale ou de régulation. Cependant, l'existence d'un produit de transcription n'assure pas que ce transcrit possède une fonction biologique propre. En effet, la transcription d'une région en amont d'un gène peut influencer la synthèse des ARNm issus de ce gène sans que le transcrit issu de cette région amont ait une fonction

apparente. Cela a été observé, chez la levure, dans le cas du gène SER3 dont la transcription est réprimée par la synthèse de l'ARNnc SRG1 à partir d'une région chevauchant le gène SER3 (Martens et al., 2004). L'ARN polymérase, en cours de transcription, empêche les facteurs d'activation de la transcription de reconnaître le promoteur du gène SER3 (Martens et al., 2004; Schmitt et Paro, 2004). De plus, une partie de ces transcrits peut provenir d'une activité transcriptionnelle non spécifique résultant du bruit de fond de la transcription. En effet, comme tous processus biologiques, l'initiation de la transcription n'est pas absolument fidèle et peut se réaliser au niveau de promoteurs cryptiques ou dans une région où la structure de la chromatine est ouverte. Chez la levure, on estime que seulement 10 % des ARN polymérases en phase d'élongation ont initié au niveau d'un promoteur conventionnel (Struhl, 2007).

Des analyses plus approfondies sur ces différents transcrits nouvellement identifiés ont mis en évidence qu'une grande proportion d'entre eux provenaient de régions génomiques chevauchantes. Ces transcrits chevauchants peuvent être produits à partir du même brin d'ADN ou à partir des deux brins opposés. L'étude du niveau de transcription sur 30% du génome humain a révélé que 61% des loci analysés présentaient des transcrits chevauchants (Cheng et al., 2005). De nombreux nouveaux transcrits, colocalisant avec des introns de gènes annotés, sont issus de régions d'ADN positionnées dans la même orientation ou en orientation inverse par rapport au sens de transcription du gène (Rinn et al., 2003). Un exemple de cette organisation chevauchante est donné dans la Figure 9 avec le gène PISD codant la phosphatidylsérine décarboxylase. Dans ce cas, on observe neuf transcrits issus de régions localisées sur les brins sens et antisens de ce gène. Ces transcrits chevauchent à la fois des exons et des introns du pré-ARNm du gène PISD.

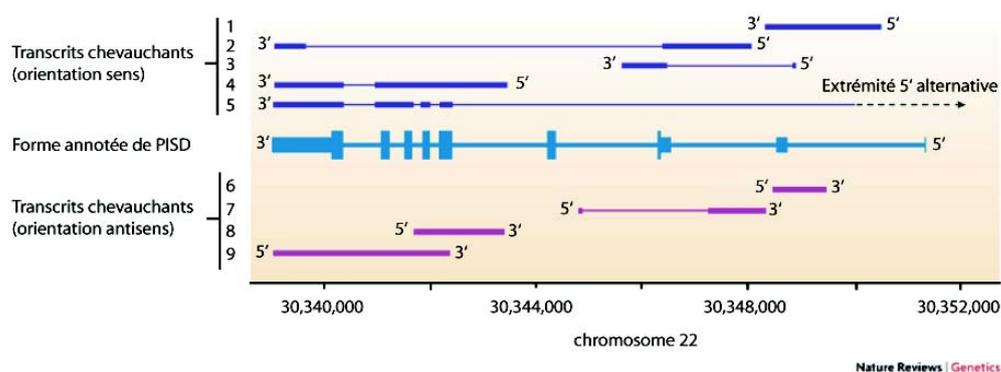


Figure 9 : Organisation chevauchante des unités de transcription (d'après Kapranov et al., 2007b).

De nombreux transcrits chevauchent la forme annotée du gène PISD, codant la phosphatidylsérine décarboxylase (en bleu clair). Les introns sont représentés par un trait fin, les exons par des rectangles épais alors que les rectangles plus fins correspondent aux régions 5' et 3'UTR de l'ARN. Cinq transcrits sont synthétisés dans la même orientation que le gène PISD (bleu foncé) alors que quatre autres transcrits colocalisent avec des régions en orientation inverse (en violet). Dans ce cas, les traits épais représentent l'information génomique retrouvée dans les molécules d'ARNs. Cette architecture a été caractérisée en combinant les résultats d'expériences de RACE avec ceux d'expériences d'hybridation sur des puces à ADN (Kapranov et al., 2005).

L'analyse du transcriptome murin montre que pour 72% des régions transcrites, il est possible d'observer également un transcrit en orientation inverse (Katayama et al., 2005). Une analyse informatique prédit l'existence de 2667 paires de transcrits sens-antisens chez l'homme (Yelin et al., 2003). Des expériences d'hybridation sur des puces à oligonucléotides et de northern-blot ont validé une majorité de ces paires (Yelin et al., 2003). La majorité des régions donnant naissance à ces paires de transcrits sens-antisens est organisée en orientation opposée en possédant des extrémités 5' chevauchantes chez l'homme et la souris tandis que ces régions sont majoritairement orientées de façon convergente chez la drosophile et *C. Elegans* (Katayama et al., 2005; Zhang et al., 2006). Ainsi les nombreux transcrits identifiés par les analyses à haut débit peuvent être fonctionnellement importants pour la cellule puisque le mécanisme de transcription sens-antisens participe à la régulation de l'expression des gènes chez les eucaryotes (pour une revue voir (Lavorgna et al., 2004). En effet, les transcrits antisens peuvent s'apparier avec le brin sens et interférer avec la transcription, l'épissage, la localisation et la stabilité de l'ARNm sens.

1.3.3. Les promoteurs alternatifs

Les promoteurs alternatifs sont des éléments importants dans la régulation de l'expression génique et ils peuvent participer à la complexité du protéome des mammifères. Un gène possède un promoteur alternatif lorsqu'il produit deux transcrits avec des extrémités 5' différentes à partir de sites d'initiation de la transcription séparés de plus de 500 pb. De nombreuses études se sont intéressées au nombre de promoteurs alternatifs présents dans le génome. Certaines études ont évalué que 20% des gènes possèdent au moins un promoteur alternatif (Landry et al., 2003; Trinklein et al., 2003; Cooper et al., 2006). L'analyse d'un nombre plus important de données permet d'estimer que plus de la moitié des gènes humains présente cette caractéristique (Carninci et al., 2006; Kimura et al., 2006; Baek et al., 2007; Tsuritani et al., 2007). Les gènes peuvent être sous le contrôle de plusieurs promoteurs alternatifs avec une moyenne de 3,1 promoteurs alternatifs par gène (Kimura et al., 2006). Par exemple, le gène codant l'UDP-glucuronyl transférase possède sept promoteurs différents. Ces types de promoteur sont majoritairement retrouvés associés à des régions pauvres en boîte TATA et en îlots CpG (Kimura et al., 2006). Ils sont essentiellement utilisés lors de la régulation de l'expression des gènes au cours du développement précoce et au sein de tissus spécifiques comme le cerveau, le cœur et le foie (Baek et al., 2007) La Figure 10 présente un résumé de la répartition des gènes renfermant des promoteurs alternatifs comparée à celle des gènes avec des promoteurs uniques.

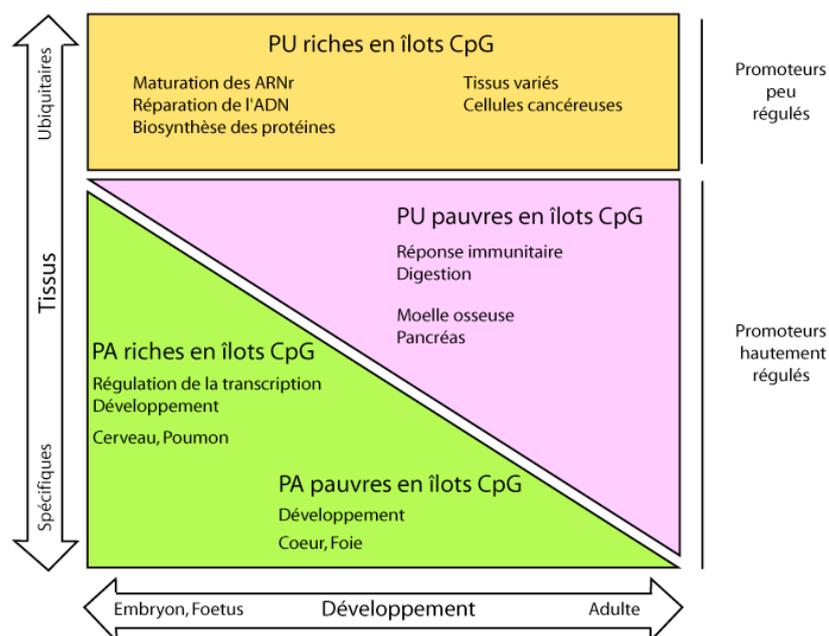


Figure 10 : Fonctions et expressions des gènes associés à des promoteurs alternatifs et à des promoteurs uniques (d'après Baek et al., 2007).

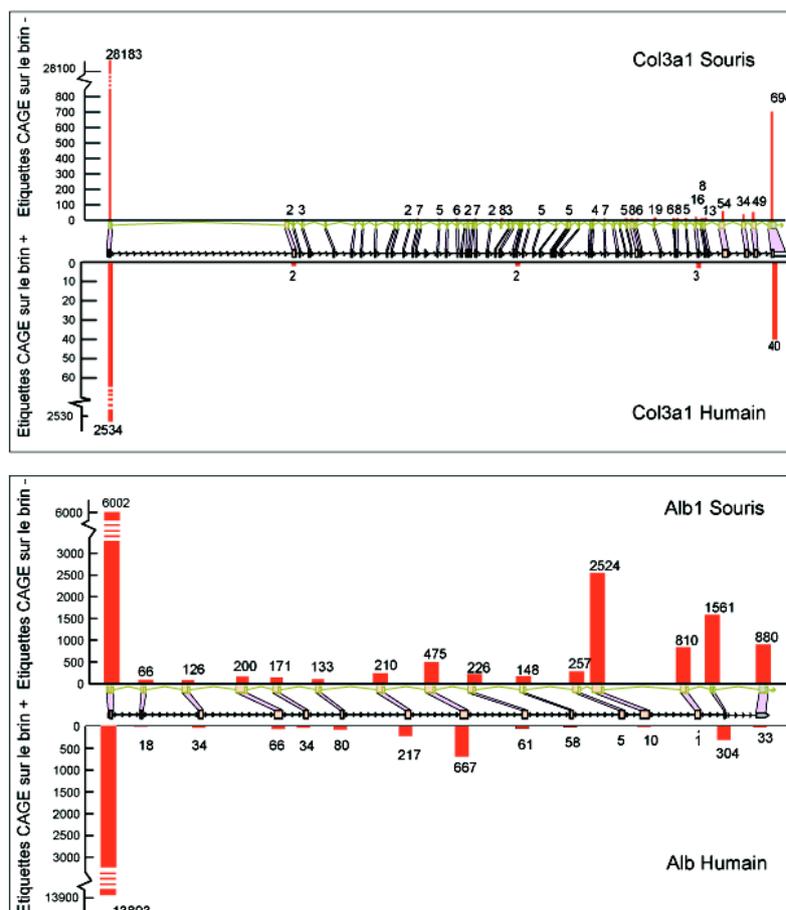
Représentation de la répartition des gènes associés à des promoteurs alternatifs (PA) ou à des promoteurs uniques (PU) en fonction du stade de développement, des tissus et de la richesse en îlots CpG de ces promoteurs.

Pour la majorité des ARNm issus des gènes à promoteurs alternatifs, les protéines synthétisées seront identiques car le même codon d'initiation de la traduction est utilisé malgré la variabilité de la région 5'UTR. Cependant, le taux d'expression de ces protéines pourra fluctuer en fonction de la force des promoteurs ou de l'efficacité de traduction qui peut varier en fonction de la nature de la région 5' de l'ARNm. Dans d'autres cas, des promoteurs alternatifs permettront la synthèse, soit de protéines isoformes avec des régions aminoterminales différentes, soit de protéines totalement différentes (pour une revue Landry et al., 2003).

Il est admis que les promoteurs alternatifs sont majoritairement localisés en 5' ou dans les premiers exons des gènes de protéines. Cependant, l'étude à haut débit de Carninci *et al* a révélé de façon surprenante la présence en nombre significatif de TSS au niveau des régions 3' non traduites des gènes (3'UTRs) (Carninci et al., 2005). Ces TSSs peuvent être considérés comme une forme extrême de promoteurs alternatifs. Contrairement aux TSS classiques

caractérisés par un dinucléotide YR_(-1/+1), un triplet de guanine est présent aux positions -2 à +1 dans 59% des TSS observés dans les régions 3'UTRs. En plaçant quatre de ces régions devant un gène rapporteur Carninci *et al* ont démontré leur capacité à diriger la transcription (Carninci et al., 2006). Les transcrits ainsi synthétisés sont dans la même orientation que ceux produits par les gènes concernés portant les régions 3'UTR étudiées. Ces transcrits pourraient réguler l'expression d'autres gènes chevauchants localisés sur le brin opposé par un mécanisme de régulation sens-antisens.

De manière tout aussi surprenante, 21% des TSS analysés par Carninci *et al* se localisent dans les exons des gènes (Carninci et al., 2005; Carninci et al., 2006). Ces TSSs peuvent être considérés comme des promoteurs alternatifs exoniques. Ces promoteurs exoniques sont majoritairement localisés dans des régions pauvres en îlots CpG et dirigent l'initiation de la transcription au niveau d'un site unique. La répartition des TSS exoniques varie selon les gènes et semble être tissu-spécifique. Par exemple, le gène ubiquitaire COL3A1 montre un niveau de TSS exonique négligeable alors que le gène ALB, exprimé presque exclusivement dans le foie, présente un taux important d'initiation exonique de la transcription (Figure 11).



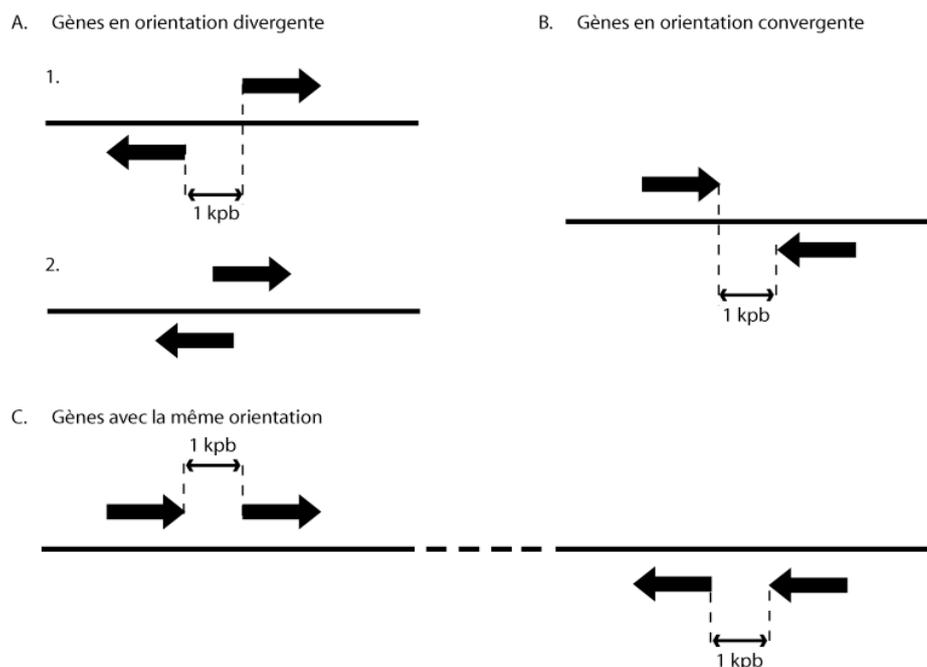


Figure 12 : Organisations de deux gènes adjacents sur le génome.

Trois organisations possibles sont représentées. Les gènes peuvent être orientés de façon divergente et partager une région promotrice commune (A.1.) ou posséder deux régions promotrices distinctes (A.2.). Ils peuvent aussi être dans une configuration convergente lorsque leurs extrémités 3' sont proches (B.) ou s'organiser dans la même orientation comme représenté en C. Les lignes continues représentent le génome, les pointillés indiquent une interruption dans le génome et les flèches montrent le sens de transcription des gènes.

Cette organisation des gènes dans le génome ne semble pas être due au hasard. En effet, de nombreuses études ont montré que les gènes sont majoritairement orientés de manière divergente (Adachi et Lieber, 2002; Takai et Jones, 2004; Trinklein et al., 2004; Carninci et al., 2006). En analysant la distance séparant les TSS de deux gènes les plus proches à partir d'une banque de données de 23 752 gènes, Trinklein *et al* ont mis en évidence que 11,8% des gènes en orientation divergente, 7,4% des gènes en orientation convergente et 7,4% des gènes dans la même orientation étaient séparés par moins de 1 kpb (Trinklein et al., 2004). De plus, l'étude des positions sur le génome de 41 118 ADNc complets humains a montré que pour les paires de gènes séparés par une région intergénique de taille inférieure à 1 kpb, 47% possèdent des orientations divergentes, 26% des orientations convergentes et 27% ont la même orientation (Koyanagi et al., 2005). Une étude ciblée sur les chromosomes 21 et 22 montre que cette organisation n'est pas liée à la densité génique (Adachi et Lieber, 2002). Cependant l'enrichissement en gènes divergents n'est observé que chez les mammifères et ceci même dans le cas de gènes ancestraux (Koyanagi et al., 2005).

Deux gènes en orientation divergente partagent une région promotrice commune, appelé promoteur bidirectionnel, lorsque la distance les séparant est inférieure à 1 kpb. Trinklein *et al* ont mis en évidence que plus de 10% des gènes humains présentent des promoteurs bidirectionnels et que 90% de ces promoteurs permettent l'initiation de la transcription dans les deux orientations (Trinklein *et al.*, 2004). Ces promoteurs bidirectionnels sont majoritairement associés à des îlots CpG et contiennent peu de boîte TATA. Il n'existe pas de motifs de régulation spécifiques à ce type de promoteurs mais il a été observé qu'un jeu restreint de facteurs de transcription était employé pour réguler les promoteurs bidirectionnels (Lin *et al.*, 2007).

Ce type de promoteur régule majoritairement l'expression de gènes mitochondriaux ainsi que de gènes impliqués dans la réparation de l'ADN ou codant des protéines chaperonnes et des hélicases à boîtes DEAD (Adachi et Lieber, 2002; Trinklein *et al.*, 2004). Certaines études ont montré que des gènes associés à un promoteur bidirectionnel étaient coexprimés (Shimada *et al.*, 1989; Ame *et al.*, 2001; Adachi et Lieber, 2002; Trinklein *et al.*, 2004; Bellizzi *et al.*, 2007). Cependant cette coexpression peut varier selon le type cellulaire analysé (Trinklein *et al.*, 2004).

Certains promoteurs bidirectionnels possèdent des éléments de régulation en commun qui sont efficaces pour la transcription de deux gènes adjacents. C'est le cas des gènes codant pour l'adénosine kinase et l'AP3m1 (Carninci *et al.*, 2006). Dans ce cas, la région promotrice centrale dirige l'initiation de la transcription de part et d'autre de cette région centrale (Figure 12.A.1). Le mécanisme dirigeant le sens de la transcription dans ce type de promoteur n'est pas connu à ce jour. D'autres promoteurs bidirectionnels sont composés de deux promoteurs distincts et indépendants, comme cela a été observé dans le cas des gènes GABPA-ATP5J (Carninci *et al.*, 2006). Dans ce type d'organisation, les sites d'initiation de la transcription d'un gène peuvent se trouver en amont des TSS du gène adjacent (Figure 12.A.2.). Ceci peut conduire à la production de paires de molécules d'ARN capables de s'associer par leur partie 5'. Ainsi la proportion importante de promoteur bidirectionnel pourrait s'expliquer par leur rôle dans la régulation complexe de l'expression des gènes. Ceci se faisant, soit par un mécanisme de transcription sens-antisens, soit par l'utilisation d'un seul promoteur régulant deux gènes voisins. Ces promoteurs bidirectionnels pourraient également jouer un rôle protecteur des gènes constitutifs vis-à-vis de l'insertion de séquences d'ADN répétées. En

effet, il a été observé une très faible proportion d'éléments répétés, comme les éléments LTR, dans les promoteurs bidirectionnels par rapport aux promoteurs de gènes divergents séparés par une distance supérieure à 1 kpb (Takai et Jones, 2004).

Ces analyses à haut débit ont mis en évidence la complexité du transcriptome des mammifères. De la même manière, nous allons aborder dans la deuxième partie de mon introduction les résultats surprenants obtenus lors de l'analyse à grande échelle des sites de fixation des facteurs de transcription régulant l'efficacité de la transcription de la Pol II chez les mammifères.

2. Régulation de l'initiation de la transcription dépendante de l'ARN polymérase II

Chez les eucaryotes, il existe plusieurs milliers de gènes présentant leur propre programme de transcription en fonction du stade de développement, du tissu et des conditions physiologiques. La régulation de l'expression de ces gènes se réalise par l'intermédiaire de différents facteurs de transcription qui se lient à la région promotrice du gène à réguler en reconnaissant des sites de fixations spécifiques. L'étude du réseau d'interaction de ces différents facteurs et de la localisation de leur site de fixation sur le génome permet de mieux comprendre le mode de régulation de l'expression des gènes. Ces études ont été majoritairement réalisées sur des promoteurs isolés. Ces expériences restent nécessaires pour l'analyse fine du rôle des facteurs de transcription au niveau d'un gène particulier. Cependant, ces dernières années, des analyses à haut débit ont permis de réaliser ces types d'études au niveau d'un génome entier. Ces analyses ont montré une large distribution des sites de fixation des facteurs de transcription dans le génome y compris, de manière surprenante, en dehors des régions situées en amont de gènes connus. Ces études ont également permis la découverte de sites de liaison fonctionnels non canoniques ainsi que l'établissement de réseaux d'interaction entre différents facteurs de transcription.

2.1. Les activateurs de la transcription

L'initiation de la transcription est dépendante de la reconnaissance de la région promotrice par les facteurs généraux de la transcription. Ces facteurs, TFIIA, TFIIB, TFIID, TFIIE, TFIIIF et TFIIH, permettent le recrutement de la Pol II au niveau du promoteur de base des gènes codant des protéines (pour une revue voir Orphanides et al., 1996). Seuls les facteurs généraux de la transcription TFIIA, TFIIB, TFIIIF, TFIIE en association avec le complexe SNAPc sont impliqués dans la reconnaissance du promoteur de base des gènes codant les ARNs et les ARNsno (pour une revue voir Hernandez, 2001). La formation du complexe de pré-initiation (PIC) au niveau de ces promoteurs, présentée dans les revues citées précédemment, ne sera pas abordée dans mon introduction. Les travaux de ma thèse ont concerné l'étude des gènes cibles de l'activateur de la transcription Staf, c'est pourquoi je souhaite me focaliser sur les facteurs participant à l'activation de la transcription.

Les activateurs de la transcription augmentent le niveau basal de la transcription en régulant, directement ou via des cofacteurs, l'activité du complexe d'initiation de la transcription en réponse aux différents stimuli cellulaires liés au stade de développement, à la nature de tissus ou aux conditions physiologiques. Leur propre activité peut être elle-même régulée par des modifications post-traductionnelles qui peuvent être par exemple une phosphorylation, une ubiquitination, une sumoylation ou une glycosylation.

Les activateurs de la transcription ont une structure modulaire composée d'un domaine d'activation de la transcription (AD) et d'un domaine de liaison à l'ADN (DBD). Le domaine de liaison à l'ADN est nécessaire mais non suffisant à l'activation de la transcription. Des expériences de fusions protéiques ont montré que la transcription peut être activée par un facteur de transcription chimère composé de l'AD d'un activateur X avec le DBD d'un facteur Y (Brent et Ptashne, 1985).

Initialement, il a été suggéré que les activateurs de la transcription interagissaient directement avec le complexe de transcription au niveau de la région promotrice (pour une revue voir Ptashne et Gann, 1997). Ensuite, étant donné la diversité des gènes à réguler et l'existence d'un répertoire limité d'éléments régulateurs en *cis* et de facteurs généraux de la transcription, il est apparu évident que la complexité de l'expression des gènes est le fruit de l'intervention de facteurs intermédiaires, nommés coactivateurs, qui relaient l'information entre les activateurs et la machinerie transcriptionnelle (pour une revue voir Lemon et Tjian, 2000).

Les activateurs peuvent recruter des coactivateurs de plusieurs types (pour une revue voir Lemon et Tjian, 2000). En effet, les activateurs peuvent, par exemple, interagir avec des coactivateurs qui sont intrinséquement associés aux composants de la machinerie transcriptionnelle comme certaines TAFs du complexe TFIID (Figure 13). Certains coactivateurs peuvent être des molécules qui jouent un rôle strict d'adaptateur entre les activateurs et le PIC. De plus, les activateurs peuvent réguler la transcription des gènes en interagissant avec des complexes comme le Médiateur composés de multiples coactivateurs de nature différente de celles des deux précédents types évoqués (Figure 13).

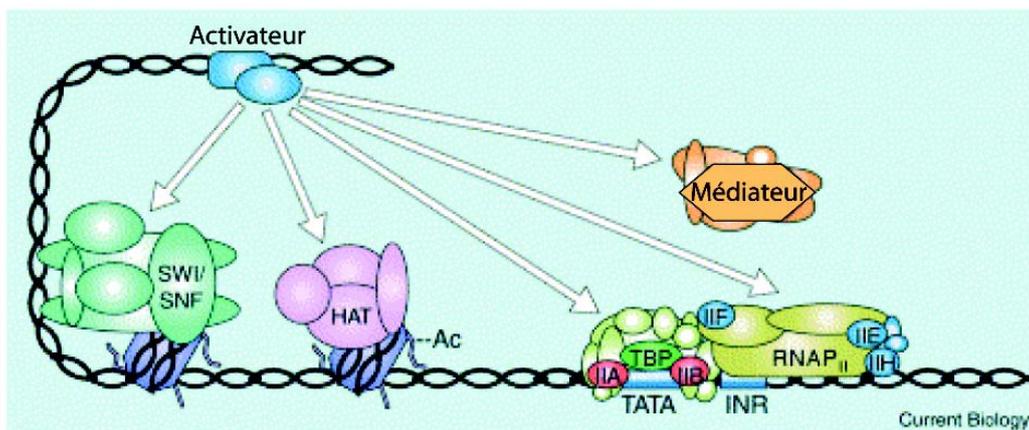


Figure 13 : Représentation schématique des modes d'actions possibles d'un activateur de la transcription (d'après Fry et Peterson, 2001).

Un activateur peut réguler la transcription en stabilisant la machinerie transcriptionnelle en interagissant avec la Pol II, les facteurs généraux de la transcription ou des coactivateurs comme le complexe du Médiateur. De plus, il peut augmenter l'accessibilité de la chromatine en interagissant avec des complexes de remodelage comme SWI/SNF ou de modifications des histones comme les histones acétyl-transférases (HAT).

D'autre part, un des moyens d'activer la transcription des gènes est d'augmenter l'accessibilité de la région promotrice en changeant la structure de la chromatine (pour une revue voir Workman et Kingston, 1998). Pour cela, les activateurs peuvent recruter des complexes de remodelage de la chromatine ou de complexes de modification des histones qui jouent le rôle de coactivateurs en induisant le relâchement de la structure chromatinienne au niveau des régions promotrices des gènes à réguler (Figure 13) (pour des revues voir Smith et Peterson, 2005; Kouzarides, 2007).

En plus de la diversité des éléments promoteurs, ces facteurs induisent un niveau de complexité supplémentaire dans le mécanisme de transcription. En effet, plusieurs activateurs de la transcription peuvent interagir avec de nombreux coactivateurs de nature différente. Cependant, il est encore difficile de montrer quelle interaction est critique pour l'activation de la transcription et de savoir si l'ordre de recrutement des différents facteurs est important pour le mécanisme (Fry et Peterson, 2001).

La communauté scientifique s'est toujours intéressée à l'étude du réseau complexe de régulation de l'expression des gènes. Pour cela, de nombreuses études ont été réalisées à l'échelle d'un gène d'intérêt. Il est maintenant possible d'appréhender l'interactome se

formant entre les promoteurs et les facteurs de transcription à l'échelle d'un génome entier. Ceci est réalisé en cartographiant les sites de fixation de facteurs de transcription le long du génome.

2.2. Cartographie des sites de fixation de facteurs de transcription

La cartographie des sites de fixation d'un facteur de transcription sur un gène donné a été traditionnellement réalisée par des expériences *in vitro* de liaison de facteurs recombinants ou par une approche *in vivo* utilisant des gènes rapporteurs. L'augmentation du nombre de génomes séquencés permet d'identifier des éléments régulateurs conservés en comparant les séquences des promoteurs orthologues. Il est également possible de localiser les sites de fixation d'un facteur de transcription au niveau d'un génome entier en utilisant des techniques novatrices couplées à l'immunoprécipitation de chromatine (ChIP).

2.2.1. Identification des sites de fixation de facteurs de transcription à l'échelle d'un gène

L'étude de la régulation de l'expression d'un gène d'intérêt se réalise en étudiant la constitution de sa région promotrice. La présence d'éléments régulateurs dans la région promotrice peut être recherchée par des analyses *in silico*, *in vitro* et *in vivo*.

Il a été observé que les régions fonctionnellement importantes du génome présentaient un plus faible taux de mutations que les régions non fonctionnelles. Il existe donc une pression de sélection pour maintenir des éléments importants du génome au cours de l'évolution. Ainsi, on peut identifier des éléments régulateurs en comparant les régions intergéniques de différentes espèces. Ceci est la base de l'analyse par empreinte phylogénique (pour une revue voir Wasserman et Sandelin, 2004). Dans ce type d'analyse, les espèces comparées ne doivent pas être trop éloignées ni trop proches d'un point de vue phylogénique afin de détecter des similitudes significatives. Dans le cas d'un gène particulier, les séquences en amont des gènes orthologues de différentes espèces sont alignées en utilisant des algorithmes tels que ClustalW (Thompson et al., 1994) ou Multalin (Corpet, 1988). Les régions conservées peuvent correspondre à des sites de fixation de facteurs de transcription. Afin de vérifier cette hypothèse et d'identifier le facteur de transcription impliqué, ces analyses sont réalisées

conjointement à l'utilisation de banques de données de facteurs de transcription telles que la banque Transfac (Wingender et al., 2000) ou l'algorithme MatInspector qui prédit la localisation de sites de fixation de facteurs de transcription connus (Quandt et al., 1995).

Le caractère fonctionnel de sites de liaison du facteur de transcription est généralement établi *in vivo* par des expériences de transfections transitoires de cellules eucaryotes. La région promotrice d'intérêt est clonée en amont d'un gène rapporteur. L'influence de mutations au niveau de sites de fixation potentiels est observée par la variation de la synthèse du produit du gène rapporteur. Ceci permet de déterminer les éléments importants dans la régulation du gène d'intérêt.

Des expériences *in vitro* peuvent également permettre de mettre en évidence la présence de sites de fixation de facteurs de transcription. Un extrait nucléaire peut être utilisé lors d'expériences de retard sur gel réalisées avec un fragment d'ADN renfermant un motif régulateur potentiel conservé identifié *in silico* ou préalablement détecté par des expériences *in vivo*. L'observation d'un complexe retardé suggère qu'un facteur de transcription se lie à la séquence d'ADN étudiée.

2.2.2. Identification des sites de fixation de facteurs de transcription à l'échelle d'un génome

Cette approche est basée sur la sélection des régions d'ADN associées à un facteur de transcription donné par immunoprécipitation de chromatine couplée à une analyse à grande échelle de ces fragments. Cette analyse globale utilise des approches dérivées de celles développées pour l'analyse à haut débit du transcriptome. Ainsi elle peut être basée sur l'hybridation des fragments immunoprécipités sur une puce (ChIP on chip) ou sur le séquençage d'une courte séquence du fragment d'ADN immunoprécipité (ChIP-SAGE, ChIP-Seq) ou de leurs extrémités 5' et 3' (ChIP-PET) (Figure 14).

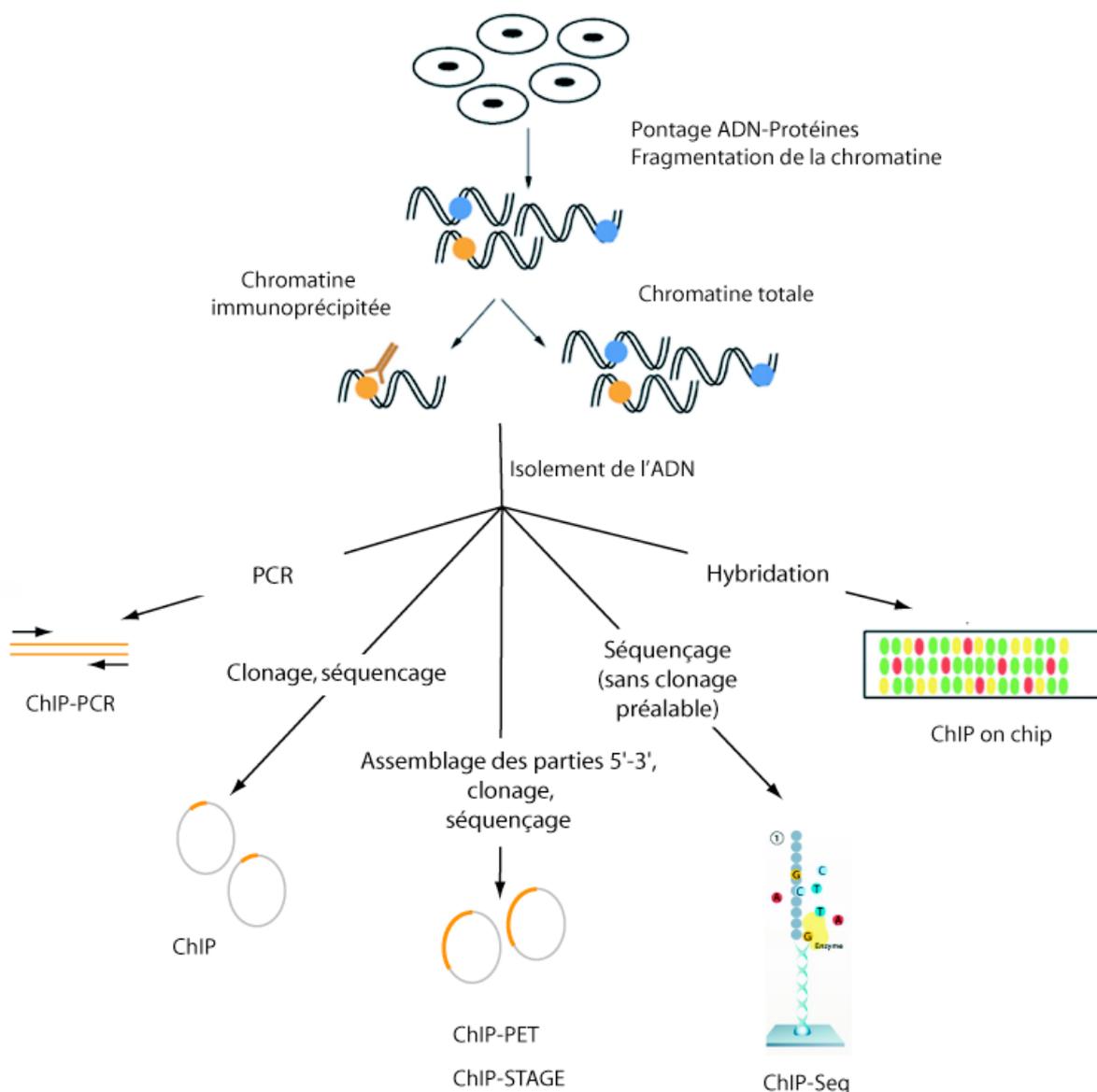


Figure 14 : Différentes stratégies d'identification des sites de fixation d'un facteur de transcription basées sur l'immunoprécipitation de chromatine (adapté d'après Hawkins et Ren, 2006).

La chromatine est fragmentée après pontage des protéines à l'ADN. Les fragments obtenus sont immunoprécipités avec des anticorps dirigés contre un facteur de transcription spécifique. L'ADN est isolé et analysé selon différentes approches. Pour la technique de ChIP-PCR, les régions d'intérêt sont recherchées par PCR dans l'ADN immunoprécipité en utilisant des couples d'oligonucléotides spécifiques. Les fragments sont clonés puis séquencés dans le cas de l'approche classique du ChIP (Weinmann et al., 2001). Dans le cas des approches de ChIP-PET et ChIP-STAGE, les parties 5' et/ou 3' issues de différents fragments immunoprécipités sont associées, clonées puis séquencées (Kim et al., 2005a; Wei et al., 2006). La technique de ChIP-Seq permet d'analyser directement les fragments d'ADN immunoprécipités par séquençage à haut débit sans effectuer de clonage (Robertson et al., 2007). Enfin, l'analyse par ChIP on chip est réalisée en hybridant les fragments immunoprécipités sur des puces (Weinmann et al., 2002; Cawley et al., 2004; Lee et al., 2006).

Actuellement, l'immunoprécipitation de chromatine est une technique de choix pour mettre en évidence l'interaction d'un facteur de transcription au niveau d'une région d'ADN d'intérêt *in vivo*. Dans cette approche, les facteurs de transcription sont pontés à la chromatine par un traitement au formaldéhyde puis après fragmentation de la chromatine, les fragments liant le facteur d'intérêt sont immunoprécipités par des anticorps spécifiques. Après dénaturation des protéines, les fragments d'ADN peuvent être caractérisés par différentes approches.

Dans le cas du ChIP-PCR, la région promotrice d'intérêt est recherchée par PCR avec un couple d'oligonucléotides amplifiant cette région promotrice (Figure 14). Les résultats obtenus sont comparés avec ceux provenant d'un échantillon immunoprécipité avec des anticorps issus d'un sérum pré-immun.

L'analyse des fragments immunoprécipités peut être également réalisée en hybridant ces fragments sur des puces (Figure 14). Cette approche, nommée CHIP on chip, a débuté par l'utilisation de puces contenant des régions promotrices (Ren et al., 2002) ou des îlots CpG (Weinmann et al., 2002). Cependant cette approche ne couvre qu'un petit échantillon de régions potentiellement régulatrices. Les analyses suivantes se sont basées sur des puces couvrant des chromosomes entiers (Martone et al., 2003; Cawley et al., 2004). Puis les puces représentant les séquences non redondantes du génome humain sont apparues permettant alors d'analyser la distribution globale des sites de liaison d'un facteur de transcription (Lee et al., 2006). Les fragments immunoprécipités, ainsi que les fragments issus de l'échantillon global, sont marqués par deux fluorochromes différents et hybridés sur la puce. L'enrichissement en une région d'intérêt est détecté en comparant les intensités des deux fluorochromes pour chaque point. Ce type de technique est plus facilement applicable aux études répétées dans différentes conditions physiologiques.

Il est possible de cloner les fragments d'ADN immunoprécipités et d'identifier de nouveaux sites de fixation d'un facteur de transcription d'intérêt en séquençant ces fragments (Figure 14) (Weinmann et al., 2001). Les gènes potentiellement régulés par ces sites seront identifiés en localisant les fragments obtenus sur la séquence génomique. Cependant pour appliquer cette approche à l'échelle d'un génome entier, il est nécessaire que l'ensemble des fragments d'ADN immunoprécipités soient séquencés.

Des améliorations ont été apportées afin d'obtenir lors du séquençage une bonne représentativité de l'ensemble des fragments d'origine. La technique de STAGE (Sequence Tag Analysis of Genomic Enrichment), dérivant de la technique de SAGE, permet d'analyser rapidement de nombreux fragments de 21 pb représentatifs des fragments immunoprécipités par ChIP (Kim et al., 2005a). De même, l'approche couplant les techniques de ChIP et de PET permet cette analyse à partir des extrémités 5' et 3' de ces fragments (Figure 14) (Wei et al., 2006). Ces techniques sont rapidement applicables à l'analyse de tout organisme dont le génome est séquencé. De plus, ces approches permettent d'obtenir en une seule réaction de séquence des informations sur plusieurs fragments immunoprécipités.

Très récemment, une approche basée sur une technique innovante de séquençage a été développée et a permis de déterminer rapidement la séquence des fragments immunoprécipités. Cette technique a été nommée ChIP-Seq (Figure 14) (Robertson et al., 2007). Les fragments immunoprécipités sont fixés par l'intermédiaire d'un adaptateur sur une puce et sont amplifiés en phase solide. Chaque échantillon est séquencé en utilisant la technique de séquençage Solexa (<http://www.illumina.com/pages.ilmn?ID=203>). Cette technique est basée sur l'utilisation de quatre dNTP terminateurs marqués avec un fluorochrome différent. La lecture de la plaque est réalisée après excitation par un laser. Une courte séquence est obtenue pour chaque fragment après plusieurs cycles d'addition des dNTPs et est localisée sur la séquence génomique afin de permettre l'identification du site de fixation du facteur de transcription. Comme dans les autres approches, l'enrichissement en une région est évalué en analysant l'ADN immunoprécipité avec des anticorps issus d'un sérum pré-immun. Cette approche est intéressante car elle utilise peu de matériel et ne nécessite pas la construction de banques comme dans le cas du ChIP-PET ou du ChIP-STAGE. De plus, elle apparaît moins coûteuse que la technique de ChIP on chip et peut être appliquée rapidement à l'étude de génome d'organisme variés dès que leur génome est séquencé.

Des études ont comparé les résultats obtenus avec ces différentes techniques afin d'évaluer leur efficacité (Ji et al., 2006; Euskirchen et al., 2007). Il apparaît qu'une majorité de sites communs est retrouvée avec les différentes approches. Cependant, certains sites ne sont identifiés que par l'une de ces techniques. Ainsi ces différentes approches s'avèrent complémentaires pour la construction d'une carte globale des sites de fixation d'un facteur de transcription.

2.3. Large distribution des sites de fixation des facteurs de transcription dans le génome

Ces dernières années de nombreuses équipes se sont intéressées à la cartographie globale des sites de liaison d'un certain nombre de facteurs de transcription avec pour objectif futur de mieux comprendre la régulation de l'expression des gènes (Tableau 3). Ces analyses ont été réalisées pour différents facteurs de transcription sur différents types cellulaires ou différents organismes. Cependant, tous les résultats obtenus montrent que la localisation de ces sites ne se limite pas aux régions situées juste en amont de gènes identifiés. Ceci suggère que le génome pourrait être transcrit plus abondamment que ce qui était admis jusqu'à présent.

Approches	Facteurs	Organismes	Types cellulaires Tissu	Références
ChIP-chip				
Chr22	p65 (NF-κB)	Homme	HeLa	(Martone et al., 2003)
Chr21 et 22	Sp1, c-Myc, p53	Homme	Lymphocytes T	(Cawley et al., 2004)
Chr22	CREB	Homme	Placenta	(Euskirchen et al., 2004)
Promoteur	Oct4, Sox2, Nanog	Homme	ES	(Boyer et al., 2005)
Ilots CpG	NF-Y	Homme	Foie, Leucocytes	(Testa et al., 2005)
Génome	SUZ12	Homme	ES	(Lee et al., 2006)
ChIP-SAGE	CREB	Rat	Rein	(Impey et al., 2004)
	E2F4	Homme	Fibroblastes	(Kim et al., 2005a)
ChIP-PET	Oct4, Nanog	Souris	ES	(Loh et al., 2006)
	p53	Homme	Colon	(Wei et al., 2006)
ChIP-Seq	NRSF	Homme	Lymphocytes T	(Johnson et al., 2007)
	STAT1	Homme	HeLa	(Robertson et al., 2007)

Tableau 3 : Exemples de facteurs de transcription dont les sites de liaison ont été identifiés par des analyses à haut débit.

Des approches variées ont permis d'analyser globalement la répartition des sites de fixation d'un certain nombre de facteurs de transcription. Ces analyses ont été réalisées sur différents types cellulaires de trois organismes différents.

2.3.1. Localisation classique des sites de liaison de facteurs de transcription

Ces approches ont confirmé la présence de sites de fixation identifiés lors de précédentes études et ont permis de localiser de nouveaux sites dans les régions situées en amont de gènes

connus. Cependant, les proportions sont variables selon le facteur de transcription analysé. Par exemple, 95% des sites de liaison du facteur SUZ12 (une sous-unité du complexe PRC₂) ont été observés dans une région de 1 kpb autour de TSS connus ou prédits (Lee et al., 2006). 40% des sites de fixation du facteur CREB et 22% des sites pour les facteurs Sp1, cMyc et p53 ont été identifiés à proximité de gènes connus (Cawley et al., 2004; Impey et al., 2004).

Ces approches à haut débit ont également permis de caractériser de nouveaux sites de liaison de facteurs de transcription. Ainsi, parmi les 542 sites de fixation de p53 identifiés, 98 représentent de nouveaux sites de liaison (Wei et al., 2006). Certains des sites de fixation identifiés ont une séquence non canonique (Martone et al., 2003; Johnson et al., 2007; Wang et al., 2007). Dans le cas du facteur AR (Androgen Receptor), la séquence canonique du site de liaison est AGAACAn₃TGTTCT. Cependant, 68% des sites identifiés correspondent soit au demi-site de fixation TGTTCT soit à une nouvelle architecture du site de fixation (AGAACAn₀₋₈TGTTCT ; TGTTCTn₀₋₈AGAACA ; AGAACAn₀₋₈AGAACA) (Wang et al., 2007). Une analyse plus précise des résultats obtenus par ces études globales a montré un enrichissement de certains sites de fixation dans des régions promotrices particulières. Par exemple, 40% des sites de fixation du facteur SUZ12 sont localisés dans des îlots CpG (Lee et al., 2006). De même, la distribution des sites de fixation du facteur CREB est hautement corrélée à la présence d'îlots CpG et 25% des sites de fixation de ce facteur font partie de promoteurs bidirectionnels (Impey et al., 2004). Impey *et al* ont également mis en évidence la présence de sites de fixation de CREB en aval de certains TSS précédemment identifiés, ceci laisse supposer l'existence probable de promoteurs alternatifs pour ces gènes.

2.3.2. Répartition inattendue des sites de fixation de facteurs de transcription le long du génome

De façon surprenante, de nombreux sites de fixation se localisent en dehors des régions situées juste en amont de gènes connus. En effet, de nombreux sites sont retrouvés à l'intérieur des gènes et dans des régions non annotées du génome (Martone et al., 2003; Euskirchen et al., 2004; Impey et al., 2004; Loh et al., 2006; Wei et al., 2006; Robertson et al., 2007). Il apparaît qu'une proportion non négligeable (environ 35% en moyenne) de sites de liaison est localisée dans les introns. De plus, on observe un plus grand nombre de sites dans le premier intron des gènes que dans les autres introns. Par exemple, 12% des sites p65 sont

localisés dans le premier intron tandis que 28% sont situés dans l'ensemble des autres introns (Martone et al., 2003).

Peu de sites sont retrouvés au niveau des exons des gènes. Les différentes études montrent que seulement environ 2% de sites de fixation sont présents dans ces structures. On peut même remarquer qu'apparemment aucun site de liaison du facteur p53 n'est présent dans les exons des gènes (Wei et al., 2006).

Des sites de fixation des facteurs de transcription sont également identifiés dans les régions à proximité de l'extrémité 3' de gènes connus. Ceci a été observé pour 36% des sites Sp1, cMyc et p53 identifiés sur les chromosomes 21 et 22 (Cawley et al., 2004). Dans 27% des cas, les sites identifiés dans la région 3' d'un gène sont également situés dans la région 5' d'un autre gène adjacent (Cawley et al., 2004). Une grande majorité des sites identifiés dans cette étude est également localisée en amont de régions transcrites nouvellement caractérisées par Kapranov *et al* (Kapranov et al., 2002).

Des sites de fixations sont également retrouvés dans des régions non annotées du génome et ceci dans des proportions variables selon les études. Dans le cas du facteur p65, 28% des sites de liaison identifiés remplissent cette caractéristique. Parmi ces sites, 22% se situent à proximité des régions transcriptionnellement actives identifiées par Rinn *et al* et les autres sont localisés dans des régions du génome actuellement sans aucune annotation (Martone et al., 2003; Rinn et al., 2003).

Ainsi la répartition des sites de fixation des facteurs de transcription dans le génome est plus large qu'attendue. On peut se demander si les sites localisés à des positions non canoniques sont fonctionnels. La présence de sites de fixation en amont de régions transcriptionnellement actives, identifiées lors d'études du transcriptome, semble indiquer qu'une large proportion de ces sites sont fonctionnels. Les sites de fixation localisés dans la région 3' de gènes connus peuvent participer à la régulation de l'expression de gènes voisins ou chevauchants. Cawley *et al* ont vérifié par RT-PCR et northern-blot l'existence de transcrits dont la transcription est potentiellement régulée par les facteurs Sp1, cMyc et p53 et pour lesquels les sites de liaison sont localisés dans la région 3' de gènes connus (Cawley et al., 2004).

La mise en évidence de la présence de la Pol II au niveau des régions renfermant certains sites de fixation est un argument en faveur du possible caractère fonctionnel de ces sites. Par

exemple, des expériences de ChIP ont montré que la Pol II est associée à 50% des régions renfermant des sites de liaison du facteur AR (Wang et al., 2007). Il a déjà été établi, par d'autres études, que des facteurs de transcription peuvent également participer à la régulation de l'expression d'un gène à longue distance. Cela a été observé en particulier pour des sites de fixation de p65 localisés dans des régions introniques (Schjerven et al., 2001; Guo et al., 2003). L'analyse comparative, entre différentes espèces, des régions renfermant de tels sites et la mise en évidence de la conservation de ces motifs peuvent montrer l'importance de ces régions et renforcer leur caractère fonctionnel probable.

Il est néanmoins possible que certains sites de fixation qui apparaissent atypiques soient non fonctionnels. Les facteurs de transcription peuvent se lier au niveau de sites de fixation isolés apparus suite à une mutation mais ne pas permettre le recrutement de la machinerie transcriptionnelle.

2.3.3. Détermination de réseaux de régulation

La cartographie des sites de fixation des facteurs de transcription à l'échelle des génomes est requise pour l'étude des réseaux de régulation se mettant en place entre ces différents facteurs et dans différentes conditions physiologiques. Les régions régulatrices sont généralement composées d'un regroupement de différents éléments régulateurs. Certaines études se sont donc intéressées à l'identification d'un site de liaison pour un facteur X à proximité des sites de fixation d'un facteur de transcription d'intérêt (Martone et al., 2003; Cawley et al., 2004; Boyer et al., 2005; Wang et al., 2007). C'est ainsi que Wang *et al* ont identifié trois motifs à proximité de sites de liaison du facteur AR (Wang et al., 2007). Des expériences de ChIP-PCR ont montré que les facteurs FoxA1, GATA2 et Oct1 se lient à 80% des régions renfermant également un site de liaison du facteur AR (Wang et al., 2007). De multiples expériences ont montré que ces facteurs, à l'exception de FoxA1, jouent un rôle important, en collaboration avec le facteur AR, dans la régulation de l'expression des gènes en réponse à des hormones androgènes (Wang et al., 2007). Un réseau de régulation peut s'établir via les interactions se formant entre divers facteurs de transcription mais peut également dépendre du produit des gènes régulés. Différentes études ont illustré ce dernier point (Euskirchen et al., 2004; Boyer et al., 2005; Lee et al., 2006; Loh et al., 2006; Johnson et al., 2007). Deux d'entre elles ont montré que les facteurs de transcription Oct4 et Nanog permettent de maintenir les

cellules ES (Embryonic Stem) dans un état pluripotent en reprimant des gènes codant des facteurs de transcription clés dans le mécanisme de différenciation et en activant des gènes essentiels au maintien de l'état indifférencié (Boyer et al., 2005; Loh et al., 2006).

On peut penser que ces études globales de la répartition des sites de liaison de facteurs de transcription vont largement se multiplier dans les années à venir. Les données obtenues lors de ces études fourniront une base pour établir et compléter des réseaux d'interaction des facteurs de transcription. Ceci nous permettra de mieux comprendre les mécanismes qui régissent le bon fonctionnement d'une cellule via un niveau de transcription des gènes finement contrôlé.

Les deux premières parties de ce mémoire avaient pour objectifs de mettre en évidence la grande complexité au niveau de l'expression des génomes de mammifères. Ainsi, le génome n'est pas constitué de gènes indépendants répartis le long des chromosomes. En effet, il apparaît que de nombreux gènes peuvent se chevaucher, partager des promoteurs bidirectionnels ou être régulés par des promoteurs alternatifs. Les analyses du transcriptome ainsi que les cartographies des sites de liaison de facteurs de transcription montrent que le génome est plus transcrit qu'attendu. Cette grande complexité d'expression peut expliquer que des organismes eucaryotes, tels que la levure et la drosophile, possédant environ le même nombre de gènes que les mammifères présentent de grandes différences de phénotypes avec ces derniers.

Les travaux effectués au cours de ma thèse ont permis de caractériser de nouveaux gènes cibles du facteur de transcription Staf. Ces travaux se sont articulés autour d'une identification globale des sites de fixation de ce facteur dans les promoteurs des gènes de protéines ainsi que sur des analyses plus fines de promoteurs de trois gènes spécifiques. Nous allons dans le prochain chapitre nous intéresser plus particulièrement aux caractéristiques et aux fonctions de ce facteur de transcription.

3. Staf, un activateur de la transcription à multiples facettes

Le facteur de transcription Staf a été identifié au laboratoire comme étant l'activateur de la transcription du gène de l'ARN de transfert de la sélénocystéine (ARNt^{Sec}) chez *Xenopus laevis*. Il a été ensuite montré que ce facteur était également impliqué dans l'activation de la transcription, par l'ARN polymérase II ou III, d'une variété de gènes codant des ARNc tels que les ARNsn U1-U5 ainsi que les ARNs 7SK, Y, de la MRP et de la RNase P. De façon surprenante, Staf peut également stimuler la transcription de gènes codant des protéines. Nous allons donc nous intéresser à l'architecture des promoteurs régulés par ce facteur ainsi qu'aux caractéristiques de la protéine Staf qui permettent une telle diversité de reconnaissance de gènes cibles.

3.1. Activateur de la transcription du gène de l'ARN de transfert de la sélénocystéine

Le promoteur du gène de l'ARNt^{Sec} et les promoteurs des gènes codant les autres ARNt cellulaires ne présentent pas la même architecture bien qu'ils soient tous deux reconnus par la Pol III. Les promoteurs des gènes codant pour les ARNt généraux sont constitués d'une boîte A et d'une boîte B localisées dans la région transcrite du gène (Figure 15).

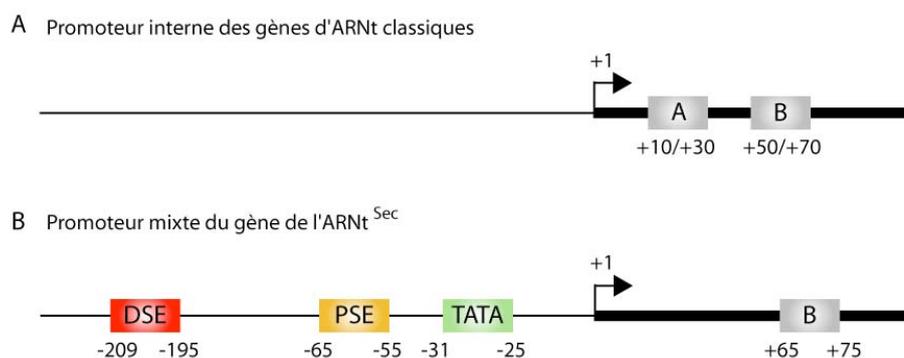


Figure 15 : Architecture des promoteurs des gènes des ARNs de transfert chez les vertébrés.

Les positions des éléments du promoteur de base (A : boîte A, B : boîte B, PSE : proximal sequence element, TATA : boîte TATA) et de l'élément activateur (DSE : distal sequence element) sont indiquées. La région transcrite est représentée par un trait épais et la flèche indique le site d'initiation de la transcription. Les positions indiquées pour l'ARNt^{Sec} sont celles de l'ARNt^{Sec} de *Xenopus laevis*.

Le promoteur de base du gène de l'ARNt^{Sec} est, quant à lui, un promoteur mixte. Il est composé d'une boîte interne B et de deux éléments externes : une boîte TATA et un élément PSE (Proximal Sequence Element) (Carbon et Krol, 1991). L'activité optimale du gène de l'ARNt^{Sec} est dépendante d'un élément activateur appelé DSE (Distal Sequence Element) (Myslinski et al., 1992). Les caractéristiques de ces éléments promoteurs, DSE, PSE et boîte TATA, seront développées dans la partie suivante de mon manuscrit.

Par la suite, il a été montré qu'une protéine présente dans des extraits nucléaires de cellules HeLa reconnaît spécifiquement l'élément activateur DSE (Myslinski et al., 1992). Le criblage d'une banque d'expression avec un oligonucléotide portant la séquence de cet élément (CCAGCATGCCTCGCG) a permis d'identifier l'ADNc codant cette protéine (Schuster et al., 1995). Des expériences de micro-injection dans des oocytes de xénopes ont démontré que cette protéine était l'activateur de la transcription du gène de l'ARNt^{Sec} (Schuster et al., 1995). Pour cela, un vecteur spécifiant une protéine chimère a été préparé en échangeant le DBD de la protéine isolée avec celui de la protéine Krox-20. L'expression de cette protéine chimère permet d'activer la transcription d'un gène d'ARNt^{Sec} muté qui renferme la séquence de reconnaissance du facteur de transcription Krox-20.

La protéine identifiée, permettant l'activation de la transcription du gène de l'ARNt^{Sec}, a été nommée Staf (pour **S**elenocysteine **t**RNA gene transcription **a**ctivating **f**actor). Staf est une protéine de 600 acides aminés composée de trois domaines (Figure 16). Le domaine d'activation est situé dans la partie amino-terminale et renferme deux sous-domaines. Le premier est composé de quatre répétitions d'un motif de quinze acides aminés. Le deuxième correspond à une région de dix-huit acides aminés. Le domaine central contient le domaine de liaison à l'ADN composé de sept doigts de zinc de type C₂H₂. Aucune fonction n'a été attribuée à la partie carboxy-terminale de la protéine à ce jour (Schuster et al., 1995; Schuster et al., 1998).

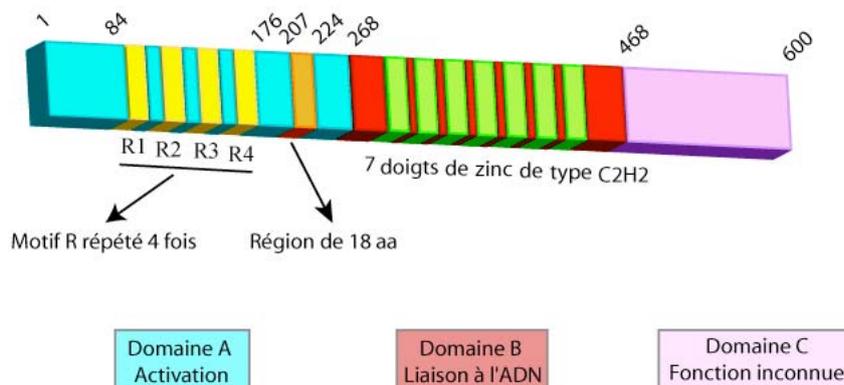


Figure 16 : Représentation schématique des domaines structuraux et fonctionnels de l'activateur de la transcription Staf (Schuster et al., 1995; Schuster et al., 1998).

Le domaine A contenant les deux sous-domaines d'activation (en jaune et orange) est représenté en bleu. Le domaine B de liaison à l'ADN est indiqué en rouge tandis que les sept doigts de zinc sont représentés en vert. Le domaine C de fonction inconnue est coloré en rose. Les positions de chaque région sont indiquées.

3.2. Activateur de la transcription de gènes d'ARNs non codants et d'ARNs messagers

Chez les vertébrés, les gènes codant les ARNsnc peuvent être transcrits par les ARN polymérase II ou III selon l'architecture de leur promoteur (Figure 17) (pour une revue voir Hernandez, 2001)). D'autres gènes de petits ARNnc, comme les gènes codant les ARNsno U3, U8 et U13 ou l'ARN 7SK et Y4 sont respectivement transcrits par la Pol II ou la Pol III (Figure 17). Ils possèdent les mêmes éléments promoteurs que ceux qui sont présents dans les gènes d'ARNsn (pour des revues voir Paule et White, 2000; Hernandez, 2001; Schramm et Hernandez, 2002). Ces gènes seront appelés gènes de type ARNsnc dans la suite de mon manuscrit.

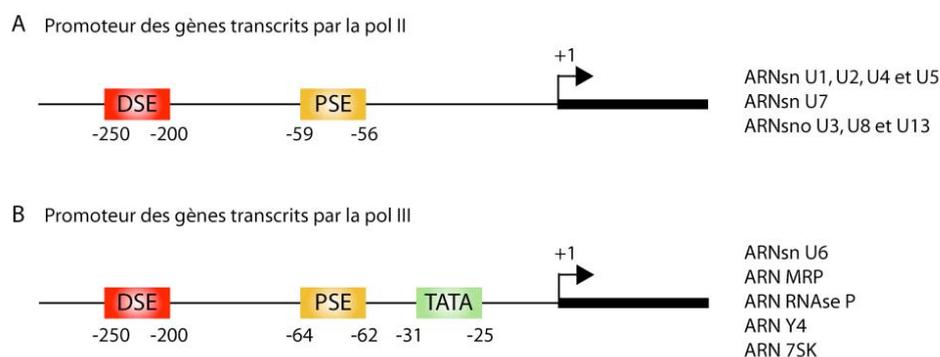


Figure 17 : Architecture des promoteurs des gènes d'ARNsn et de type ARNsn transcrits par la Pol II (A) et la Pol III (B).

Les positions des éléments du promoteur de base (PSE : proximal sequence element, TATA : boîte TATA) et de l'élément activateur (DSE : distal sequence element) sont indiquées. La région transcrite est représentée par un trait épais et la flèche indique le site d'initiation de la transcription. Les gènes contrôlés par chaque type de promoteurs sont indiqués.

Les promoteurs de ces gènes sont totalement externes à la région transcrite et possèdent deux éléments communs : un élément activateur distal, la DSE et un élément proximal, constituant du promoteur de base, la PSE (Figure 17) (Murphy et al., 1987; Dahlberg et Schenborn, 1988; Hernandez et Lucito, 1988).

La DSE est localisée entre les positions -250/-200 dans les promoteurs Pol II et Pol III (Lobo et Hernandez, 1994). La DSE n'a pas de rôle dans la spécificité de transcription par la Pol II ou III et est interchangeable d'un type de promoteur à l'autre (Ciliberto et al., 1987). Cet élément est capable de stimuler la transcription du promoteur de base d'un facteur cinq à vingt quelque soit son orientation (Ares et al., 1985; Ciliberto et al., 1985; Mattaj et al., 1985; McNamara et al., 1987). L'activité de la DSE est généralement dépendante de sa position au sein de la région promotrice puisque cet élément est inactif lorsqu'il se trouve trop proche de la PSE ou en aval de la région codante (Mattaj et al., 1985). Néanmoins, il a été observé que les promoteurs des gènes de l'ARN H1 humain, de l'ARN 7SK du fugu et de l'ARNsn U7 de xénope présentent une architecture compacte inhabituelle dans laquelle la DSE est adjacente à la PSE (Watkins et al., 1992; Myslinski et al., 2001; Myslinski et al., 2004).

La DSE renferme deux motifs régulateurs : un motif octamère (ATGCAAAT) ou une séquence très proche appelée octamer-like et un site de liaison de Staf (Parslow et al., 1984; Carbon et al., 1987; McNamara et al., 1987; Tebb et Mattaj, 1989; Roebuck et al., 1990; Myslinski et al., 1992; Lobo et Hernandez, 1994; Schuster et al., 1995; Schaub et al., 1997). Ces éléments peuvent être présents soit simultanément soit séparément. Ils sont présents simultanément dans quasiment 100%, 80% et 46% des cas chez les mammifères, le poulet et le xénope respectivement (Hernandez, 1992; Schaub et al., 1997; Gu et al., 1998; Schramm et Hernandez, 2002). Chez le fugu, cette proportion s'inverse puisque parmi les 28 gènes d'ARNsn identifiés comme possédant une DSE, seuls les gènes codant l'ARNsn U4.1 et U7 possèdent une DSE composée d'un site de liaison de Staf et d'un octamère (Myslinski et al., 2004). De façon surprenante, sept sur trente-cinq des gènes d'ARNsn identifiés dans cette espèce ne présentent pas d'élément DSE. Il s'agit des gènes codant les ARNsn U1.3, U2.2, U2.4, U4.3, U4.4, U5.4 et l'ARNt^{Sec} (Myslinski et al., 2004). Ces gènes, à l'exception du gène de l'ARNt^{Sec}, sont présents en copies multiples dans le génome du fugu et compensent sûrement l'absence de l'élément activateur par leur nombre de copies. Le gène de l'ARNt^{Sec}, quant à lui, peut également compenser cette absence par la présence de l'élément promoteur interne B. En effet, il a été montré, chez le xénope, que cet élément permettait de stimuler la transcription par la Pol III du gène de l'ARNt^{Sec} et d'un gène d'ARNsn chimère dans lequel cet élément B a été introduit artificiellement (Carbon et Krol, 1991; Lescure et al., 1991). On peut noter que le motif octamère n'est pas spécifique du promoteur des gènes d'ARNsn et de type ARNsn. En effet, il est également retrouvé dans les régions activatrices des promoteurs de gènes d'ARNm exprimés de façon ubiquitaire ou tissu-spécifique (Parslow et al., 1984; Fletcher et al., 1987; Luo et al., 1992).

L'élément promoteur PSE du gène de l'ARNsn U6 ainsi que des gènes de type ARNsn transcrits par la Pol III (ARN 7SK, MRP, Y...) possède les mêmes caractéristiques structurales que la PSE des promoteurs des gènes transcrits par la Pol II décrites au paragraphe 1.1.2 de la première partie de mon introduction (Murphy et al., 1986; Krol et al., 1987; Bennett et al., 1992; Hernandez, 1992; Lobo et Hernandez, 1994; Maraia et al., 1994). Ces éléments PSE sont fonctionnellement interchangeable entre les promoteurs des gènes pol II et pol III (Bark et al., 1987; Kunkel et Pederson, 1988; Lobo et Hernandez, 1989; Parry et al., 1989). Cependant, leur position est différente. En effet, l'élément PSE est centré sur les positions -64/-62 dans les promoteurs des gènes transcrits par la Pol III. Il est décalé d'environ cinq paires de base en moyenne en amont de la position observée dans les gènes

transcrits par la Pol II (-59/-56). Ce décalage est important, mais pas suffisant, pour la détermination de la spécificité de la transcription par la Pol III (Lescure et al., 1991).

Une boîte TATA, en position -31/-25, est associée à la PSE dans le promoteur de base des gènes dépendants de la Pol III (Figure 17). Cette boîte est parfaitement similaire en séquence et en position à celle des gènes d'ARNm (Murphy et al., 1986; Carbon et al., 1987; Krol et al., 1987; Baer et al., 1990; Hernandez, 1992; Maraia et al., 1994). Cependant, cette boîte n'est pas présente dans les promoteurs des gènes d'ARNsn et de type ARNsn reconnus par la Pol II. En association avec la PSE, cet élément TATA est indispensable pour la transcription de base et est un déterminant majeur de la spécificité de transcription par la Pol III (Kunkel et Pederson, 1988; Mattaj et al., 1988; Lobo et Hernandez, 1989; Lobo et al., 1990; Lescure et al., 1991). En effet, le gène de l'ARNsn U6, initialement dépendant de la Pol III, est transcrit par la Pol II lorsqu'on supprime la boîte TATA de son promoteur. Inversement, l'addition de cette boîte TATA en position -31/-25 dans le promoteur du gène de l'ARNsn U2 permet la transcription de ce gène par la Pol III (Mattaj et al., 1988; Lobo et Hernandez, 1989).

On peut remarquer que le promoteur des gènes transcrits par la Pol III présente la même architecture que le promoteur externe du gène de l'ARNt^{Sec} (Figure 15 et Figure 17). Il a été montré que la PSE et la boîte TATA de ce gène sont fonctionnellement équivalentes à celles qui sont présentes dans le promoteur du gène de l'ARNsn U6 (Carbon et Krol, 1991; Myslinski et al., 1993). L'étude de la transcription de gènes d'ARNsn et de type ARNsn a mis en évidence que Staf participait à l'activation de la transcription des gènes des ARNs U4, Y4 et 7SK humains, des ARNs U1, U2, U5 et de la MRP chez le xénope et du gène murin et humain de l'ARNsn U6 (Schaub et al., 1997). Ainsi, Staf est un activateur de la transcription des gènes d'ARNsn et de type ARNsn transcrits par la Pol II et la Pol III.

Staf est également capable d'activer la transcription d'un gène rapporteur d'ARNm (Schuster et al., 1995). En effet, ce facteur permet l'activation de la transcription à partir d'un promoteur chimère, composé du site de liaison de Staf fusionné au promoteur de base du gène de la thymidine kinase et placé en amont du gène rapporteur chloramphénicol acétyl-transférase.

Deux régions distinctes du domaine amino-terminal (domaine A) de Staf permettent d'activer spécifiquement la transcription de ces deux catégories de gènes (Schuster et al., 1998). Le sous-domaine responsable de l'activation de la transcription des gènes d'ARNsn et de type

ARNsn est composé de dix-huit acides aminés localisés des positions 207 à 224 (Figure 18) (Schuster et al., 1998). La séquence de cette région est PDTISALEQYAAKVSIEG. Il a été montré que la leucine en position 213 de cette région est essentielle à l'activation de la transcription des gènes d'ARNsn et de type ARNsn (Schuster et al., 1998).

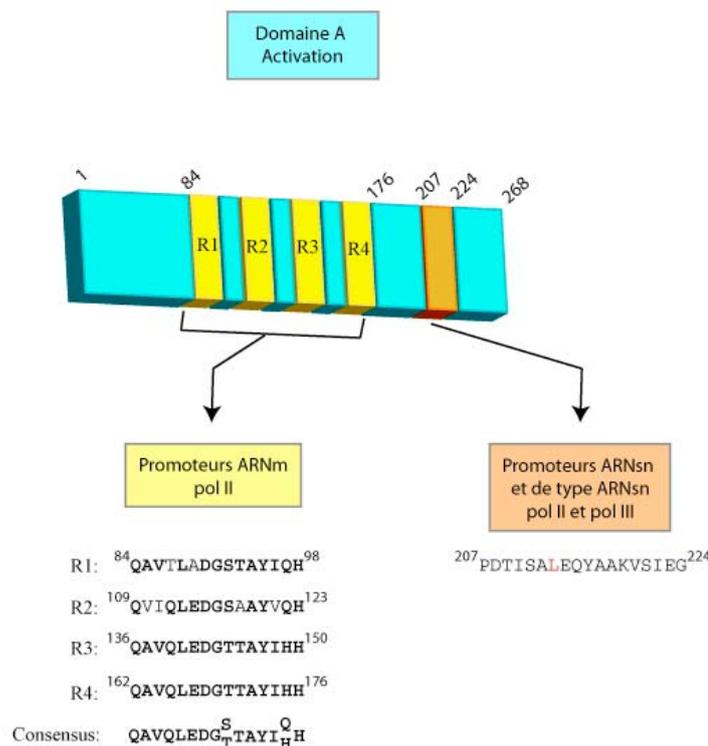


Figure 18 : Caractéristiques structurales du domaine d'activation de la transcription de Staf (Schuster et al., 1998).

Le domaine d'activation de la transcription des gènes d'ARNm, constitué de 4 répétitions d'un motif R, est indiqué en jaune. La séquence et les positions des quatre motifs répétés R1, R2, R3 et R4 ainsi que la séquence consensus sont indiquées. Le domaine d'activation de la transcription des gènes d'ARNsn et de type ARNsn est représenté en orange. La séquence et les positions de cette région de dix-huit acides aminés sont indiquées. La leucine essentielle est en rouge.

La région responsable de l'activation de la transcription de gènes d'ARNm est localisée entre les acides aminés 84 et 176 (Schuster et al., 1998). Elle est composée de quatre répétitions d'un motif R de quinze acides aminés séparées par dix à douze acides aminés (Figure 18). La

seule présence du motif R1 permet l'activation de la transcription. Cependant, l'efficacité d'activation augmente lorsque les autres motifs sont ajoutés un à un. L'efficacité d'activation de Staf est maximale lorsque les motifs R1 à R4 sont présents (Schuster et al., 1998). Les résidus leucine, aspartate et thréonine localisés respectivement aux positions 166, 168 et 170 du motif R4 contribuent de manière importante à l'activité de ce sous-domaine d'activation (Schuster et al., 1998).

Ainsi le facteur de transcription Staf possède deux régions d'activation physiquement et fonctionnellement distinctes. On peut noter que le facteur d'activation Oct1 peut également activer à la fois la transcription des gènes d'ARNsn et d'ARNm en se fixant sur le site octamère de la DSE (Tanaka et al., 1992). Oct1 possède deux domaines d'activation. Le premier, localisé du côté amino-terminal, est riche en résidus glutamine alors que le second, localisé du côté carboxy-terminal, est riche en résidus sérine et thréonine (Tanaka et al., 1992). Cependant, les domaines précisément requis pour l'activation de la transcription des gènes d'ARNsn ou de type ARNsn et des gènes d'ARNm n'ont pas été définis. De manière surprenante, il a été observé que le DBD de Oct1 joue un rôle important dans l'activation de la transcription des gènes d'ARNsn. En effet, le domaine POU-spécifique de ce DBD permet le recrutement coopératif de Oct1 au motif octamère et du facteur général de la transcription SNAPc sur la PSE (Murphy et al., 1992; Mittal et al., 1996; Ford et Hernandez, 1997; Murphy, 1997; Ford et al., 1998). La stabilisation de SNAPc au niveau de la PSE permet une formation efficace du complexe d'initiation de la transcription au niveau du promoteur (Mittal et al., 1999).

3.3. Les protéines homologues de Staf chez l'homme

L'interrogation des banques de données de protéines de l'institut européen de bioinformatiques (EBI) avec la séquence de Staf de *Xenopus laevis* a révélé une homologie de séquence significative entre la protéine de xénope et deux protéines humaines de fonction inconnue (Myslinski et al., 1998). Ces deux protéines, ZNF143 et ZNF76, présentent respectivement 84% et 64% d'identité avec Staf et sont constituées respectivement de 626 et 515 acides aminés (Figure 19).

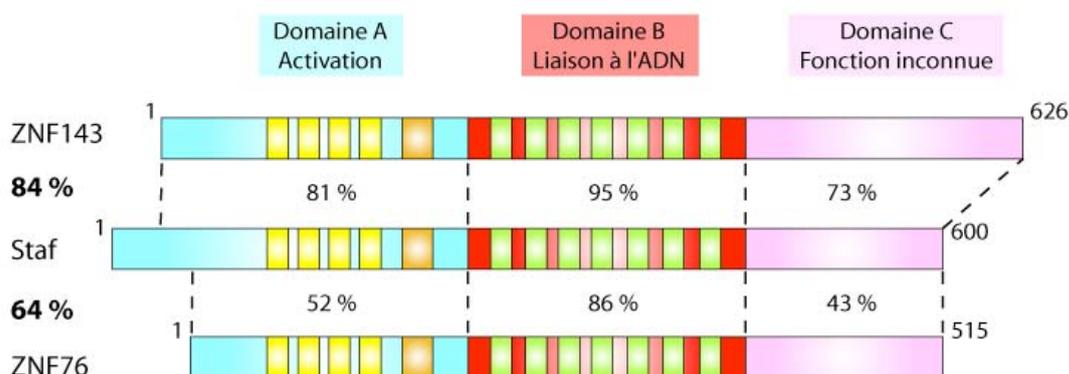


Figure 19 : Représentation schématique des protéines homologues ZNF143, Staf et ZNF76 (d'après Myslinski *et al.*, 1998).

La taille des trois protéines est indiquée. Les trois domaines A, B et C sont représentés respectivement en bleu, rouge et rose. Les deux sous-domaines d'activation sont indiqués en jaune et orange et les sept doigts de zinc sont en vert. Le pourcentage d'identité entre Staf et ZNF143 et Staf et ZNF76 est indiqué en gras. Les pourcentages d'identité entre les domaines A, B et C de Staf et ZNF143 et de Staf et ZNF76 sont indiqués.

Ces protéines peuvent être divisées en trois domaines équivalents à ceux de Staf (Myslinski *et al.*, 1998). Le domaine central de ZNF143 et ZNF76 contient sept doigts de zinc de type C_2H_2 et présente une forte identité de séquence avec celui de Staf (Figure 19). Les deux sous-domaines d'activation spécifiques de la transcription des gènes d'ARNsn et d'ARNm sont également retrouvés dans les deux protéines ZNF143 et ZNF76 (Figure 19). La comparaison de la séquence de ces trois protéines a également permis de mettre en évidence une région de 24 acides aminés très conservée dans le domaine carboxy-terminal. En effet, 16 acides aminés sur 24 sont identiques dans la région comprise entre les positions 561-584 dans Staf, 469-492 dans ZNF76 et 587-610 dans ZNF143 (Myslinski *et al.*, 1998). Cependant, des différences entre ces trois protéines sont également observées. En effet, ZNF143 et ZNF76 possèdent un domaine amino-terminal plus court respectivement de 45 et 77 acides aminés par rapport à celui de Staf. Le domaine carboxy-terminal de ZNF143, quant à lui, présente une insertion de 53 acides aminés entre les positions 498 et 551 alors que celui de Staf et de ZNF76 ont la même taille (Figure 19) (Myslinski *et al.*, 1998).

Les similarités observées entre ces protéines laissent supposer que ZNF143 et ZNF76 sont des homologues humains de la protéine Staf de xénope. Il a été montré que les DBD de ces deux protéines reconnaissent avec une affinité comparable les séquences cibles de Staf présentes

dans les promoteurs des gènes de l'ARNt^{Sec} de xénope et de l'ARNsn U6 humain (Myslinski et al., 1998). De plus, des expériences de dominants négatifs dans des oocytes de xénope ont montré que les DBD de ZNF143 et ZNF76 réduisent l'efficacité de transcription du gène de l'ARNt^{Sec} en entrant en compétition avec la protéine Staf endogène (Myslinski et al., 1998). Il a également été montré que les protéines ZNF143 et ZNF76 peuvent activer la transcription des gènes de l'ARNsn U1 et de l'ARNt^{Sec} de xénope ainsi que la transcription d'un gène rapporteur d'ARNm (Myslinski et al., 1998). L'ensemble de ces résultats ainsi que la comparaison des séquences montrent que ZNF143 est l'orthologue humain de Staf et que ZNF76 est une protéine homologue à ZNF143 et Staf.

Les transcrits codant ces deux protéines ont été retrouvés en proportion variable dans plusieurs tissus par northern-blot (Myslinski et al., 1998). De plus, de nombreuses EST spécifiques de ces transcrits sont détectées au sein de nombreux tissus et à différents stades du développement (EST Profile Viewer, UniGene, NCBI). L'homologue de Staf chez la souris a également été identifié et nommé mStaf (Adachi et al., 1998). Des comparaisons de séquences ont permis d'identifier chez le poisson zèbre et le fugu des protéines montrant un fort pourcentage d'identité avec Staf (Myslinski et al., 2004). Cependant aucune étude fonctionnelle n'a été réalisée avec ces homologues potentiels. Staf est donc un facteur de transcription présent dans de nombreuses espèces de vertébrés. Des recherches bioinformatiques complémentaires, réalisées au laboratoire, montrent que la protéine Staf est présente dans d'autres espèces faisant partie de l'embranchement des chordées comme *Ciona intestinalis* (résultats non publiés).

Récemment, sept gènes de protéines ont été identifiés comme étant régulés par ZNF143. Chez l'homme, il s'agit des gènes codant le facteur de régulation de l'interféron 3 (IFR3) (Mach et al., 2002), la transaldolase (TALDO1) (Grossman et al., 2004), l'oxyde nitrique synthase neuronale (NOS1) (Saur et al., 2002), la protéine ribosomique mitochondriale S11 (MRPS11) (Ishiguchi et al., 2004), la synaptobrevine-like 1 (SYBL1) (Di Leva et al., 2004) et l'aldéhyde réductase (AKR1A1) (Barski et al., 2004). La transcription de ce dernier gène a également été identifiée comme étant dépendante de Staf chez la souris ainsi que celle du gène codant une sous unité d'un complexe de chaperons cytosoliques (TCP1) (Kubota et al., 2000).

C'est en recherchant des gènes dont l'expression était induite par traitement au cisplatine que Ishiguchi *et al* ont identifié le gène MRPS11 (Ishiguchi et al., 2004). Puisque ZNF143 régule

l'expression de ce gène, ces auteurs se sont donc intéressés aux effets sur ZNF143 de l'administration de cisplatine à des cellules. Ils ont observé que l'expression de ZNF143 augmente dans des cellules cancéreuses sensibles au cisplatine et que la sensibilité de ces cellules augmente lorsque l'expression de ZNF143 est presque totalement éliminée par des expériences d'interférence à l'ARN (Ishiguchi et al., 2004; Wakasugi et al., 2007). De plus, lors de ces expériences d'interférence, on observe une resensibilisation des cellules résistantes au cisplatine (Wakasugi et al., 2007).

Des expériences de coimmunoprécipitation et de retard sur gel ont montré que ZNF143 peut interagir avec la protéine suppresseur de tumeur p73 et que ce facteur p73 peut stimuler la fixation de ZNF143 sur sa séquence cible ainsi que sur une séquence riche en GC modifiée par du cisplatine (Wakasugi et al., 2007). ZNF143 pourrait avoir un rôle potentiel dans la réparation de l'ADN en activant spécifiquement des gènes impliqués dans ce mécanisme en réponse à un traitement au cisplatine. Pour identifier ces gènes, les auteurs ont recherché des sites de liaison de ZNF143 dans le promoteur d'une série de gènes impliqués dans le mécanisme de réparation de l'ADN. Ils ont identifiés un grand nombre de gènes potentiellement régulés par ZNF143 (Wakasugi et al., 2007). Cependant, leur recherche est biaisée car ils ont considéré que le site de liaison de ZNF143 se réduisait à quatre résidus cytosines en position 3-5 et 12. Nous verrons dans la partie 3.3 de mon introduction que ZNF143 reconnaît un site de liaison certes dégénéré mais tout de même plus complexe (Schaub et al., 1997; Schaub et al., 1999a).

De récentes études réalisées par Zheng et Yang ont montré que ZNF76 peut également avoir une fonction de répresseur de la transcription (Zheng et Yang, 2004). En effet, la surexpression par transfections transitoires de ZNF76 dans différents types cellulaires entraîne l'inhibition de l'expression de gènes cibles du facteur de transcription p53. Cette réduction a été observée pour un gène rapporteur sous le contrôle d'un site de liaison du facteur p53 ainsi que pour le gène endogène codant la protéine p21 (Zheng et Yang, 2004). Des expériences de double-hybride et de coimmunoprécipitation ont montré que ZNF76 peut interagir, via ses domaines amino et carboxy-terminaux, avec TBP (Zheng et Yang, 2004). La surexpression par transfections transitoires de ZNF76 muté dans son domaine carboxy-terminal et des expériences de dominant négatif réalisées avec le domaine carboxy-terminal sauvage montrent que ZNF76 perd la capacité d'inhiber l'expression de gènes cibles du facteur de transcription p53. Ceci suggère que ZNF76 inhibe la transcription en se liant à TBP.

Il a été montré que l'activité de ZNF76 peut être régulée par des modifications post-traductionnelles (Zheng et Yang, 2004; Zheng et Yang, 2006). ZNF76 présente différents isoformes lorsque deux constituants du système de sumoylation, les protéines PIAS1 et SUMO-1, sont surexprimées avec ZNF76 dans des cellules HEK293 (Zheng et Yang, 2004). La sumoylation de ZNF76 inhibe son interaction avec TBP et entraîne une réduction drastique de son activité de répression de la transcription de gènes cibles de p53 (Zheng et Yang, 2004). Le domaine amino-terminal de ZNF76 peut, quant à lui, être acétylé par la protéine p300 et déacétylé par HDAC1 dans des conditions de transfections transitoires (Zheng et Yang, 2006). Cette acétylation entrerait en compétition avec la sumoylation de ZNF76 et entraînerait une diminution de l'efficacité d'interaction de ZNF76 avec TBP. Zheng et Yang ont également observé qu'il existe deux isoformes de 515 et 570 acides aminés de la protéine ZNF76 issus d'un épissage alternatif. Les auteurs ne se sont intéressés qu'à l'interaction de ces deux isoformes avec TBP. Ils ont montré que ZNF76₅₇₀ permet de coimmunoprécipiter une quantité plus importante de TBP (Zheng et Yang, 2006).

3.4. Mode de liaison de Staf à l'ADN

Staf se lie à l'ADN au niveau d'un site appelé SBS (pour Staf Binding Site) par l'intermédiaire de son domaine central, composé de motifs en doigt de zinc de type C₂H₂. Des expériences de sélection de sites *in vitro* par la technique SELEX (Systematic Evolution Ligands by EXponential enrichment) ont permis de définir la séquence consensus reconnue par le DBD de Staf (Schaub et al., 1997). Cette séquence de 18 pb est YWCCCRNMATSCMYRRCR (Figure 20A). Ce site de liaison présente un haut niveau de variation puisque 12 des 18 positions sont dégénérées. En effet, la position 7 de ce site est totalement dégénérée et les positions 1, 2, 6, 8, 11, 13-18 sont moins conservées que les positions 3-5, 9, 10, 12. La comparaison de la séquence de sites naturels de liaison de Staf validés expérimentalement chez l'homme, le xénope et la souris a également permis de déterminer une séquence consensus spécifiquement pour les gènes d'ARNsn et de type ARNsn ainsi que pour les gènes d'ARNm (Figure 20B). Les deux séquences consensus obtenues sont proches de celle identifiée par la technique de SELEX. La séquence consensus obtenue à partir des gènes d'ARNsn et de type ARNsn a permis d'identifier par comparaison

de séquence d'autres gènes d'ARNsn et de type ARNsn potentiellement régulés par Staf comme les gènes codant

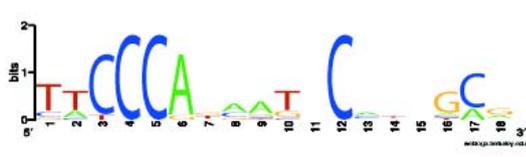
A. Détermination par SELEX

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
A%	6.8	55.0	1.6	0.0	0.0	83.9	3.2	54.8	91.8	1.4	0.0	1.4	74.6	23.4	1.9	29.3	3.2	23.1
C%	20.5	3.3	95.2	100.0	100.0	0.0	38.7	35.6	1.4	0.0	19.2	97.2	15.5	34.4	46.3	2.4	77.4	7.7
G%	6.8	10.0	0.0	0.0	0.0	14.5	30.7	9.6	6.8	1.4	75.3	1.4	2.9	0.0	3.7	58.5	9.7	61.5
T%	65.9	31.7	3.2	0.0	0.0	1.6	27.4	0.0	0.0	97.2	5.5	0.0	7.0	42.2	48.1	9.8	9.7	7.7
Number of sequences	44	60	62	62	62	62	62	73	73	73	73	73	71	64	54	41	31	26
Consensus	Y	A/t	C	C	C	A/g	N	A/c	A	T	G/c	C	A/c	Y	Y	R	C	R

B. Détermination par comparaison de séquences

1. Gènes d'ARNsn et de type ARNsn

PolII	hU4C	TCCCCAGCGTCCCAAGCG
	hU4ATAC	TACCCAGAAGACCCGCG
	xU1b1	CTCCCATAGTTCATTGCA
	xU2	TTCCCGACTGCCCCGGCA
	xU5	TACCCATGCTGCATTAAG
PolIII	hY4	TTCCCATCATGCAACTAC
	h7SK	TATCCAGAATGCCTTGCA
	hH1	TTCCCAAGACACATAGCG
	hU6	TTCCCATGATTCCCTTCAT
	htRNasec	TTCCCAAGAAATGCGTGGCG
	xMRP	GTCCCATCATGCAAAGCG
	mU6	CTCCACAAGTCTGTGCG
consensus	TTCCANNATNCNNNGCG	



2. Gènes d'ARNm

h IRF-3	CTCCCAGCATGCCTCTGG
h MRP S11	CTCCACAATGCACCGCG
m Ccta site 1	TTCCCGGCAGACCTTGCG
m Ccta site 2	TACCCAGCAGGCCCGCG
h SYBL1 site 1	CTCCCAAGAGGCACGCG
h SYBL1 site 2	CTCCCAAGAGGCACGCG
h TAL-H	GGCCACAATGCCCGCG
h AKR1A1	GGCCACATTGCACCGCG
m AKR1A4	GGCCACAGTGCACCGCG
h nNOS1c	CTCCCAGCTGCCCTGGG
consensus	NNCCANNAKGCMMNYGCG



Figure 20 : Détermination de la séquence consensus du site de liaison de Staf.

A. Tableau récapitulatif des fréquences des bases obtenues pour chaque position du SBS par la technique SELEX permettant de déterminer la séquence consensus indiquée sur la dernière ligne du tableau (Schaub et al., 1997). **B.1.** Alignement des séquences SBS de promoteurs de gènes d'ARNsn et de type ARNsn validées expérimentalement (Schaub et al., 1997; Myslinski et al., 2001). **B.2.** Alignement des séquences SBS de promoteurs de gènes d'ARNm validées expérimentalement (Kubota et al., 2000; Mach et al., 2002; Saur et al., 2002; Barski et al., 2004; Di Leva et al., 2004; Grossman et al., 2004; Ishiguchi et al., 2004). La représentation

des séquences consensus obtenues par comparaison de séquences a été réalisée grâce au logiciel WebLogo (Crooks et al., 2004).

l'ARN Y1 chez la souris ou l'ARNsn U2 chez l'homme, la souris, le rat et le poulet (Schaub et al., 1997).

Le facteur de transcription Staf reconnaît son site de liaison par l'intermédiaire de son domaine de liaison à l'ADN composé de sept doigts de zinc de type C₂H₂. Le motif structural en doigt de zinc a été identifié pour la première fois au sein du facteur de transcription TFIID (Miller et al., 1985; Diakun et al., 1986). Ce motif est retrouvé dans de nombreuses protéines impliquées dans la régulation de l'expression des gènes et le nombre de doigts de zinc dans une protéine est variable. La comparaison des séquences de nombreux motifs en doigt de zinc du type C₂H₂ a permis de déduire une séquence consensus de celui-ci (Figure 21A) (Angrand, 1993; Desjarlais et Berg, 1993). Il apparaît que certains acides aminés sont conservés notamment deux cystéines et deux histidines qui sont invariantes. De plus, une phénylalanine ou une tyrosine ainsi qu'une leucine sont respectivement retrouvées quatre et dix résidus après la deuxième cystéine invariante. La séquence de liaison, située entre deux doigts de zinc, est également très conservée.

Figure 20 : Le domaine de liaison à l'ADN de Staf est constitué de sept doigts de zinc de type C₂H₂.

A. Séquence consensus d'un motif en doigt de zinc de type C₂H₂. **B.** Le domaine de liaison à l'ADN de Staf ainsi que les sept doigts de zinc sont indiqués respectivement en rouge et en vert. Les séquences et les positions des sept doigts de zinc (Zf1-Zf7) sont indiqués. La séquence consensus d'un motif en doigt de zinc de type C₂H₂ est rappelée en dessous de l'alignement. Les acides aminés impliqués dans les feuillets β et l'hélice α sont indiqués par des flèches. Pour A et B : Les deux cystéines (C) et les deux histidines (H) invariantes sont indiquées en rouge et les acides aminés conservés sont en gras. La lettre X représente n'importe quel acide aminé. **C.** Représentation schématique de la structure tertiaire d'un doigt de zinc de type C₂H₂ établie par RMN (illustration tirée de (Branden et Tooze, 1991) adaptée de (Lee et al., 1989)). L'hélice α et les feuillets β sont

Figure 21 : Le domaine de liaison à l'ADN de Staf est constitué de sept doigts de zinc de type C2H2.

Les sept doigts de zinc du type C2H2 de Staf sont semblables à ceux initialement caractérisés dans TFIIIA et leur séquence est proche de la séquence consensus établie pour ce type de motif (Figure 21B). La première structure tertiaire d'un doigt de zinc a été réalisée par RMN à partir de la protéine Xfin de xénope (Lee et al., 1989). Cette structure montre que ce motif est constitué de deux brins β antiparallèles de trois résidus chacun et d'une hélice α composée de douze acides aminés (Figure 21C). Les deux résidus cystéines et histidines sont localisés respectivement sur le feuillet β et sur l'hélice α . Ces quatre acides aminés forment des liaisons de coordination avec un atome de zinc. Ceci permet le rapprochement du feuillet β et de l'hélice α et la formation d'une structure globulaire en forme de doigt.

La détermination de la première structure tridimensionnelle d'un complexe ADN-protéine à doigts de zinc a été réalisé à partir de la protéine Zif268 de souris associée à sa séquence cible (Pavletich et Pabo, 1991). Cette structure a permis de visualiser le mode d'interaction d'un doigt de zinc avec les bases de l'ADN. Une protéine contenant plusieurs doigts de zinc s'enroule autour de l'hélice d'ADN. Le motif en doigt de zinc contacte le grand sillon de la molécule d'ADN via son hélice α et interagit avec trois paires de base contiguës appelées triplet. Cette reconnaissance se réalise majoritairement de façon antiparallèle. En effet, les résidus -1, 3 et 6 par rapport au premier résidu de l'hélice α contactent respectivement la base 3', médiane et 5' du triplet. Cependant, un contact supplémentaire est observé entre le résidu en position 2 de l'hélice α et une base localisée sur l'autre brin d'ADN au niveau du triplet précédent.

L'emploi de techniques d'interférence de liaison a permis d'identifier les bases essentielles pour la fixation de la forme entière et tronquée du DBD de Staf sur la séquence cible présente dans les promoteurs des gènes de l'ARNt^{Sec} de xénope et de l'ARNsn U6 humain (Schaub et al., 2000). Il apparaît que les doigts de zinc de Staf interagissent avec l'ADN en suivant le mode mis en évidence avec Zif268 à savoir un doigt de zinc reconnaissant trois paires de base. L'utilisation de formes tronquées du DBD de Staf a permis de montrer qu'un sous-domaine uniquement composé des doigts de zinc 3 à 6 était suffisant, *in vitro*, pour la liaison au SBS du gène de l'ARNsn U6 humain et de l'ARNt^{Sec} de xénope (Schaub et al., 2000).

Des études complémentaires montrent que les doigts de zinc d'une protéine n'interagissent pas nécessairement tous avec l'ADN (Pavletich et Pabo, 1993; Nolte et al., 1998). Ceci a également été observé dans le cas du facteur Staf. En effet, le septième doigt de zinc du DBD de Staf ne contacte pas spécifiquement les bases du SBS présent dans les promoteurs des gènes de l'ARNsn U6 humain et de l'ARNt^{Sec} de xénope (Schaub et al., 1999a). Ceci explique la longueur de dix-huit paires de bases observée pour le site de liaison de Staf dans la mesure où six des sept doigts de zinc établissent des contacts avec trois paires de bases consécutives. Cependant, le septième doigt de zinc apparaît important dans la stabilisation de la liaison de Staf sur son site puisque l'absence de ce motif diminue l'affinité de Staf pour son SBS (Schaub et al., 1999a). Le doigt de zinc numéro 1 présente, quant à lui, une flexibilité d'interaction puisqu'il contacte le SBS du gène de l'ARNt^{Sec} mais pas celui du gène de l'ARNsn U6 (Schaub et al., 1999a). L'activité transcriptionnelle maximale de ces deux gènes est dépendante de la présence ou non des interactions établies par ce premier doigt de zinc (Schaub et al., 1999b). En effet, il a été montré que la transcription optimale du gène codant l'ARNsn U6 humain nécessite l'absence d'interaction du doigt de zinc numéro 1 avec l'ADN afin de permettre la formation d'un complexe entre Oct1 et sa séquence cible. Contrairement à ce gène, le promoteur du gène codant l'ARNt^{Sec} de xénope permet la fixation du doigt de zinc 1. Cette flexibilité se reflète au niveau de la séquence des SBS de ces deux gènes. En effet, l'extrémité 3' du SBS du gène de l'ARNsn U6 diverge par rapport à la séquence consensus alors que celle du gène de l'ARNt^{Sec} est conservée (Figure 20B).

RÉSULTATS ET DISCUSSIONS

CHAPITRE 1

Identification des sites de liaison de hStaf/ZNF143 dans les promoteurs des gènes de protéines humaines

CHAPITRE 1 : Identification des sites de liaison de hStaf/ZNF143 dans les promoteurs des gènes de protéines humaines

1. Introduction :

Comme cela a été présenté dans la troisième partie de mon introduction aux paragraphes 3.1 et 3.2, le facteur de transcription Staf a été caractérisé au laboratoire comme étant l'activateur de la transcription du gène de l'ARN de transfert de la sélénocystéine chez le xénope (Schuster et al., 1995). Par la suite, il a été montré que ce facteur activait la transcription des gènes d'ARNsn et de type ARNsn par les ARN polymérase II et III chez de nombreux vertébrés (Schaub et al., 1997; Myslinski et al., 2001).

De façon surprenante, Staf possède également la capacité de stimuler la transcription d'un gène rapporteur d'ARNm à partir d'un promoteur chimère composé de trois sites de liaison de Staf clonés en amont du promoteur de base du gène de la thymidine kinase (Schuster et al., 1995). De plus, de récentes études ont montré que l'orthologue humain de Staf, hStaf/ZNF143, pouvait activer la transcription de sept gènes de protéines (Myslinski et al., 1998; Kubota et al., 2000; Mach et al., 2002; Saur et al., 2002; Barski et al., 2004; Di Leva et al., 2004; Grossman et al., 2004; Ishiguchi et al., 2004).

Nous nous sommes demandés si Staf était impliqué dans la transcription d'autres gènes de protéines. Pour cela, nous avons recherché le site de liaison du facteur hStaf/ZNF143 (SBS) dans les régions promotrices des gènes de protéines chez l'homme. Ceci a pu être réalisé en interrogeant la banque de données DBTSS (Suzuki et al., 2002; Yamashita et al., 2006). Cette banque de données rassemble des informations sur de nombreux ADNc complets humains et murins. De ce fait, il est possible d'accéder aux sites d'initiation de la transcription ainsi qu'aux régions promotrices localisées en amont de ces sites. Lors de notre étude, nous avons utilisé la version 4 de cette banque de données qui contenait 8793 et 6875 séquences promotrices chez l'homme et la souris respectivement. Nous avons interrogé cette banque de données avec la séquence consensus du SBS obtenue par comparaison des séquences des sites observés dans les promoteurs des sept gènes de protéines connus comme régulés par Staf (YYCCANNRNRRCNNYRCR). Cette analyse bioinformatique a été complétée par des interrogations avec dix autres séquences divergentes établies à partir des résultats obtenus lors

de la détermination de la séquence consensus du SBS par des expériences de SELEX (Schaub et al., 1997; Schaub et al., 1999a).

Cette recherche nous a permis d'identifier des promoteurs contenant des SBS potentiels. Nous avons ensuite recherché la présence de hStaf/ZNF143 *in vivo* au niveau de ces promoteurs par des expériences d'immunoprécipitation de chromatine de cellules HeLa couplées à des réactions de polymérisation en chaîne (ChIP-PCR).

2. Résultats :

2.1. Résumé des principaux résultats

L'interrogation de la base de données DBTSS avec les onze séquences différentes a permis d'identifier chez l'homme 1488 sites potentiels de liaison de hStaf/ZNF143. Seuls 648 de ces sites (43,5%) sont conservés chez la souris. Ces sites conservés ont été retenus pour la suite de notre étude. L'analyse des régions promotrices renfermant ces SBS potentiels a montré qu'un motif de sept paires de bases de séquence ACTACAN était retrouvé juste en amont de ces sites dans 47% des cas. De ce fait, nous avons complété notre étude en interrogeant la banque DBTSS avec une séquence de 19 pb composée du motif ACTACAN suivi des 12 premières paires de bases des séquences SBS employées initialement. Nous avons ainsi identifié 204 nouveaux SBS potentiels conservés chez la souris. L'analyse manuelle des régions promotrices renfermant des sites SBS potentiels a permis de mettre en évidence 41 SBS additionnels.

Les 893 SBS ainsi identifiés et conservés chez la souris se répartissent parmi 716 promoteurs (Figure 22). 86,5% de ces régions promotrices sont retrouvés dans le génome du chien et du rat et la quasi-totalité des SBS (97%) de ces régions est également conservée dans ces génomes.

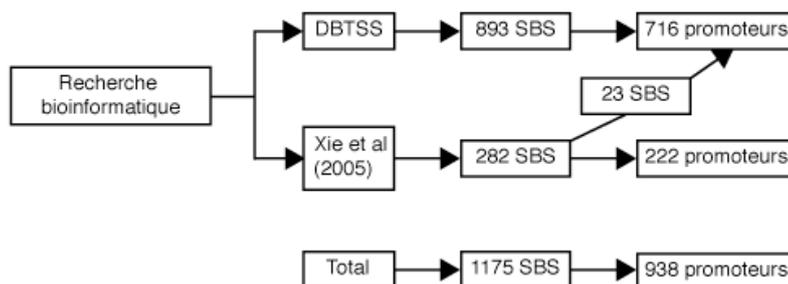


Figure 22 : Représentation schématisée des résultats obtenus lors de la recherche bioinformatique de SBS dans les promoteurs de gènes de protéines humaines (d’après Myslinski et al., 2006).

La recherche bioinformatique a été réalisée à partir de la banque de données DBTSS et des résultats publiés par Xie *et al.* Le nombre de SBS et le nombre de promoteurs renfermant ces sites identifiés par ces deux approches sont indiqués.

Au cours de notre étude, Xie *et al.* ont identifié, en comparant les séquences des régions promotrices de plusieurs génomes de mammifères, 174 motifs régulateurs conservés dans ces promoteurs (Xie et al., 2005) Parmi ces motifs, le motif M4 ACTAYRNNNCCCR est particulièrement intéressant car il renferme dans sa partie 5’ la séquence ACTACAN, identifiée dans notre étude immédiatement en amont de nombreux SBS. Nous avons donc examiné si ces motifs M4, allongés de 12 pb en 3’, renfermaient des SBS potentiels. De cette manière, nous avons identifié 282 nouveaux SBS potentiels dont 23 sont présents dans les promoteurs précédemment identifiés. Les autres SBS se répartissent dans 222 nouvelles régions promotrices (Figure 22). La presque totalité (97%) des 282 nouveaux sites est conservée chez la souris, le rat et le chien. Ainsi, cette analyse bioinformatique nous a permis d’identifier au total 1175 SBS potentiels conservés répartis dans 938 régions promotrices de gènes d’ARNm de mammifères (Figure 22).

Les sites potentiels de liaison de hStaf/ZNF143 identifiés lors de notre étude sont majoritairement placés en orientation directe par rapport au gène adjacent et sont, dans 62% des cas, localisés dans une région de 200 pb en amont du TSS. Un cinquième des régions promotrices identifiées possède plusieurs SBS. Bien que le promoteur du gène NCLN renferme six SBS, la majorité des promoteurs à SBS multiples contient deux sites. De façon surprenante, 15% des promoteurs identifiés renferment des SBS communs. Dans ce cas de figure, les deux gènes sont en orientation divergente et les TSS de ces gènes sont séparés de

moins de 1 kpb. Ainsi, Staf pourrait intervenir dans la transcription de gènes contrôlée par des promoteurs bidirectionnels. Il a également été observé que les promoteurs caractérisés lors de cette analyse sont essentiellement localisés dans des îlots CpG et ne renferment pas de boîte TATA.

Afin de tester si les nombreux sites potentiels de liaison de hStaf/ZNF143 identifiés *in silico* correspondent à une réalité biologique, nous avons examiné si ces sites étaient occupés par hStaf/ZNF143 *in vivo*. Pour cela, dans un premier temps l'ADN associé à hStaf/ZNF143 a été isolé par immunoprécipitation de chromatine (ChIP). L'immunoprécipitation de chromatine a été réalisée à partir de cellules HeLa et avec des anticorps spécifiques du facteur hStaf/ZNF143. Nous avons ensuite recherché dans l'ADN ainsi isolé, par PCR semi-quantitative avec des couples d'oligonucléotides spécifiques, la présence de fragments contenant des SBS potentiels. 295 promoteurs représentant 56% des promoteurs à sites multiples et 25% à sites uniques ont ainsi été testés. Des résultats positifs sont observés dans 90% des cas et ceci suggère très fortement que les SBS identifiés lors de notre analyse *in silico* constituent de véritables sites de fixation pour hStaf/ZNF143. On peut noter que les régions non amplifiées lors de ces PCR ne sont pas nécessairement des régions ne liant pas Staf puisque la technique de ChIP est dépendante de l'accessibilité de l'anticorps à l'épitope de la protéine dans la structure chromatinienne. De façon intéressante, nous avons observé que 15% des sites de liaison de hStaf/ZNF143, identifiés suite à l'analyse des résultats obtenus par Xie *et al*, présentaient un nucléotide autre que la cytidine conservée en position 12 du site consensus du SBS. Les expériences de ChIP-PCR réalisées sur cinq de ces sites non canoniques ont montré que cette substitution n'avait pas d'influence sur la fixation de hStaf/ZNF143 sur ces régions.

Les régions promotrices identifiées lors de notre étude sont majoritairement localisées dans des îlots CpG et dépourvues de boîte TATA. Nous nous sommes demandés si Staf pouvait activer la transcription de gènes d'ARNm en l'absence d'éléments du promoteur de base. Pour cela, nous avons testé, par des expériences de transfections transitoires, la capacité de Staf à activer la transcription d'un gène rapporteur placé sous le contrôle de trois sites SBS identiques. Nous avons observé que Staf est capable de diriger la transcription à partir de ce type de promoteur. Par une approche équivalente, nous avons montré que le motif ACTACAN, retrouvé juste en amont de nombreux SBS, est important dans l'activation de la transcription. En effet, l'absence de cette séquence entraîne une diminution de 50% du taux de

transcription d'un gène rapporteur placé sous la dépendance de SBS associés aux motifs ACTACAN.

Cette étude nous a permis d'identifier un grand nombre de sites de liaison de Staf en amont de gènes d'ARNm chez les mammifères. 30% de ces gènes codent pour des protéines dont la fonction n'est pas attribuée. Les autres gènes ont été classés par groupe selon leur fonction sur la base des informations obtenues dans NCBI et la banque de données GOA (Gene Ontology Annotation, <http://www.ebi.ac.uk/GOA/>) (Camon et al., 2003; Camon et al., 2004a; Camon et al., 2004b). Aucune classe de gènes de protéines n'apparaît être préférentiellement régulée par hStaf/ZNF143. Cependant, les classes prépondérantes, représentant chacune environ 20% des promoteurs totaux, sont celles qui regroupent les gènes codant des protéines de liaison à l'ADN et des facteurs de transcription ainsi que les gènes codant des protéines impliquées dans la synthèse, la maturation et la dégradation des protéines.

2.2. Publication : « A genome scale location analysis of human Staf/ZNF143-binding sites suggests a widespread role for human hStaf/ZNF143 in mammalian promoters »

Evelyne Myslinski, Marie-Aline Gérard, Alain Krol, and Philippe Carbon

The Journal of Biological Chemistry, 2006

A Genome Scale Location Analysis of Human Staf/ZNF143-binding Sites Suggests a Widespread Role for Human Staf/ZNF143 in Mammalian Promoters

Evelyne MYSLINSKI, Marie-Aline GÉRARD, Alain KROL, and Philippe CARBON

Journal of biological chemistry, 2006, Vol. 281, N° 52, Pages 39953-39962

Pages 85-94 :

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

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

<http://dx.doi.org/10.1074/jbc.M608507200>

La version imprimée de cette thèse peut être consultée à la bibliothèque ou dans un autre établissement via une demande de prêt entre bibliothèques (PEB) auprès de nos services :

<http://www-sicd.u-strasbg.fr/services/peb/>

3. Discussion

L'identification des sites de liaison de hStaf/ZNF143 à l'échelle du génome humain a révélé un nombre inattendu et élevé de SBS conservé dans les régions promotrices de quatre espèces de mammifères. En effet, 1175 SBS ont été identifiés dans 938 promoteurs de gènes de protéines. Les expériences de ChIP-PCR réalisées sur 295 régions promotrices ont démontré la liaison de hStaf/ZNF143 dans 90% des promoteurs testés. Ces résultats soulignent l'efficacité de la recherche bioinformatique qui a donné peu de faux positifs. De plus, la technique de ChIP est dépendante de l'accessibilité de l'anticorps à l'épitope de la protéine cible dans la chromatine. De ce fait, hStaf/ZNF143 peut être associé à certaines régions promotrices bien qu'aucune amplification de ces régions n'ait été obtenue lors des PCR réalisées sur l'ADN immunoprécipité.

L'utilisation de la banque de données DBTSS a permis d'identifier 893 SBS répartis dans 716 régions promotrices. Les 222 autres promoteurs identifiés résultent de l'exploitation des données publiées par Xie *et al* (Xie et al., 2005). Lors de notre étude, la banque DBTSS contenait 8793 promoteurs humains alors que le génome humain renferme environ 25 000 gènes de protéines (Consortium, 2004). L'extrapolation de nos données à ces 25 000 gènes permet d'estimer que le génome humain doit renfermer environ 2500 SBS se répartissant dans 2000 régions promotrices de gènes codant des protéines. Cette valeur est très similaire au nombre de site de liaison du facteur Sp1 dans les régions promotrices de ce type de gènes (Cawley et al., 2004). Ces résultats sont sûrement sous-estimés puisque l'interrogation de la banque de données a été réalisée avec la séquence consensus du SBS et quelques variants. Cependant, d'autres variations du site SBS peuvent être fonctionnelles. Ainsi, on remarque que les promoteurs des gènes IRF3 et AKR1A1, présents dans la banque DBTSS et connus comme étant régulés par hStaf/ZNF143, n'ont pas été retrouvés lors de notre étude car ils possèdent des SBS différents des séquences utilisées lors de notre analyse (Mach et al., 2002; Barski et al., 2004).

De plus, notre étude s'est limitée à l'analyse des régions promotrices alors qu'il a été montré récemment que des sites de liaison de facteurs de transcription peuvent être localisés dans d'autres régions du génome (Martone et al., 2003; Cawley et al., 2004; Euskirchen et al., 2004; Impey et al., 2004; Boyer et al., 2005; Lee et al., 2006; Loh et al., 2006; Wei et al.,

2006; Johnson et al., 2007; Robertson et al., 2007). Les travaux réalisés sur le génome humain entier par Yannick-Noël Anno, dans le cadre de son Master de Génomique Structurale et Bioinformatique (co-direction Philippe Carbon et Olivier Poch, IGBMC, Illkirch), montrent que c'est également le cas pour hStaf/ZNF143 puisqu'un enrichissement en SBS est observé à l'extrémité 5' du premier intron ainsi que globalement dans le dernier exon des gènes.

De manière surprenante, notre analyse a mis en évidence la présence d'une séquence ACTACAN juste en amont de 58% des sites SBS identifiés. Il est possible que cette séquence constitue le site de fixation pour un facteur de transcription non caractérisé à ce jour. Il est aussi envisageable que ce motif participe à une structuration de l'ADN facilitant la liaison de hStaf/ZNF143 sur son site. L'importance de cette séquence est également suggérée par plusieurs travaux publiés récemment (Prakash et Tompa, 2005; Xie et al., 2005; Lin et al., 2007). Dans l'étude réalisée par Xie *et al*, la comparaison des séquences des promoteurs de plusieurs génomes de mammifères a permis de mettre en évidence 174 motifs conservés et fortement retrouvés dans les promoteurs de mammifères. Les six premiers motifs M1 à M6 identifiés et les facteurs associés à ces motifs sont M1 : NRF-1, M2 : MYC, M3 : ELK-1, M4 : facteur inconnu, M5 : NFY et M6 : Sp1 (Xie et al., 2005). Le motif M4 de séquence ACTAYRNNNCCCR est particulièrement intéressant car notre étude montre clairement que ce motif, le 4^{ème} motif le plus conservé et le plus retrouvé dans les promoteurs de mammifères, résulte de l'association de la séquence ACTACAN avec les six premiers nucléotides du site SBS (Figure 23).

ACTACAN+SBS	ACTACANYYCCCANNRNRRCNNYRCR
Motif M4	ACTAYRNNNCCCR
Motif n°27	ACTACAANTCCCA
Motif LM4	GAACTACAAGTCCCAGCATGC

Figure 23 : Comparaison de 4 motifs conservés et fréquemment retrouvés dans le génome de mammifères avec le motif ACTACAN+SBS.

La séquence des motifs M4, n°27 et LM4 est comparée à la séquence du motif associant la séquence ACTACAN, identifiée dans notre étude, avec un site SBS (Prakash et Tompa, 2005; Xie et al., 2005; Myslinski et al., 2006; Lin et al., 2007; Xie et al., 2007).

De même, une autre étude a permis d'identifier, par comparaison des séquences de régions promotrices orthologues chez les mammifères, un motif proche en séquence du motif M4 ; il s'agit du motif conservé numéro 27 de séquence ACTACAANTCCCA (Figure 23) (Prakash et Tompa, 2005).

Le degré de conservation moindre de la partie 3' du site SBS peut expliquer qu'aucune de ces deux études n'aient permis de mettre en évidence cette séquence en aval des motifs M4 et n° 27. Très récemment, Xie *et al* ont publié les résultats d'une nouvelle étude dans laquelle le motif M4 a été étendu de deux nucléotides (GA) dans sa partie 5' et de six nucléotides (GCATGC) dans sa partie 3' pour donner le motif LM4 (Figure 23) (Xie et al., 2007). La séquence des douze derniers nucléotides du motif LM4 est très similaire à la séquence des douze premiers nucléotides du site SBS. Le motif LM4 semble donc encore mieux correspondre au motif ACTACAN-SBS identifié dans les analyses réalisées au laboratoire (Figure 23).

Nous avons observé que 42% des SBS identifiés dans notre étude ne possèdent pas une séquence adjacente de type ACTACAN. Ceci suggère que le motif correspondant au site SBS, associé ou non à la séquence ACTACAN, doit être encore plus abondant dans le génome que le motif M4. Ainsi le site de liaison de hStaf/ZNF143 semble être un des sites de liaison de facteurs de transcription les plus fréquents dans le génome humain.

Les régions promotrices identifiées au cours de cette étude se localisent majoritairement dans des îlots CpG et ne possèdent pas de boîte TATA. La grande proportion de SBS localisés à proximité des TSS suggère que hStaf/ZNF143 peut avoir un rôle direct dans l'initiation de la transcription. Nous avons montré que la seule présence de trois SBS en amont du gène de la luciférase est suffisante pour son expression. Ainsi, le facteur hStaf/ZNF143 peut diriger la transcription en recrutant directement ou indirectement la machinerie de transcription. Il a été montré, chez les mammifères, qu'environ 10% des gènes possèdent des promoteurs qualifiés de bidirectionnels. Dans ce type d'organisation, les points d'initiation de la transcription sont distants de moins de 1 kpb et un même promoteur peut diriger la transcription de deux gènes dans des directions opposées (Adachi et Lieber, 2002; Trinklein et al., 2004). De manière surprenante, 15% des régions promotrices identifiées dans notre étude sont orientées de manière divergente et renferment un SBS. Ceci suggère que Staf pourrait être préférentiellement impliqué dans les phénomènes de transcription bidirectionnelle. Des études

supplémentaires sont à réaliser afin de déterminer la fonctionnalité des sites SBS présents dans ces régions et la fonction du facteur hStaf/ZNF143 dans la transcription des gènes organisés de cette façon particulière.

Parmi les promoteurs identifiés, 30% sont associés à des gènes codant des protéines sans fonction attribuée lors de notre étude. L'analyse des fonctions des autres gènes montre que le facteur hStaf/ZNF143 ne semble pas activer préférentiellement des gènes codant des protéines impliquées dans une fonction spécifique. Cependant, les deux classes fonctionnelles regroupant les gènes codant des protéines de liaison à l'ADN et des facteurs de transcription ainsi que les gènes codant des protéines impliquées dans la synthèse, la maturation et la dégradation des protéines sont prépondérantes et représentent chacune environ 20% des gènes à fonctions attribuées.

CHAPITRE 2

hStaf/ZNF143 participe à la régulation de l'expression du gène TFAM humain

CHAPITRE 2 : hStaf/ZNF143 participe à la régulation de l'expression du gène TFAM humain

1. Introduction :

Parallèlement à l'analyse de la distribution des sites de liaison de hStaf/ZNF143 dans le génome humain, nous nous sommes intéressés à la caractérisation fine d'un promoteur potentiellement régulé par hStaf/ZNF143. Il a été observé que le promoteur du gène TFAM possède un site potentiel de fixation de hStaf/ZNF143 (Ishiguchi et al., 2004). Le gène nucléaire TFAM code une protéine de 25 kDa, appelée mtTFA ou Tfam (Transcription Factor Mitochondrial A), initialement identifiée comme étant un facteur de transcription spécifique de l'ADN mitochondrial (ADNmt) chez les mammifères (Fisher et Clayton, 1985; Fisher et Clayton, 1988). Cette protéine possède deux domaines caractéristiques de la famille des protéines HMG (High Mobility Group) (Parisi et Clayton, 1991).

L'ADNmt est circulaire, en copies multiples et porte les gènes de 2 ARNr, de 22 ARNt et de 13 des 90 protéines impliquées dans la chaîne respiratoire (pour une revue voir (Falkenberg et al., 2007). L'un des brins de la molécule d'ADN est nommé brin lourd ou H et l'autre brin léger ou L de part leur différence de densité sur un gradient de chlorure de césium. Chez l'homme, chaque brin porte un seul promoteur : LSP (Light-Strand Promoteur) et HSP (High-Strand Promoter) spécifique du brin L et H respectivement. Les gènes mitochondriaux sont transcrits sous forme de deux longs ARN polycistroniques à partir de ces deux promoteurs. Ces transcrits sont ensuite maturés afin de libérer les ARNm, ARNr et ARNt matures. La transcription des gènes mitochondriaux nécessite la présence de l'ARN polymérase mitochondriale (POLRMT), du facteur de transcription Tfam et de l'un des deux facteurs de transcription mitochondrial B (TFB1M ou TFB2M) (Fisher et Clayton, 1988; Parisi et Clayton, 1991; Falkenberg et al., 2002; McCulloch et Shadel, 2003).

Tfam se lie spécifiquement à l'ADNmt sur une séquence localisée juste en amont des deux promoteurs et modifie la conformation de cette région (Fisher et al., 1989; Fisher et al., 1992). Le facteur Tfam est également essentiel à la réplication de l'ADNmt puisque celle-ci s'initie à partir d'une amorce issue de la maturation du transcrit généré à partir du promoteur LSP (Chang et Clayton, 1985; Chang et al., 1985). Les 1000 à 10 000 copies d'ADNmt présentes

dans une cellule somatique de mammifères sont regroupées au sein de plusieurs centaines de structures compactes appelées nucléoides (Garrido et al., 2003). Vingt et une protéines dont Tfam font parties des nucléoides humains (Alam et al., 2003; Wang et Bogenhagen, 2006). Parallèlement à son rôle dans la transcription et dans le compactage de l'ADNmt, Tfam intervient également dans la régulation du nombre de copies de l'ADNmt chez les mammifères (Ekstrand et al., 2004).

De précédentes études réalisées sur la région promotrice -86/+20 du gène humain TFAM ont montré que l'activité de ce promoteur était hautement dépendante de la présence de sites de fixation des facteurs Sp1 et NRF-1 et 2 (Nuclear Respiratory Factors 1 et 2) (Virbasius et Scarpulla, 1994). Cependant, l'observation, par une autre équipe, de la présence d'un site SBS potentiel dans ce promoteur, nous a amené à examiner le rôle de hStaf/ZNF143 dans la régulation de la transcription de ce gène (Ishiguchi et al., 2004). Des expériences de transfections transitoires nous ont permis de déterminer une région minimale, de taille supérieure à celle analysée lors de l'étude de Virbasius *et al*, permettant une transcription efficace du gène rapporteur codant la luciférase. L'examen de l'empreinte phylogénique a permis de mettre en évidence la présence de sept motifs conservés dans les promoteurs orthologues de mammifères. La recherche *in silico* de sites de liaison de facteurs de transcription a révélé la présence des sites de fixation des facteurs Sp1 et NRF-2 précédemment identifiés ainsi que cinq autres motifs conservés dont deux sites potentiels pour hStaf/ZNF143. Des expériences de retard sur gel, d'empreintes à la DNase I et d'immunoprécipitation de chromatine ont permis de montrer que hStaf/ZNF143 se lie au promoteur du gène TFAM. Le rôle des différents motifs conservés identifiés a été déterminé en transfectant des cellules COS7 avec des constructions renfermant la région promotrice minimale du gène TFAM mutée dans ces motifs.

2. Résultats :

2.1. Résumé des principaux résultats

La détermination du site d'initiation de la transcription (TSS) du gène TFAM a été réalisée en utilisant la banque de données des sites d'initiation de la transcription DBTSS. Cette analyse a permis d'identifier plusieurs TSS répartis dans une fenêtre de 41 pb. De ce fait, la

numérotation utilisée pour décrire les différentes positions des éléments du promoteur est établie à partir du codon d'initiation de la traduction. Les TSS sont observés dans une région localisée entre les positions -147 et -107 avec un enrichissement en TSS à la position -138. Afin de déterminer la région promotrice minimale capable de diriger efficacement la transcription du gène TFAM, le fragment -1463/+102, ainsi que des sous-fragments correspondant à des délétions progressives au niveau de la région 5', ont été clonés en amont du gène rapporteur de la luciférase. Leur activité respective a été analysée par des expériences de transfections transitoires dans des cellules COS7. La région -416/+102 est suffisante pour permettre la transcription efficace de la luciférase. Les autres régions (-1463/+102 et -1120/+102) présentent une activité équivalente mais inférieure à la région -416/+102. Ceci laisse supposer la présence d'éléments répresseurs dans la région comprise entre les positions -1120 et -417.

L'identification d'éléments régulateurs potentiels dans la région promotrice minimale a été réalisée par l'examen des empreintes phylogéniques. Les régions du génome du chimpanzé, du chien, de la souris et du rat, orthologues à la région non traduite -416/-1 en amont du gène TFAM humain ont été alignées. La comparaison des séquences a mis en évidence six éléments conservés dont les sites de liaison des facteurs Sp1 et NRF-2 précédemment identifiés (Figure 24) (Virbasius et Scarpulla, 1994). Le site de fixation de NRF-1 n'a pas été identifié par cette approche car sa position est différente chez la souris et le rat par rapport aux trois autres organismes (Piantadosi et Suliman, 2006).

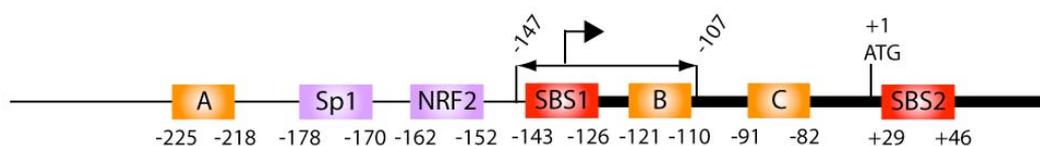


Figure 24 : Représentation schématique de la région promotrice du gène TFAM humain (d'après Gerard et al., 2007b).

Les sites de liaison de hStaf/ZNF143, de Sp1 et NRF-2 ainsi que les motifs conservés A et C sont respectivement représentés en rouge, violet et orange. La région transcrite est représentée par un trait épais et la double flèche indique la région contenant les différents sites d'initiation de la transcription. La numérotation utilisée est établie à partir du codon d'initiation de la traduction.

L'élément conservé aux positions -143/-126, nommé SBS1, correspond à la séquence consensus du site de fixation de hStaf/ZNF143 (Schaub et al., 1997; Schaub et al., 1999b). Trois autres éléments conservés, nommés A, B et C, sont observés respectivement aux positions -225/-218, -121/-110 et -91/-82 (Figure 24). L'analyse de la région -416/-1 par le logiciel MatInspector n'a pas permis d'associer ces motifs conservés à des sites de fixation de facteurs de transcription connus (Quandt et al., 1995). L'analyse de la séquence +1/+102 montre la présence d'un deuxième élément conservé correspondant à la séquence consensus d'un site de fixation de Staf. Ce site, nommé SBS2, est localisé en orientation inverse dans la région codante (Figure 24). On peut noter que nous n'avons pas identifié de boîte TATA, d'Inr et DPE dans la région promotrice du gène TFAM et que la région comprise entre les positions -584/+28 est présente dans un îlot CpG.

La capacité de hStaf/ZNF143 à se fixer à la région promotrice du gène TFAM a été testée *in vitro* par des expériences de retard sur gel et d'empreintes à la DNase I. L'association de hStaf/ZNF143 *in vivo* au promoteur a été testée par immunoprécipitation de chromatine. Les expériences de retard sur gel ont été réalisées avec un fragment d'ADN renfermant les deux sites de fixation putatifs SBS1 et SBS2. Nous avons montré que le DBD de hStaf/ZNF143 se lie à la région promotrice en formant deux complexes retardés différents C1 et C2. Ces complexes sont spécifiques car la liaison de hStaf/ZNF143 au fragment est abolie par l'ajout d'un fragment d'ADN contenant la séquence consensus du SBS tandis qu'un fragment d'ADN non spécifique n'influence pas cette liaison. Le profil de retard sur gel suggère que le complexe C2 contient deux copies du DBD de hStaf/ZNF143. Des expériences similaires ont été réalisées avec des sondes portant des mutations soit dans le site SBS1 (SBS1sub) ou SBS2 (SBS2sub) soit dans les deux sites simultanément (SBS1sub/SBS2sub). Les résultats obtenus montrent que la liaison de hStaf/ZNF143 sur la région promotrice du gène TFAM *in vitro* dépend de l'intégrité des deux sites SBS1 et SBS2.

Afin de localiser expérimentalement les sites de liaison de hStaf/ZNF143 sur cette région promotrice, des expériences d'empreintes à la DNase I ont été réalisées avec le DBD de hStaf/ZNF143 et un fragment d'ADN refermant les deux sites de liaison SBS1 et SBS2 sauvages ou mutés. Une empreinte de 21 pb comprise entre les positions -146/-128 est observée avec les sondes sauvages et SBS2sub et correspond au site de liaison SBS1. Aucune empreinte n'est détectée pour les expériences faites avec la sonde SBS1sub. Ceci est

concordant avec les expériences de retard sur gel pour lesquelles on observe que le DBD se lie de façon spécifique mais avec un taux nettement inférieur sur le SBS2 par rapport au SBS1. Ce faible taux de fixation ne permet pas une identification du site de liaison par la technique d'empreinte à la DNase I.

Après ces expériences *in vitro*, nous avons détecté la liaison de hStaf/ZNF143 sur la région promotrice dans le contexte chromatinien par des expériences d'immunoprécipitation de chromatine à partir de cellules HeLa. L'analyse par PCR semi-quantitative de l'ADN immunoprécipité par des anticorps dirigés contre hStaf/ZNF143 montre que la région promotrice du gène TFAM est sélectionnée spécifiquement. De ce fait, hStaf/ZNF143 est capable de se lier à la région promotrice du gène TFAM.

Nous avons ensuite examiné si les sites de liaison de hStaf/ZNF143 sont requis pour l'activité du promoteur du gène TFAM. La région promotrice -416/+102 sauvage ou portant les mutations individuelles ou simultanées des sites SBS1 et SBS2 a été clonée en amont du gène rapporteur de la luciférase. Des cellules COS7 ont été transfectées avec les différentes constructions et la mesure de l'expression de la luciférase reflète l'effet des mutations sur l'activité de la région promotrice. Nous avons observé que la simple mutation dans le site SBS1 ou SBS2 entraîne une chute de l'activité du promoteur de 63% et 55% respectivement par rapport à l'activité du promoteur sauvage. Une diminution drastique (90%) est obtenue lorsque le promoteur est muté dans les deux sites SBS. Ces résultats montrent que l'activité du promoteur de base du gène TFAM est dépendante des deux sites de fixation de hStaf/ZNF143.

L'analyse comparative des séquences orthologues de la région promotrice -416/+1 a montré l'existence de cinq autres motifs conservés. Deux de ces motifs, correspondant aux sites de fixation de Sp1 et NRF-2, participent à la régulation de l'activité du promoteur du gène TFAM (Virbasius et Scarpulla, 1994). Nous avons étudié le rôle des trois autres éléments conservés par des expériences de transfections transitoires dans des cellules COS7. La région -416/+102 sauvage ou mutée individuellement dans les motifs A, B, C ainsi que dans les sites de fixation de Sp1 et NRF-2 a été clonée en amont du gène de la luciférase. La mesure de l'expression de la luciférase montre, comme attendu, que la mutation des sites de fixation de Sp1 et NRF-2 entraîne une chute de l'activité du promoteur et cette activité est respectivement de 46% et 57% par rapport à l'activité du promoteur sauvage. Les

constructions mutées dans les motifs A ou C présentent un taux d'expression de la luciférase correspondant à 70 et 75% de l'activité de la construction sauvage. Il faut noter que la mutation du motif B n'influence pas l'activité du promoteur. Ainsi, l'activité du promoteur du gène TFAM est contrôlée par une combinaison d'au moins six motifs régulateurs.

2.2. *Publication : « Transcription factor hStaf/ZNF143 is required for expression of the human TFAM gene »*

*Marie-Aline Gérard, Alain Krol, and Philippe Carbon
Gene, 2007*

Transcription factor hStaf/ZNF143 is required for expression of the human TFAM gene

Marie-Aline GÉRARD, Alain KROL and Philippe CARBON

Gene, 2007, Vol. 401, Pages 145-153

Pages 107-115 :

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

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

<http://dx.doi.org/10.1016/j.gene.2007.07.011>

La version imprimée de cette thèse peut être consultée à la bibliothèque ou dans un autre établissement via une demande de prêt entre bibliothèques (PEB) auprès de nos services :

<http://www-sicd.u-strasbg.fr/services/peb/>

3. Discussion

Nous avons observé qu'une région de 416 pb comprise entre les positions -416/+102 est essentielle pour l'activité du promoteur de base du gène TFAM. Une analyse par mutagénèse montre que six motifs A, C, SBS1, SBS2, Sp1 et NRF-2, conservés entre l'homme, le chimpanzé, le chien, la souris et le rat, sont des éléments régulateurs fonctionnels.

L'implication des sites de liaison des facteurs Sp1 et NRF-2 dans l'activité du promoteur du gène TFAM a été précédemment observée en analysant une région promotrice de 106 pb comprise entre les positions -218/-133 selon notre numérotation (Virbasius et Scarpulla, 1994). Un site de liaison de NRF-1, important dans la régulation du gène TFAM, avait également été identifié dans cette précédente étude (Virbasius et Scarpulla, 1994). Notre approche basée sur l'empreinte phylogénique ne nous a pas permis de mettre en évidence ce site NRF-1 car sa position n'est pas conservée dans le promoteur du gène TFAM des cinq mammifères analysés. La position du site NRF-1 est équivalente chez l'homme, le chien et le chimpanzé mais ce site est situé environ 100 pb en amont dans le promoteur de souris et de rat (Piantadosi et Suliman, 2006).

La séquence de l'élément A (₋₂₁₈GTGGGCGG₂₂₄) ressemble à la séquence consensus du site de fixation de Sp1 GGGGCGGGG (Li et al., 2004). Il se peut donc que le motif A soit un site de reconnaissance potentiel pour Sp1. Le motif C ne présente pas de similitude avec la séquence d'un site de fixation d'un facteur de transcription connu.

Nous avons également mis en évidence la présence de deux sites de liaison de hStaf/ZNF143, SBS1 et SBS2, fonctionnellement importants pour la régulation de l'activité du promoteur du gène TFAM. Cependant, lors des expériences d'empreinte à la DNase I, seul le site SBS1 est protégé. Ceci suggère que la fixation de hStaf/ZNF143 sur le site SBS2 est stabilisée par des interactions avec d'autres facteurs cellulaires.

Ainsi, l'analyse fonctionnelle du promoteur du gène TFAM montre l'importance du facteur de transcription hStaf/ZNF143 dans la régulation de la transcription du gène TFAM. Récemment, il a été montré que hStaf/ZNF143 activait la transcription de sept gènes de protéines et notre analyse à l'échelle du génome humain a permis d'identifier 1175 SBS

conservés dans 938 régions promotrices chez l'homme, la souris, le rat et le chien (Kubota et al., 2000; Mach et al., 2002; Saur et al., 2002; Barski et al., 2004; Di Leva et al., 2004; Grossman et al., 2004; Ishiguchi et al., 2004; Myslinski et al., 2006). Parmi ces 938 régions promotrices, dix-huit sont en amont de gènes nucléaires codant des protéines mitochondriales (Myslinski et al., 2006). Il s'agit de douze protéines ribosomiques, du facteur de terminaison de la traduction 1, de l'ARN polymérase mitochondriale et de trois translocases des membranes mitochondriales internes et externes. L'implication de hStaf/ZNF143 dans la régulation de l'expression du gène TFAM, ainsi que ces dix-huit gènes de protéines potentiellement régulés par ce facteur, suggère que le facteur de transcription hStaf/ZNF143 joue un rôle important dans le contrôle nucléaire du fonctionnement des mitochondries des cellules mammifères.

CHAPITRE 3

**hStaf/ZNF143 est impliqué dans la transcription
du gène BUB1B humain
régulé au cours du cycle cellulaire**

CHAPITRE 3 : hStaf/ZNF143 est impliqué dans la transcription du gène BUB1B humain régulé au cours du cycle cellulaire

1. Introduction :

Le promoteur du gène BUB1B renferme plusieurs sites potentiels de liaison pour des facteurs de transcription variés dont le facteur de transcription hStaf/ZNF143 (Seike et al., 2002). Nous nous sommes donc intéressés à la caractérisation fine du promoteur de ce gène qui code une protéine nommée BubR1 (Hardwick et al., 2000). Cette protéine BubR1 est un des facteurs essentiels du mécanisme de contrôle de la fidélité de ségrégation des chromosomes au cours de la division cellulaire (Millband et Hardwick, 2002). Elle fait partie du complexe de contrôle mitotique MCC (Mitotic Checkpoint Complex) qui renferme également les protéines Mad2, Bub3 et Cdc20 (Sudakin et al., 2001).

Au cours de la mitose, la transition de la métaphase à l'anaphase est assurée par un complexe protéique, nommé APC (Anaphase-Promoting Complex), renfermant une activité ubiquitine ligase E3 (pour une revue voir (Peters, 2006). Ce complexe, activé par le cofacteur Cdc20, ubiquitine certains facteurs clés de la régulation du cycle cellulaire, tels que la cycline B et la sécurine, entraînant leur dégradation par le protéasome (pour des revues voir (Peters, 2006; Musacchio et Salmon, 2007). La séparation des chromatides sœurs ne se réalise que lorsque tous les chromosomes sont alignés au niveau de la plaque équatoriale. Ceci se traduit par une certaine tension mécanique au niveau des centromères qui entraîne la déstabilisation du complexe MCC, la libération de Cdc20 et l'activation du complexe APC (Nicklas et al., 1995; Skoufias et al., 2001). Une récente étude a montré que ce complexe APC peut également ubiquitiner Cdc20 ce qui entraîne la dissociation du complexe MCC (Reddy et al., 2007; Stegmeier et al., 2007). Un dysfonctionnement de ce point de contrôle peut conduire à une mauvaise répartition des chromosomes et à une instabilité chromosomique (pour une revue voir (Bharadwaj et Yu, 2004).

Le gain ou la perte d'un chromosome sont souvent observés dans de nombreux types de cancers (Lengauer et al., 1998). Il a été observé chez la souris que l'absence d'un des deux allèles du gène BUB1B est associée au développement de cancers du poumon et du colon (Dai et al., 2004; Rao et al., 2005). L'analyse du niveau d'expression du gène BUB1B dans des

cellules synchronisées a montré que le transcrit de ce gène n'est pas détectable au cours de la phase G1 du cycle cellulaire alors qu'il est très abondant en fin de phase G2 et au début de la mitose (Davenport et al., 1999; Whitfield et al., 2002). L'expression dépendante du cycle cellulaire explique que la distribution du transcrit se limite aux tissus se divisant abondamment comme le thymus ou la moelle osseuse (Li et al., 1999). BUB1B est donc un gène dont la transcription est régulée durant le cycle cellulaire. Nous nous sommes donc intéressés à l'expression de ce gène potentiellement dépendant du facteur hStaf/ZNF143.

Des expériences de transfections de cellules en culture avec différentes constructions renfermant des promoteurs tronqués nous ont permis de déterminer la région minimale du promoteur requise pour la transcription du gène BUB1B. La recherche *in silico* et par la méthode d'empreinte phylogénique d'éléments promoteurs dans cette région a révélé la présence d'un élément Inr, de deux sites potentiels pour le facteur hStaf/ZNF143 ainsi que deux éléments homologues aux éléments répresseurs CDE (cell cycle dependant element) et CHR (cell cycle genes homology regions). Des expériences de retard sur gel et d'immunoprécipitation de chromatine ont permis de montrer que hStaf/ZNF143 se lie au promoteur du gène BUB1B. Le rôle de ce facteur et des éléments CDE et CHR, dans la régulation de la transcription du gène BUB1B au cours du cycle cellulaire, a été analysé par des expériences de transfections transitoires dans des cellules COS7 et SL2.

2. Résultats :

2.1. Résumé des principaux résultats

La détermination du site d'initiation de la transcription du gène BUB1B a été réalisée en interrogeant la banque de données de sites de transcription DBTSS (Suzuki et al., 2002). Il est apparu que ce gène présente de multiples TSS répartis dans une région de 75 pb. De ce fait, la numérotation utilisée pour décrire les différentes positions des éléments du promoteur est établie à partir du site d'initiation de la traduction. Ces différents TSS sont localisés entre les positions -198/-124 avec un enrichissement à la position -178. Afin de déterminer la région minimale suffisante pour une activité efficace du promoteur du gène BUB1B, la région -1185/-31, ainsi que des fragments correspondant à cette région progressivement délétée au niveau de l'extrémité 5', ont été clonés en amont du gène rapporteur codant la luciférase. Les

transfections transitoires de ces différentes constructions dans des cellules COS7 montrent que les régions -1185/-585 et -305/-107 sont importantes pour l'expression du gène BUB1B.

La comparaison des séquences des régions du génome de la souris, du rat et du chien homologues à la région -1185/-31 du promoteur humain du gène BUB1B a mis en évidence la présence d'éléments transposables Alu et MIR dans la région -1185/-585 (Figure 25). L'analyse de cette région avec le logiciel MatInspector n'a pas permis d'identifier de sites de liaison potentiels pour des facteurs de transcription connus (Quandt et al., 1995).

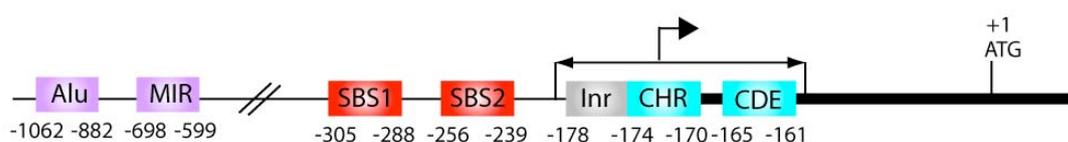


Figure 25 : Représentation schématique de la région promotrice du gène BUB1B humain (d'après Myslinski et al., 2007).

Les éléments transposables Alu et MIR, les sites de liaison de hStaf/ZNF143, l'initiateur et les éléments répresseurs CHR et CDE sont respectivement représentés en violet, rouge, gris et bleu. La région transcrite est représentée par un trait épais et la double flèche indique la région contenant les différents sites d'initiation de la transcription. La numérotation utilisée est établie à partir du codon d'initiation de la traduction.

Par contre, cinq éléments conservés chez les quatre mammifères ont été observés dans la région -464/-31 (Figure 25). Il s'agit de deux sites de liaison potentiels pour le facteur hStaf/ZNF143 localisés aux positions -305/-288 et -256/-239 nommés respectivement SBS1 et SBS2. Une séquence ACTACAA conservée est présente immédiatement en 5' de chacun de ces sites. Deux motifs possédant une forte identité de séquence avec les éléments répresseurs CDE et CHR sont localisés aux positions -174/-170 et -165/-161 respectivement. Ces éléments sont impliqués dans la répression de l'expression de plusieurs gènes au cours du cycle cellulaire (Lucibello et al., 1995; Zwicker et al., 1995b; Huet et al., 1996; Uchiumi et al., 1997; Fajas et al., 2000; Lange-zu Dohna et al., 2000; Tanaka et al., 2002). De manière surprenante, alors que l'élément CDE est localisé juste en amont de l'élément CHR dans le promoteur de ces gènes, il est situé juste en aval dans le cas du gène BUB1B. La majorité des TSS se situe au niveau de la position -178 qui correspond probablement à un élément Inr car

la séquence $_{-180}\text{TTAAATT}_{-174}$ est similaire à la séquence consensus de l'Inr ($\text{YYA}_{+1}\text{NWYY}$) (Smale et Kadonaga, 2003). On peut noter que nous n'avons pas identifié de boîte TATA et DPE dans la région promotrice du gène BUB1B et que la région comprise entre les positions -415/-50 est présente dans un îlot CpG.

Nous avons tout d'abord étudié la capacité de liaison de hStaf/ZNF143 sur les deux sites SBS potentiels présents dans la région promotrice du gène BUB1B. Pour cela, des expériences de retard sur gel ont été réalisées avec un fragment d'ADN renfermant les sites SBS1 et SBS2. Deux complexes retardés se forment lorsque le DBD de hStaf/ZNF143 est mis en présence de la sonde marquée. Ces deux complexes sont spécifiques puisque l'ajout d'un fragment non marqué renfermant le SBS présent dans le promoteur du gène de l'ARNt^{Sec} n'inhibe pas la formation des complexes tandis qu'un compétiteur non spécifique n'interfère pas avec leur formation. Des expériences similaires ont été réalisées avec trois sondes mutées soit dans le site SBS1 (mSBS1) ou SBS2 (mSBS2) soit dans les deux sites (mSBS1-2). Les résultats obtenus montrent que la liaison de hStaf/ZNF143 est dépendante de l'intégrité des deux sites. Des expériences complémentaires réalisées avec la forme entière du facteur de transcription confirment ces observations.

La liaison de hStaf/ZNF143 sur le promoteur du gène BUB1B dans le contexte chromatinien a été observée par immunoprécipitation de chromatine. Les analyses par PCR montrent que l'ADN immunoprécipité par des anticorps dirigés contre l'extrémité carboxy-terminale de hStaf/ZNF143 renferme le promoteur BUB1B contrairement à l'ADN immunoprécipité par des anticorps issus de l'échantillon pré-immun. Ainsi, hStaf/ZNF143 interagit spécifiquement avec la région promotrice de base du gène BUB1B *in vitro* et *in vivo*.

Nous avons voulu ensuite déterminer l'importance des sites SBS1 et SBS2 et de l'élément similaire à un Inr dans l'activité du promoteur du gène BUB1B. Pour cela, la région promotrice -464/-31 sauvage et mutée dans les sites SBS1 et SBS2 (mSBS1, mSBS2 et mSBS1-2) ou au niveau de l'Inr a été clonée en amont du gène rapporteur de la luciférase. Des cellules COS7 ont été transfectées avec les différentes constructions et la mesure de l'expression de la luciférase reflète l'effet des mutations sur l'activité de la région promotrice. Nous avons observé une diminution de l'activité du promoteur de 22% et 39%, par rapport au promoteur sauvage, lorsque les sites SBS1 et SBS2 sont respectivement mutés. Une chute de 81% de l'activité est obtenue lorsque la région promotrice est mutée au niveau des deux sites

SBS. La mutation de la séquence $_{-180}\text{TTAAATT}_{-174}$ correspondant à un initiateur potentiel entraîne, quant à elle, une réduction de moitié de l'activité du promoteur. La capacité de hStaf/ZNF143 à activer la transcription du gène BUB1B a été confirmée par des expériences de transfections transitoires dans des cellules SL2 de drosophiles dans lesquelles il n'existe pas d'homologues de ce facteur de transcription. Le plasmide rapporteur renfermant la région promotrice sauvage -464/-31 a été cotransfecté avec un vecteur d'expression renfermant la phase codante de hStaf/ZNF143. L'expression de la luciférase est dépendante de la quantité de facteur de transcription présent et peut être 24 fois supérieure au niveau de base obtenu avec un vecteur d'expression vide. Toutes ces données montrent que l'activité du promoteur de base du gène BUB1B est dépendante des deux sites de fixation de hStaf/ZNF143 et d'un élément conservé similaire à un Inr.

L'analyse du niveau d'expression du gène BUB1B dans des cellules synchronisées a montré que le transcrit de ce gène n'est pas détectable au cours de la phase G1 du cycle cellulaire alors qu'il est très abondant au cours de la transition G2/M (Davenport et al., 1999; Whitfield et al., 2002). Nous avons donc analysé l'activité du promoteur du gène BUB1B en fonction du cycle cellulaire. Pour cela, le vecteur rapporteur renfermant la région promotrice sauvage -464/-31 a été introduit dans des cellules COS7 synchronisées à partir du stade G1/S du cycle cellulaire et l'expression de la luciférase a été mesurée au cours du temps après réactivation des cellules. On observe que l'activité du promoteur du gène BUB1B augmente au cours du cycle cellulaire et atteint un taux maximal lorsque les cellules sont en phase G2/M pour ensuite diminuer. Lorsque des expériences similaires sont réalisées avec une construction renfermant la région promotrice -464/-196 contenant uniquement les deux sites SBS, l'expression de la luciférase est élevée dès la phase G1/S et reste constante au cours du cycle cellulaire. Des expériences de retard sur gel montrent que la fixation de hStaf/ZNF143 sur le promoteur du gène BUB1B est similaire avec des extraits nucléaires de cellules en phase G1/S ou G2/M. De plus, l'expression du gène codant hStaf/ZNF143 n'est pas régulée en fonction du cycle cellulaire (Whitfield et al., 2002). Ces résultats montrent que le facteur de transcription hStaf/ZNF143 ne joue pas un rôle essentiel dans la limitation de l'expression du gène BUB1B à la phase G2/M.

Des motifs conservés hautement similaires aux éléments répresseurs CDE et CHR ont été identifiés aux positions -174/-170 et -165/-161 respectivement dans la région promotrice du gène BUB1B. L'importance de ces éléments dans la régulation de la transcription de ce gène

a été analysée en transfectant des constructions renfermant la région promotrice -464/-31 sauvage et mutée au niveau de l'élément CDE (mCDE) ou CHR (mCHR). L'expression de la luciférase est mesurée lorsque les cellules COS7 sont arrêtées en phase G1/S ou G2/M. Comme attendu, l'activité du promoteur sauvage est plus importante (2,2 fois) lorsque les cellules sont en phase G2/M par rapport aux cellules arrêtées en phase G1/S. Lorsque les cellules sont arrêtées en phase G1/S, on observe que les mutations mCDE ou mCHR entraînent une augmentation de l'expression de la luciférase de 2,8 et 3,1 fois respectivement par rapport à la construction sauvage. Les activités mesurées pour ces mutants dans des cellules arrêtées en phase G2/M sont similaires à celle mesurée avec la région promotrice sauvage. Ces résultats suggèrent fortement que les éléments CDE et CHR sont des éléments qui répriment la transcription du gène BUB1B au cours de la phase G1/S.

2.2. Publication : « Transcription of the human cell cycle regulated BUB1B gene requires hStaf/ZNF143 »

*Evelyne Myslinski, Marie-Aline Gérard, Alain Krol, and Philippe Carbon
Nucleic Acids Research, 2007*

**Transcription of the human cell cycle regulated BUB1B gene requires
hStaf/ZNF143**

Evelyne MYSLINSKI, Marie-Aline GÉRARD, Alain KROL and Philippe CARBON

Nucleic Acids Research, 2007, Vol. 35, No. 10, Pages 3453-3464

Pages 128-139 :

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

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

<http://dx.doi.org/10.1093/nar/gkm239>

La version imprimée de cette thèse peut être consultée à la bibliothèque ou dans un autre établissement via une demande de prêt entre bibliothèques (PEB) auprès de nos services :

<http://www-sicd.u-strasbg.fr/services/peb/>

3. Discussion

L'expression du gène BUB1B régulé au cours du cycle cellulaire présente un niveau de transcription indétectable en phase G1 et un fort taux de transcription durant la phase G2/M (Davenport et al., 1999; Whitfield et al., 2002). Nous avons étudié le mode de régulation de la transcription de ce gène. Deux régions comprises entre les positions -1185/-585 et -305/-107 régulent positivement la transcription du gène BUB1B. La région -1185/-585 contient deux éléments transposables mais pas de sites de fixation pour des facteurs de transcription connus. Il a été observé que les éléments transposables peuvent avoir des effets positifs ou négatifs sur la transcription selon leur site d'insertion (Brosius, 1999).

Cinq éléments conservés chez l'homme, le chien, la souris et le rat ont été identifiés dans la région -305/-107. Une analyse par mutagenèse montre que ces éléments sont des éléments régulateurs fonctionnels. Trois d'entre eux régulent positivement la transcription du gène BUB1B tandis que les deux autres éléments sont des répresseurs de la transcription dépendants du cycle cellulaire. En effet, l'élément chevauchant les TSS est probablement un élément Inr qui participe à l'initiation de la transcription. De plus, deux sites de fixation pour le facteur de transcription hStaf/ZNF143 sont localisés dans la région promotrice du gène BUB1B aux positions -305/-288 et -256/-239. Des expériences de retard sur gel, de mutagenèse et de ChIP ont montré que hStaf/ZNF143 pouvait se lier au promoteur et l'importance de ce facteur dans la régulation de la transcription du gène BUB1B a été démontrée par des expériences de transfections transitoires. Il a été observé que hStaf/ZNF143 active la transcription de ce gène indépendamment du cycle cellulaire.

Notre récente étude sur l'identification des SBS à l'échelle du génome humain a montré qu'une séquence ACTACAN est souvent associée aux sites SBS dans les promoteurs de gène de protéines (Myslinski et al., 2006). Une séquence de ce type est localisée juste en amont des deux sites SBS1 et SBS2. Cependant, sa mutation ne perturbe pas l'activité du promoteur du gène BUB1B.

Les éléments participant à la régulation de l'expression du gène BUB1B au cours du cycle cellulaire sont les éléments CDE et CHR identifiés à proximité de l'Inr. Ces éléments sont également impliqués dans la régulation de la transcription d'autres gènes régulés au cours du

cycle cellulaire (Lucibello et al., 1995; Zwicker et al., 1995b; Baldi et al., 1996; Huet et al., 1996; Uchiumi et al., 1997; Li et Altieri, 1999; Fajas et al., 2000; Lange-zu Dohna et al., 2000; Tanaka et al., 2002). De manière surprenante, l'élément CHR est en amont de l'élément CDE dans le promoteur du gène BUB1B alors que cette organisation est inversée dans les promoteurs des gènes précédemment identifiés. Nous avons observé que les gènes RBL2, PLK1 et BIRC5, régulés par les éléments CDE et CHR, possèdent également un site SBS dans leur région promotrice (Baldi et al., 1996; Uchiumi et al., 1997; Li et Altieri, 1999; Fajas et al., 2000; Myslinski et al., 2006).

Les facteurs de transcription Sp1, NF-Y et E4TF1 sont respectivement impliqués dans l'activation de la transcription des gènes CDC25C, CDC2 et aurora A en se fixant à proximité des éléments CDE-CHR (Zwicker et al., 1995a; Huet et al., 1996; Tanaka et al., 2002). L'inhibition de la transcription de ces gènes au cours du cycle cellulaire pourrait être due à la répression de la fonction de ces activateurs (Zwicker et al., 1995a; Huet et al., 1996). Des expériences d'empreinte génomique ont montré que les éléments CDE et CHR sont protégés de manière périodique. Les facteurs CDF-1 et CHF ont été proposés comme interagissant respectivement avec les éléments CDE et CHR des promoteurs des gènes CDC25C et CCNA2 (Liu et al., 1998; Philips et al., 1999). Cependant, aucune étude supplémentaire n'a été réalisée sur le mécanisme potentiel de répression via ces protéines. Récemment, il a été montré que les éléments CDE et CHR pourraient jouer un rôle dans la formation de la structure chromatinienne particulière au niveau des promoteurs des gènes régulés au cours du cycle cellulaire (Coisy et al., 2004). Nous avons démontré que les éléments CDE et CHR sont impliqués dans la répression de la transcription du gène BUB1B au cours du cycle cellulaire mais il reste à caractériser le mécanisme impliqué dans cette répression.

CHAPITRE 4

hStaf/ZNF143 participe à la transcription indépendante du gène SCARNA2

CHAPITRE 4 : hStaf/ZNF143 participe à la transcription indépendante du gène SCARNA2

1. Introduction :

Les résultats présentés dans les trois précédents chapitres, ainsi que les travaux publiés par d'autres équipes, mettent en évidence l'importance du facteur de transcription hStaf/ZNF143 dans la régulation de la transcription des gènes codant des protéines (Kubota et al., 2000; Mach et al., 2002; Saur et al., 2002; Barski et al., 2004; Di Leva et al., 2004; Grossman et al., 2004; Ishiguchi et al., 2004; Myslinski et al., 2006; Gerard et al., 2007b; Myslinski et al., 2007). Cependant, hStaf/ZNF143 a été initialement identifié comme l'activateur de la transcription de gènes d'ARNsn et de type ARNsn (Schuster et al., 1995; Schaub et al., 1997; Myslinski et al., 1998). L'observation de la présence d'un site potentiel de liaison de hStaf/ZNF143 en amont du gène SCARNA2 (mgU2-25/61) nous a amené à étudier le mode de transcription de ce gène (Tycowski et al., 2004).

Le gène SCARNA2 code un ARNnc prédit pour modifier l'ARNsn U2 (Huttenhofer et al., 2001; Tycowski et al., 2004). La O-méthylation en 2' des riboses ainsi que la pseudouridylation sont les modifications majeures présentes dans les ARNr et les ARNsn. Les modifications des ARNr et de l'ARN U6 sont réalisées, au sein du nucléole, par de petites particules ribonucléoprotéiques (RNPsno). Les RNPsno à boîtes C/D dirigent la O-méthylation en 2' tandis que les RNPsno à boîtes H/ACA dirigent la pseudouridylation (pour une revue voir Kiss, 2001). Chaque RNPsno (C/D ou H/ACA) est composée d'un ARNsno particulier qui va guider la modification sur l'ARN cible et de protéines communes soit à l'ensemble des RNPsno à boîtes C/D soit à l'ensemble des RNPsno à boîtes H/ACA. Les ARNsno guides à boîtes C/D renferment les motifs conservés C, C', D et D'. Les éléments C et D localisés respectivement aux extrémités 5' et 3' forment une structure en hélice et les motifs C' et D' sont internes à la molécule (Tycowski et al., 1996; Kiss-Laszlo et al., 1998). Les ARNsno guides à boîtes H/ACA se replient sous la forme de deux structures en tige-boucle. Les éléments conservés H et ACA se localisent respectivement dans la région simple brin intermédiaire et trois nucléotides avant l'extrémité 3' de l'ARN (Balakin et al., 1996; Ganot et al., 1997).

D'autres petits ARNnc possèdent également ces motifs. Il s'agit des petits ARNs retrouvés dans les corps de Cajal (ARNsca). Ces ARNs guident spécifiquement la méthylation en 2' et la pseudouridylation des ARNsn U1, U2, U4 et U5 ainsi que probablement celles de l'ARNsn U12 (Darzacq et al., 2002). Plus particulièrement, l'ARNsca2, produit du gène SCARNA2, est un ARN à boîte C/D prédit pour métyler en position 2' le ribose des nucléotides G25 et C61 de l'ARNsn U2 (Huttenhofer et al., 2001; Tycowski et al., 2004). Les ARNsca sont synthétisés dans le nucléoplasme puis s'accumulent dans les corps de Cajal. Les ARNsca à boîtes H/ACA possèdent une séquence signal responsable de leur localisation dans les corps de Cajal (Richard et al., 2003).

Chez les vertébrés, les gènes codant les ARNsno et ARNsca sont généralement localisés dans les introns de gènes d'ARNm. Après épissage du pré-ARNm, les ARNsno et ARNsca sont libérés lors de l'étape de débranchement de l'intron et maturés par des exonucléases (Bachellerie et al., 2002). Comme cela a été décrit dans le paragraphe 3.2 de mon introduction, les ARNsno U3, U8 et U13 sont, quant à eux, transcrits par la Pol II à partir de leur propre gène (Tyc et Steitz, 1989; Ach et Weiner, 1991). Les travaux de Tycowski *et al* suggèrent que l'ARNsca2 est également synthétisé à partir d'une unité transcriptionnelle indépendante (Tycowski et al., 2004). En effet, l'ARNsca2 coiffé est apparemment codé par une région intergénique en amont de laquelle il a été observé la présence de deux motifs conservés : un motif TATA et un site potentiel de liaison de Staf (Tycowski et al., 2004). Nous nous sommes donc intéressés à l'expression de ce gène potentiellement dépendant du facteur hStaf/ZNF143.

Nous avons montré qu'une région minimale de 161 pb en amont du gène SCARNA2 permettait la transcription indépendante de ce gène. Cette région renferme trois éléments essentiels pour la transcription : un motif TATA, un site de liaison de hStaf/ZNF143 et un nouveau motif qui ne correspond à aucun site de liaison pour un facteur de transcription connu. Des expériences de retard sur gel et d'immunoprécipitation de chromatine ont permis de montrer que hStaf/ZNF143 se lie au promoteur du gène SCARNA2. L'importance du site SBS et des deux autres éléments conservés chez les mammifères dans la transcription du gène SCARNA2 a été montrée en introduisant, dans des cellules COS7, différentes constructions renfermant la région promotrice portant des mutations dans ces éléments conservés. Nous avons également observé que la transcription du gène SCARNA2 se poursuit au moins 112 nts après la position correspondant à l'extrémité 3' de l'ARNsca2 mature.

2. Résultats :

2.1. Résumé des principaux résultats

L'extrémité 5' de l'ARN^{sca2} mature a été déterminée par des expériences d'extension d'amorce et le site potentiel d'initiation de la transcription a pu ainsi être localisé sur le génome (Tycowski et al., 2004). La capacité de la région localisée en amont du gène SCARNA2 à diriger la transcription a été étudiée par des expériences de transfections transitoires de cellules COS7. La région comprise entre les positions -250 et +145 par rapport au TSS, ainsi que des sous-fragments correspondant à des délétions progressives au niveau de la région 5', ont été clonés en amont du gène de la luciférase. Les activités luciférase mesurées pour les fragments -250/+145 et -161/+145 sont équivalentes et environ 150 fois plus élevées que l'activité obtenue avec le vecteur témoin. Une très forte chute d'activité est observée avec la construction renfermant uniquement la région -88/+145. L'expression de la luciférase est égale à celle observée avec le vecteur témoin lorsque la région en amont du gène SCARNA2 est réduite à 25 pb. De ce fait, la région adjacente en 5' au gène SCARNA2 possède une activité promotrice. La région minimale permettant de diriger la transcription est une région de 161 pb et les éléments promoteurs sont localisés au sein d'un fragment de 136 pb compris entre les positions -161 et -26.

Afin de déterminer la nature des éléments promoteurs impliqués dans la transcription du gène SCARNA2, la région -250/-1 du gène SCARNA2 humain a été comparée aux régions orthologues des gènes du chimpanzé, du chien, de la souris et du rat. L'alignement obtenu met en évidence la présence de quatre éléments conservés chez ces cinq mammifères. Comme attendu, le site potentiel de liaison de Staf ainsi que le motif TATA, précédemment identifiés par Tycowski *et al.*, ont été respectivement retrouvés aux positions -70/-53 et -31/-26 (Tycowski et al., 2004) (Figure 26).

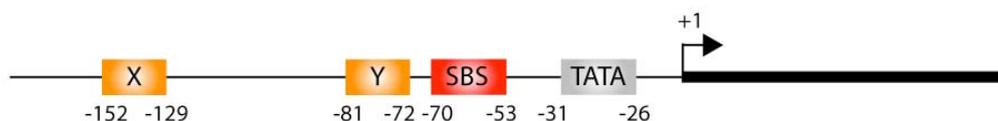


Figure 26 : Représentation schématique de la région promotrice du gène SCARNA2 humain (d'après Gerard et al., 2007a soumis pour publication).

Les éléments conservés X et Y, le site de liaison de hStaf/ZNF143 et l'élément TATA sont respectivement représentés en orange, rouge et gris. La région transcrite est représentée par un trait épais et la flèche indique le site d'initiation de la transcription.

Deux nouveaux motifs conservés, nommés X et Y, ont été respectivement identifiés aux positions -152/-129 et -81/-72 dans la région promotrice (Figure 26). L'analyse de la région -250/-1 par le logiciel MatInspector n'a pas permis d'associer des sites de liaison de facteurs de transcription connus à ces motifs (Quandt et al., 1995). On peut noter que l'analyse de la région à proximité du TSS n'a pas mis en évidence la présence d'un Inr ou d'un élément promoteur DPE.

Des expériences de retard sur gel et d'immunoprécipitation de chromatine ont été réalisées afin de déterminer si le facteur de transcription hStaf/ZNF143 est capable de se lier au SBS potentiel identifié aux positions -70/-53 dans la région promotrice du gène SCARNA2. Dans le cas des expériences *in vitro*, un complexe retardé est observé lorsque hStaf/ZNF143 est mis en présence d'un fragment d'ADN marqué renfermant le site SBS. La liaison est spécifique puisque le complexe disparaît lorsqu'on ajoute un compétiteur spécifique non marqué renfermant un site SBS consensus tandis que la liaison n'est pas affectée par l'ajout d'un fragment ne renfermant pas de sites SBS. L'intégrité du site SBS est requise pour la liaison de hStaf/ZNF143 puisque la mutation de ce site empêche la formation du complexe retardé.

La recherche de la liaison *in vivo* de hStaf/ZNF143 en amont du gène SCARNA2 a été réalisée par ChIP-PCR à partir de cellule HeLa. La région -87/+228, renfermant le SBS identifié, est recherchée par PCR dans l'ADN immunoprécipité par des anticorps dirigés contre l'extrémité carboxy-terminale du facteur de transcription et dans l'ADN traité sans anticorps. On observe que cette région est spécifiquement amplifiée par PCR à partir du

premier échantillon. Ces deux types d'expériences montrent que hStaf/ZNF143 se lie *in vitro* et *in vivo* au SBS localisé aux positions -70/-53 dans le promoteur du gène SCARNA2.

Nous avons voulu ensuite vérifier la fonctionnalité du SBS présent dans la région promotrice du gène SCARNA2. Pour cela, la construction contenant la région -250/+145 clonée en amont du gène de la luciférase a été mutée au niveau du site SBS (SBSsub). L'effet de cette mutation sur l'activité du promoteur est mesuré en dosant les variations d'expression de la luciférase dans les extraits de cellules COS7 transfectées. On observe une chute de 70% de l'activité du promoteur muté par rapport à l'activité obtenue avec la construction renfermant la région sauvage. Ceci démontre que le SBS présent dans le promoteur du gène SCARNA2 est fonctionnel et que hStaf/ZNF143 intervient dans la régulation de l'expression de ce gène.

Des expériences similaires ont été réalisées afin d'évaluer l'importance des trois autres éléments conservés identifiés dans le promoteur du gène SCARNA2. Des mutations supplémentaires ont été réalisées à partir de la construction SBSsub afin d'obtenir les double, triple et quadruple mutants suivant : SBSsub-TATAsub, Ysub-SBSsub-TATAsub et Xsub-Ysub-SBSsub-TATAsub. On observe que la mutation simultanée de l'élément TATA et du SBS provoque une chute de 80% de l'activité du promoteur par rapport à la région sauvage. La construction Ysub-SBSsub-TATAsub présente une activité équivalente à celle du double mutant. Ceci indique que l'élément Y n'est pas essentiel à l'activité du promoteur. La mutation supplémentaire au niveau de l'élément X induit une réduction drastique de l'expression de la luciférase qui devient quasiment équivalente au niveau d'expression obtenu avec le vecteur contrôle. Ces résultats montrent que la transcription du gène SCARNA2 est dépendante de la présence de trois éléments régulateurs dont l'élément X et le SBS localisés respectivement aux positions -152/-129 et -70/-53.

La formation de l'extrémité 3' des ARNs U1, U2, U4 et U5 ainsi que des ARNsno U3, U8 et U13 nécessite la participation d'un élément agissant en cis, nommé boîte 3', localisé en aval de l'extrémité mature de l'ARN. Cette boîte n'est pas un élément essentiel à la terminaison de la transcription de ce type de gènes puisque la Pol II poursuit la transcription au-delà de cet élément (Cuello et al., 1999; Hernandez, 2001; Medlin et al., 2003; Baillat et al., 2005). L'analyse de la région localisée en aval de la position correspondant à l'extrémité 3' de l'ARNsca2 mature ne montre pas l'existence d'une séquence similaire à une boîte 3'.

Cependant, on observe un élément de 15 pb conservé dans les régions orthologues des cinq mammifères analysés.

Cet élément, nommé élément 3', est localisé immédiatement en aval de la séquence correspondant à l'extrémité 3' de l'ARNsca2 mature entre les positions +421 et +435. Un fragment renfermant la région promotrice de 250 pb, le gène SCARNA2 et une région de 112 pb en aval de l'extrémité 3' de l'ARNsca2 mature a été cloné afin d'analyser l'ARN synthétisé après transfections dans des cellules HeLa. Pour différencier l'ARNsca2 exogène de l'ARNsca2 endogène, le gène SCARNA2 ainsi cloné a été réduit de 152 pb par délétion de la région +197/+349 pour produire le minigène SCARNA2 (miniSCARNA2). Le rôle de l'élément 3' dans la transcription du gène SCARNA2 a été analysé en transfectant des cellules COS7 avec une construction renfermant le miniSCARNA2 délété de l'élément 3' (miniSCARNA2 Δ 3'). Les ARNs obtenus suite à la transfection des deux types de constructions, miniSCARNA2 et miniSCARNA2 Δ 3', ont été analysés par RT-PCR avec le couple d'oligonucléotides P1 et P2 spécifiques d'une région interne au miniARNsca2 et le couple d'oligonucléotides P1 et P3 amplifiant un miniARNsca2 potentiellement étendu de 112 pb en 3'. Comme attendu, une amplification est observée avec les oligonucléotides P1P2 lorsque la RT-PCR est réalisée sur des extraits cellulaires transfectés par la construction miniSCARNA2. De manière plus surprenante, un produit PCR est également amplifié avec le couple P1P3 indiquant que la transcription se poursuit au moins 112 pb en aval de l'extrémité 3' de l'ARNsca2 mature. Des signaux similaires sont obtenus avec les extraits de cellules transfectées avec la construction miniSCARNA2 Δ 3'. Ainsi, l'élément conservé en 3' du gène SCARNA2 n'intervient pas dans la terminaison de la transcription par la Pol II. Cependant, on observe une diminution d'environ 50% de la quantité de produits amplifiés par PCR lorsque l'élément 3' est délété. Ceci suggère que l'élément 3' pourrait intervenir dans l'efficacité de la transcription du gène SCARNA2 ou dans la stabilité de l'ARNsca2.

2.2. Manuscrit : « *The scaRNA2 gene is encoded by an independent unit and its transcription is regulated by hStaf/ZNF143* »

Marie-Aline Gérard, Alain Krol, and Philippe Carbon
Soumis pour publication dans Nucleic Acids Research

The scaRNA2 gene is encoded by an independent unit and its transcription is regulated by
hStaf/ZNF143

Marie-Aline Gérard, Alain Krol and Philippe Carbon*

Architecture et Réactivité de l'ARN, Université Louis Pasteur de Strasbourg, CNRS,
IBMC, 15 rue René Descartes, 67084 Strasbourg, France

*Address for correspondence: Tel: (33) 3 88417064 ; fax: (33) 3 88602218. E-mail:
P.Carbon@ibmc.u-strasbg.fr

ABSTRACT

The C/D box scaRNA2 is predicted to guide specific 2'O-methylation of U2 snRNA at positions G25 and C6. In contrast to other SCARNA genes, SCARNA2 appears to be independently transcribed. By transient expression of SCARNA2-reporter gene constructs, we have demonstrated that the elements responsible for SCARNA2 promoter activity are contained within a 161 bp region upstream of the transcription start site. We have identified three cross species conserved promoter elements, a TATA motif, an hStaf/ZNF143 binding site and a novel element that are required for full promoter activity. Binding of the human hStaf/ZNF143 transcription factor to its target sequence is required for promoter activity, suggesting that hStaf/ZNF143 is a fundamental regulator of the SCARNA2 gene. Additionally, we showed that its transcription continues past the 3'-end of the mature RNA and that it is independent of the presence of a conserved element residing beyond the mature 3' end. This work represents the first report demonstrating that a scaRNA can be produced from an independent transcription unit.

INTRODUCTION

Pseudouridines and 2'-O-methylriboses are the most two common modifications in ribosomal RNAs and small nuclear RNAs (for reviews see (1,2). In rRNAs and in U6 snRNA, they are performed in the nucleolus by small nucleolar ribonucleoprotein particles (snoRNPs) (for reviews see) ((3,4). Box C/D snoRNPs direct the 2'-O-methylation of riboses, while H/ACA snoRNPs catalyze pseudouridylation (reviewed in) ((5-8). Box C/D snoRNPs are composed of a sequence-specific guide RNA associated with a set of at least four proteins: fibrillarin, Nop56, Nop58 and 15.5 kDa (4). Each box C/D snoRNA contains the conserved C, C', D and D' box motifs. Typically, box H/ACA snoRNAs exhibit a common hairpin-hinge-hairpin-tail structure with the H motif ANANNA in the single-stranded hinge region, and the ACA triplet located three nucleotides upstream of the 3' termini (4). H/ACA snoRNAs associate with the four H/ACA snoRNP proteins, the pseudouridine synthase dyskerin/Cbf5, Gar1, Nhp2 and Nop10 (reviewed in) ((9).

U6 snRNA is synthesized by RNA polymerase III (Pol III) and is modified and assembled with proteins in the nucleus. In contrast, U1, U2, U4 and U5 (and likely U11 and U12) snRNAs are synthesized by RNA polymerase II (Pol II) and possess a distinct maturation pathway (10-12). In the nucleus, the Pol II specific snRNAs acquire a monomethyl guanosine (7mG) cap structure and are extended for at least 250 nucleotides beyond the 3'-end of the mature RNA (13). Stable pre-snRNAs intermediate bearing a short 3'-end extension are generated by a transcription-linked cleavage event that requires a cis-acting element called the 3'-box (14-16). After export to the cytoplasm, the pre-snRNA is assembled with the Sm proteins to form a pre-snRNP. Binding of the Sm proteins is needed for dimethylation of 7mG to form the 2,2,7-trimethylguanosine cap structure and for removal of the 3' trailer. Subsequently, snRNPs are imported to the nucleoplasm and transiently accumulate in Cajal bodies where the snRNAs undergo specific 2'-O-methylation and pseudouridylation directed by scaRNAs which accumulate in Cajal bodies (17-19). Vertebrate guide scaRNAs contain either one single (C/D or H/ACA) or two snoRNA boxes. The composite structures H/ACA-H/ACA, C/D-H/ACA, C/D-C/D or (C/D-C/D like box) were observed in scaRNAs harboring two associated boxes.

scaRNAs are synthesized in the nucleoplasm and specifically targeted to the Cajal bodies, by a common sequence motif (the Cajal body-specific signal) in the case of H/ACA scaRNAs (20).

In vertebrates, sequences encoding sno- and scaRNAs are generally located in introns of host genes (21). These intronic sno/scaRNAs are processed from the spliced and debranched host introns. U3 and most likely U8 and U13 snoRNAs, however, are transcribed from independent units by RNA polymerase II. In vertebrates, the basal promoter of the U3 snoRNA genes include an essential promoter sequence element (PSE) also found in the basal promoter of snRNA genes (22,10). Associated to the basal promoter of snRNA and snoRNA genes, the distal sequence element (DSE) plays a major role in transcription efficiency (10). The DSE are essentially composed of two functional submotifs, the octamer and the Staf Binding Site, that can reside simultaneously or individually (23,24,10). The capped scaRNA2 (mgU2-25/61) and scaRNA17 (mgU12-22/U4-8) are apparently encoded in intergenic regions and the presence of conserved motifs in the upstream region of the genes suggest that scaRNA2 and scaRNA17 are synthesized from independent units (25). The capped scaRNA2 contains a C/D box associated to a C/D like box domain. It is predicted to guide the specific 2'O-methylation of the U2 snRNA G25 and C61 positions (25). To understand how scaRNA2 synthesis is controlled, we have cloned the 5'-flanking region of the gene and demonstrated that 161 bp of 5'-flanking region are sufficient for efficient promoter activity. In this region, we have mapped three essential promoter elements and identified hStaf/ZNF143 as a critical transactivator interacting with the cognate cis-element. Furthermore, we have shown that transcription of the SCARNA2 gene continues for at least 112 nucleotides beyond the region encoding the mature scaRNA2.

MATERIALS AND METHODS

Reporter constructs

The human SCARNA2 promoter fragment -250/+532 was PCR amplified from genomic DNA and cloned to pGEMTeasy (Promega) to create SCARNA2 -250/+532. The SCARNA2 promoter fragments -250/+145, -161/+145, -88/+145 and -25/+145 were PCR amplified from SCARNA2 -250/+532 using direct and reverse primers incorporating SacI and BamHI sites, respectively. The PCR products were cloned 5' to the luciferase reporter gene in the SacI/BamHI cut promoterless pFLASH1 vector (Synapsis) to generate SCARNA2 -250/+145 Luc, SCARNA2 -161/+145 Luc, SCARNA2 -88/+145 Luc and SCARNA2 -25/+145 Luc. The SCARNA2 -250/+532 minigene (miniSCARNA2) was obtained as follows. BamHI sites were introduced at positions +197 and +349 of SCARNA2 -250/+532 and the resulting mutant was digested with BamHI and self-ligated under diluted conditions. Substitution and internal deletion mutants were generated with the QuickChange II XL site-directed mutagenesis kit (Stratagene). Substitution mutants giving rise to constructs Xsub, Ysub, SBSsub and TATA sub are listed in Figure 5B. Deletion of the +421 to +435 fragment (positions relative to the wt gene) in miniSCARNA2 yielded the SCARNA2 -250/+532 minigene Δ 3' (miniSCARNA2 Δ 3'). All constructs were verified by automated DNA sequencing.

Transfection and dual luciferase reporter assays

COS-7 cells (8×10^4 cells) were transfected using the Lipofectamine 2000 reagent (Invitrogen). Cells received 500 ng of the SCARNA2 promoter plasmid containing the firefly luciferase reporter with 50 ng of pRL-TK (Promega) as the renilla luciferase internal control. The total amount of DNA was kept constant at 800 ng with carrier DNA. After 48h, cells were lysed and the activities of both the firefly and renilla luciferases were determined with the dual luciferase assay kit (Promega). A minimum of three independent transfections were performed in duplicate and the firefly luciferase activity was normalized to that of the Renilla luciferase.

Transfection and RT-PCR assays

COS-7 cells (8×10^4 cells) were transfected with 500 ng of SCARNA2 promoter plasmid, 25 ng of 5S maxigene as internal control and carrier DNA to bring up the total DNA amount to 800 ng/assay. Cells

were collected after 48h and total RNA was extracted using the TRI-REAGENT (Euromedex). After DNase I treatment, half of the total RNA preparation was used for first strand DNA synthesis with the Finzymes reverse transcription reagents in the presence of N9 random oligonucleotide. Decreasing amounts of the resulting cDNA (1/10, 1/100 and 1/500) were used as the starting material for the PCR reaction. Absence of contaminating genomic DNA in the RNA preparations was verified by omitting the reverse transcriptase in control reactions. Typically, PCR reactions were performed in 25 μ l in the presence of 2 μ Ci (α -³²P)-dCTP (3000 Ci/mmol) and primer pairs for specific cDNA amplification of miniSCARNA2, miniSCARNA2 Δ 3', 5S maxi (transfection efficiency control) or U3 snoRNA (RNA recovery control). Test primer pairs for miniSCARNA2 and miniSCARNA2 Δ 3' hybridized to sequences +9/+25, +352/+374 and +9/+25, +511/+532 of the wt gene, respectively. For the PCR controls with the 5S maxi and U3 snoRNA, primer pairs hybridized to positions +2/+25, +112/+129 of 5S maxi and to positions +49/+72, +142/+165 of the U3 snoRNA gene. Cycling parameters were 94 °C for 5 min, 25 cycles at 94 °C for 30 s, 62 °C for 30 s, 72 °C for 1 min and 72 °C for 5 min. Half of the samples were resolved by native polyacrylamide gel electrophoresis and revealed by autoradiography. Results were quantitated with a Fuji Bioimage Analyzer.

hStaf/ZNF143 preparation and DNA binding assays

Full length hStaf/ZNF143 was synthesized by in vitro coupled transcription-translation with the TnT system (Promega). 50 μ l reactions were programmed with 1 μ g of pSK(-)-ZNF143 (26). The various probes containing the wild type and mutant versions of the Staf Binding Site in the SCARNA2 promoter were obtained by PCR amplification of the -88/+145 region, using ³²P-labeled primers. Gel retardation assays were performed essentially as described in (27,23) with 20 fmol of the labeled probe in the presence of 2.5 μ l or 5 μ l of programmed lysate.

ChIP assay and PCR analysis

The ChIP procedure was essentially as described in (28,29). Purified immunoprecipitated DNA was analysed in 25 μ l PCR reactions in the presence of 3 μ Ci (α -³²P) dCTP (3000 Ci/mmol) with the test primer pair CCTGTGCTCGGTGGTTTC and GCAGGAGGAGAGCTTTTCAT, specific for the SCARNA2 promoter and complementary to positions -87/-69 and +228/+247 of the promoter,

respectively. The tRNA^{Sec} gene test primer pairs hybridized to sequences -391/-365 and -205/-181. For negative controls of the ChIP assay, the PP1 primer pair hybridizing to unique regions lying 2.4 kbp upstream of the tRNA^{Sec} was used. 1/500 and 1/2000 of the immunoprecipitated DNA were used in PCR analysis. Decreasing amounts of input DNA (1/10000, 1/25000 and 1/100000) were used to determine the linear range of PCR reaction for each primer pair. Cycling parameters were 95°C for 3min, 35 cycles at 95°C for 30s, 52-72°C (depending on each primer pair) for 30s, 72°C for 30s and one cycle at 72°C for 5 min.

RESULTS

Transfection assays detect promoter activity in the 5' flanking sequence of the human SCARNA2 gene

To identify whether the 5'-flanking DNA of the SCARNA2 gene exhibits promoter activity, sequences were fused to a firefly luciferase reporter gene. As the capped 5'-end of the mature scaRNA2 was previously mapped by primer extension (25), various constructs containing the +1 of the scaRNA2 were generated. Constructs containing progressive truncated portions of the -250/+145 5'-flanking regions were transiently transfected into COS-7 cells. The luciferase activity of the resulting cell extracts was further measured (Figure 1). Analysis of the firefly luciferase values, normalized to the activity of the renilla luciferase control, revealed that similar promoter activity was observed with fragments containing 161 bp or more upstream of the +1 (compare the luciferase activity of fragments -250/+145 and -161/+145 in Figure 1). In contrast, the luciferase activity of the -88/+145 construct was about 8-fold lower than that of -161/+145. Construct -25/+145, which contains only 25 bp of the SCARNA2 gene flanking sequence, led to abrogation of the luciferase activity, showing that it is not sufficient on its own for promoter activity (Figure 1). Thus, the 5' flanking sequence of the SCARNA2 gene possesses a promoter activity and the minimal promoter sequence can be defined as extending 161 bp upstream of the +1 nucleotide. Furthermore, the elements responsible for promoter activity must be contained in a 136 bp region lying between positions -161 and -26.

Identification of putative cis-elements in the human SCARNA2 promoter

To identify putative elements in the promoter region of the SCARNA2 gene, we used the phylogenetic footprinting approach. A multiple alignment was performed with sequences of the chimpanzee, dog, mouse and rat genomes that are orthologous to the -250/-1 region of the human SCARNA2 gene. This revealed the presence of the four interspecies conserved elements X, Y, SBS and TATA at positions -152/-129, -81/-72, -70/-53 and -31/-26, respectively (Figure 2A). Sequence analysis of the region -250/-1 with the MatInspector software (30) established that the Staf Binding Site (SBS) element located at positions -70/-53 matches the consensus binding site for transcription factor hStaf/ZNF143 (Figure 2B) ((24,31). This analysis did not allow us to identify transcription factor binding sites for the conserved motifs X and Y. Furthermore, the 5'-flanking region of the SCARNA2 gene contains no obvious initiator (Inr) or downstream promoter element (DPE) in the vicinity of the transcription start site.

Collectively, this computational analysis revealed the presence of four conserved motifs in the basal promoter of the human SCARNA2 gene constituting putative elements regulating its expression.

The cellular hStaf/ZNF143 transcription factor binds the cognate element in the human SCARNA2 promoter

Gel mobility shift assays were used to ask whether hStaf/ZNF143 interacts with the putative Staf Binding Site (SBS) located at positions -70/-53 in the SCARNA2 promoter. In the binding assay, the full length hStaf/ZNF143 was used with the ³²P-labeled fragment -159/+88 containing the putative SBS. Figure 3 shows that hStaf/ZNF143 bound to the SCARNA2 promoter and generated one retarded complex (Figure 3, compare lane 1 with lanes 2 and 3). In contrast, we were unable to detect a retarded complex in the binding assay performed with unprogrammed reticulocyte lysate (compare lanes 1 with lanes 3 and 6 in Figure 3). The specificity of binding was confirmed by competition studies. As expected, competition by a 1000-fold molar excess of unlabeled oligonucleotide containing the consensus SBS (31) abolished complex formation (Figure 3, compare lanes 2 and 4). In contrast, an unrelated oligonucleotide was unable to compete for binding to the protein (Figure 3, lane 5). To demonstrate that the hStaf/ZNF143 binding pattern was strictly dependent on the SBS integrity, similar binding assays were carried out with hStaf/ZNF143 and an SBS mutant having the -70

CTCCCA -65 sequence substituted by GATATC (23,32,33). Indeed, this mutation abolished totally formation of the DNA-protein complex (Figure 3, lane 8). Together, these results strongly support the hypothesis that an SBS resides at positions -70/-53 in the promoter and that transcription factor hStaf/ZNF143 specifically binds to this element.

To determine whether hStaf/ZNF143 is associated in vivo with the active SCARNA2 promoter, we performed a chromatin immunoprecipitation assay (ChIP) in HeLa cells. HeLa cells were treated with formaldehyde to cross-link proteins to DNA. After sonication, the cross-linked chromatin was immunoprecipitated with an antipeptide antibody directed against the C-terminal part of hStaf/ZNF143. After immunoprecipitation, enrichment of the SCARNA2 promoter was monitored by semi-quantitative PCR amplification using primers amplifying the SCARNA2 promoter in the region -87 to +228 surrounding the Staf Binding Site. To confirm the specificity of binding, positive and negative controls were performed. PCR amplification of the tRNA^{Sec} promoter, which binds hStaf/ZNF143 as well, was used as the positive control. The negative control amplified a segment located several kbp upstream from the tRNA^{Sec} transcribed region, which does not harbor any hStaf/ZNF143 binding site. As a supplementary control, a similar experiment was performed without antibody. For each PCR reaction, we tested two dilutions of the DNA immunoprecipitated with anti-hStaf/ZNF143 (Figure 4, lanes 1,2) or without antibody (Figure 4, lanes 3,4). A serial dilution of the input material was also tested to demonstrate that the PCR was quantitative within a linear range of amplification (Figure 4, lanes 5-7). The SCARNA2 promoter was specifically immunoprecipitated by the hStaf/ZNF143 antibody (Figure 4A, compare lanes 1,2 and 3,4; the specific product is indicated by an arrow). As expected, the positive control with anti-hStaf/ZNF143 yielded a specific signal of intensity higher than in the no-antibody control (Figure 4B, compare lanes 1,2 and 3,4). In contrast, no specific PCR product could be observed with the PP1 primers amplifying SBS-lacking regions (Figure 4C, lanes 1-4). We conclude from these experiments that hStaf/ZNF143 is recruited to the SCARNA2 promoter in living cells.

Role of the hStaf/ZNF143 binding site in activation of the human SCARNA2 promoter

To further confirm the biological activity of the identified SBS, we generated a construct that contains the SBS GATATC mutant (substituting CTCCCA by GATATC at positions 1-6) fused to the luciferase reporter gene (SBSsub, Figure 5B). Expression of the reporter gene directed by the SCARNA2 promoter was monitored by the luciferase activity after transfection into COS-7 cells. Mutation of the SBS reduced the SCARNA2 promoter activity to about 30% of the wt level (Figure 5C, compare wt and SBSmut). These results provide strong evidence for the presence of a biologically active SBS in the SCARNA2 promoter.

Evaluation of the biological activities of other putative cis-elements in the human SCARNA2 promoter

To evaluate the role of the other three phylogenetically conserved elements, we introduced additional mutations in the construct harboring the SBS substitution. SBSsub was modified by introducing substitutions into the TATA, Y and X elements (Figure 5B). The SCARNA2 promoter activities were then assayed by quantification of the luciferase activity from COS-7 cell transfected extracts (Figure 5C). As shown in Figure 5C, the simultaneous mutations of the SBS and TATA element reduced the luciferase activity to about 20% of the wt control (compare wt, SBSsub and SBSsub-TATAsub). The reporter activity of the triple mutant Ysub-SBSsub-TATAsub was of the same order as that of the double SBSsub-TATAsub mutant, indicating that the Y element is not essential for SCARNA2 gene expression. In contrast, introduction of a fourth mutation by substitution in the X element led to basal levels of luciferase activity (compare wt, Ysub-SBSsub-TATAsub and Xsub-Ysub-SBSsub-TATAsub in Figure 5C).

Taken together, these results indicate that there are three positive regulatory elements in the human SCARNA2 promoter, the function of the X and SBS motifs being predominant for full promoter activity.

Transcription of the SCARNA2 gene continues for at least 112 nucleotides beyond the region encoding the mature scaRNA2

3'-end formation of the U1-U5 snRNAs, U3 and most likely U8 and U13 snoRNAs, requires a cis-acting sequence known as the 3'-box, located downstream of the 3'-end of the mature RNA. Transcription of the genes encoding these RNAs continues beyond the 3'-box in vivo. The 3' box is not an efficient transcription terminator but may rather act as a processing element functioning in the precursor RNA (13,10,14,16). Scrutiny of the sequence located at proximity of the 3'-end of the mature scaRNA2 did not show any motif with sequence similarity to the 3'-box of Pol II snRNA genes (typically GTTTN0-3AAARN2AGA) (34). However, a 15 bp motif located at positions +421/+435 immediately past the region encoding the 3'-part of the mature RNA is conserved in orthologous regions of mammalian SCARNA2 genes (3'-element in Figure 6A, B). To address the possibility that this motif can act as a termination signal, we deleted it in the miniSCARNA2 minigene to create miniSCARNA2 Δ 3' (Figure 6B). The miniSCARNA2 was obtained by a 152 bp deletion from positions +197 to +349 and was engineered in order to distinguish the product of the transfected minigene from the endogenous scaRNA2. The three constructs miniSCARNA2, miniSCARNA2 Δ 3' and empty vector were transfected separately into COS-7 cells in the presence of a 5S RNA maxigene as an internal control for transfection. After 48h, RNAs were extracted from transfected and untransfected cells and RNA levels were measured by RT-PCR. Four different primer pairs were used in the PCR step. The first one was devoted to amplification of the reverse transcription product of the endogenous U3 snoRNA. This assay acts as a control for RNA recovery (U3 snoRNA in Figure 6C). The second primer pair was used for PCR amplification of the 5S maxi reverse transcription product and gauged transfection efficiency (5S maxi in Figure 6C). The third primer pair (PIP2 in Figure 6B) produced a 214 bp PCR product corresponding to amplification of an internal region of the mature scaRNA2. Lastly, the fourth primer pair (PIP3 in Figure 6B) targeted a putative reverse transcription product of scaRNA2 containing at least 112 bp of 3'-end extension (position of primer P3). It generated PCR products of 372 bp and 357 bp with the miniSCARNA2 and miniSCARNA2 Δ 3' templates, respectively. We also tested serial dilutions of the reverse transcription products to ensure that the PCR was quantitative within a linear range of amplification. As expected, a

product was detected with P1P2 after transfection of miniSCARNA2 (Figure 6C, line P1P2, compare lanes 1-3 with 4-6). Surprisingly, a PCR product was also generated with P1P3, indicating that transcription continues for at least 112 nucleotides (position of primer P3) beyond the 3' end of the mature scaRNA2 (Figure 6C, line P1P3, lanes 1-3). The PCR products resulting from transfection of the miniSCARNA2 Δ 3' gene were similar to those obtained with miniSCARNA2 (Figure 6C, compare lanes 1-6 and 7-12). Interestingly, deletion of the 3'-element provoked a 2-fold drop in the yield of the PCR products (Figure 6C, compare lanes 1,2 and 7,8 for lines P1P2 and P1P3). This suggests that the 3'-element is involved in the transcription efficiency of the SCARNA2 gene and/or stability of the RNA transcript. The results also showed that the 3' conserved element does not act as a transcription termination signal and that transcription continues until at least 112 bp past the 3'-part of the mature scaRNA2.

DISCUSSION

In vertebrates, sequences encoding scaRNAs are generally located in the introns of their host genes. Intronic scaRNAs are produced after the splicing step by processing of the debranched intron (19). Here we demonstrated that the 161 bp region located upstream of the mature scaRNA2 bears sufficient information for transcription of the SCARNA2 gene. Furthermore, the present study established that transcription of the human SCARNA2 gene is governed by three cis-acting elements that are evolutionarily conserved, a TATA motif, a Staf Binding Site and a novel previously uncharacterized motif. The position of the TATA element suggests that it represents a bona fide TATA binding protein (TBP) motif, a key basal cis regulatory element shared by Pol II promoters of protein coding genes and a few RNA polymerase III promoters. The CTTAAAAG sequence of the TATA element diverges in the 5'-part from the TATAAAAG consensus (35). However, structural studies of TBP-DNA complexes concluded that such a nucleotide sequence variability does not impair TBP recognition, the protein being able to accommodate C at position 1 and T at position 2 of the TATA element (36). Indeed, we showed that nucleotide substitutions in the TATA element had only a minor impact (about 10% of the wt level) on transcription efficiency of the SCARNA2 gene. It is

likely that the function of the TATA element is to ensure correct initiation of transcription at the nucleotide corresponding to the 5'-end of the mature scaRNA2, as in other Pol II transcribed genes (37). In the SCARNA2 promoter, substitution in the TATA sequence most likely led to multiple transcription start sites as observed with protein coding genes lacking TATA elements (37).

The second motif that we characterized, called X in this study, does not correspond to any cis-acting motif previously identified. It is only conserved in mammalian SCARNA2 promoters but is surprisingly absent in *Xenopus*, pufferfish and zebrafish SCARNA2 genes. This suggests that the X element acts in a species specific manner.

The third identified cis-acting motif is specifically recognized by hStaf/ZNF143 which was found associated to the SCARNA2 promoter in vivo. Although the *Xenopus* Staf was originally identified as the transcription activator of the tRNA^{Sec} gene, it also controls expression of snRNA and snRNA-type genes transcribed by RNA polymerases II and III (23,24,26). Furthermore, hStaf/ZNF143 has been reported so far to regulate transcription of nine protein coding gene (29,32,33). Lastly and more surprisingly, a genome-wide analysis led us to recently identify 1175 hStaf/ZNF143 binding sites distributed in 938 mammalian protein gene promoters (29). This strongly suggests that hStaf/ZNF143 is involved in the transcriptional regulation of a very large number of protein coding genes.

The present work extends the role of hStaf/ZNF143 to the transcription of a scaRNA gene which harbors a promoter organization different from that of snRNA and snRNA-type genes. Interestingly, the SCA2RNA promoter does not contain the PSE motif which is a key element in the transcription of snRNA genes (10). The SCARNA2 gene, as is the case for the telomerase RNA gene (TERC) and most likely SCARNA17, carries promoter elements that look like those of protein coding genes. However, despite this resemblance, the organization of the elements is unique within the promoters of the SCARNA2, SCARNA17 and TERC genes (38,39). An important question to ask is why the scaRNA2 and snRNA genes are transcribed by different promoters. snRNA gene promoters are known to be among the most active ones, producing about $2-3 \cdot 10^6$ transcripts per cell per generation (40). In contrast, SCARNA2 may not require a promoter with such a high activity and can thus function with a promoter characteristic of protein coding genes. We therefore suggest that the relative difference in the activity levels of the promoters is related to the much higher amounts of spliceosomal snRNAs,

compared to scaRNAs, that are demanded to accomplish mRNA splicing. Another possibility, not mutually exclusive with the former, is the following. The vast majority of snRNA genes are ubiquitously expressed while, in contrast, most of the mRNA coding gene expression is temporally and spatially restricted. Thus, the presence of an mRNA-type promoter in the scaRNA2 gene can offer the possibility to also regulate its expression.

Finally, we showed that transcription of the SCARNA2 gene continues past the 3'-end of the mature RNA. This is reminiscent of the transcriptional status of snRNA genes where the snRNA precursors are extended for at least 250 nucleotides beyond the 3'-end of the mature RNA (13). Stable pre-snRNA intermediates bearing a short 3'-end extension are generated by a transcription-linked cleavage event that requires the 3'-box (14-16). However, the 3'-conserved element located immediately after the region encoding the 3'-part of the mature scaRNA2 is not structurally related to the 3'box of snRNA genes. Our work showed that this motif is not an efficient transcription terminator but its involvement as a 3'-end processing motif cannot be excluded. This will be addressed in future investigations.

ACKNOWLEDGMENTS

We thank L. Wurth and C. Allmang for U3 snoRNA oligonucleotides and A. Schweigert for excellent technical assistance.

REFERENCES

1. Limbach, P.A., Crain, P.F. and McCloskey, J.A. (1994) Summary: the modified nucleosides of RNA. *Nucleic Acids Res*, **22**, 2183-2196.
2. Maden, B.E. and Hughes, J.M. (1997) Eukaryotic ribosomal RNA: the recent excitement in the nucleotide modification problem. *Chromosoma*, **105**, 391-400.
3. Smith, C.M. and Steitz, J.A. (1997) Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell*, **89**, 669-672.
4. Reichow, S.L., Hamma, T., Ferre-D'Amare, A.R. and Varani, G. (2007) The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res*, **35**, 1452-1464.
5. Kiss-Laszlo, Z., Henry, Y., Bachellerie, J.P., Caizergues-Ferrer, M. and Kiss, T. (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*, **85**, 1077-1088.
6. Tycowski, K.T., Smith, C.M., Shu, M.D. and Steitz, J.A. (1996) A small nucleolar RNA requirement for site-specific ribose methylation of rRNA in *Xenopus*. *Proc Natl Acad Sci U S A*, **93**, 14480-14485.
7. Ganot, P., Bortolin, M.L. and Kiss, T. (1997) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell*, **89**, 799-809.
8. Ni, J., Tien, A.L. and Fournier, M.J. (1997) Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell*, **89**, 565-573.
9. Meier, U.T. (2005) The many facets of H/ACA ribonucleoproteins. *Chromosoma*, **114**, 1-14.
10. Hernandez, N. (2001) Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J Biol Chem*, **276**, 26733-26736.
11. Schramm, L. and Hernandez, N. (2002) Recruitment of RNA polymerase III to its target promoters. *Genes Dev*, **16**, 2593-2620.
12. Kiss, T. (2004) Biogenesis of small nuclear RNPs. *J Cell Sci*, **117**, 5949-5951.
13. Cuello, P., Boyd, D.C., Dye, M.J., Proudfoot, N.J. and Murphy, S. (1999) Transcription of the human U2 snRNA genes continues beyond the 3' box in vivo. *Embo J*, **18**, 2867-2877.
14. Medlin, J.E., Uguen, P., Taylor, A., Bentley, D.L. and Murphy, S. (2003) The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *Embo J*, **22**, 925-934.
15. Uguen, P. and Murphy, S. (2003) The 3' ends of human pre-snRNAs are produced by RNA polymerase II CTD-dependent RNA processing. *Embo J*, **22**, 4544-4554.
16. Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N. and Shiekhattar, R. (2005) Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell*, **123**, 265-276.
17. Darzacq, X., Jady, B.E., Verheggen, C., Kiss, A.M., Bertrand, E. and Kiss, T. (2002) Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *Embo J*, **21**, 2746-2756.
18. Kiss, A.M., Jady, B.E., Darzacq, X., Verheggen, C., Bertrand, E. and Kiss, T. (2002) A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. *Nucleic Acids Res*, **30**, 4643-4649.
19. Kiss, A.M., Jady, B.E., Bertrand, E. and Kiss, T. (2004) Human box H/ACA pseudouridylation guide RNA machinery. *Mol Cell Biol*, **24**, 5797-5807.
20. Richard, P., Darzacq, X., Bertrand, E., Jady, B.E., Verheggen, C. and Kiss, T. (2003) A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *Embo J*, **22**, 4283-4293.
21. Filipowicz, W. and Pogacic, V. (2002) Biogenesis of small nucleolar ribonucleoproteins. *Curr Opin Cell Biol*, **14**, 319-327.
22. Ach, R.A. and Weiner, A.M. (1991) Cooperation between CCAAT and octamer motifs in the distal sequence element of the rat U3 small nucleolar RNA promoter. *Nucleic Acids Res*, **19**, 4209-4218.

23. Schuster, C., Myslinski, E., Krol, A. and Carbon, P. (1995) Staf, a novel zinc finger protein that activates the RNA polymerase III promoter of the selenocysteine tRNA gene. *Embo J*, **14**, 3777-3787.
24. Schaub, M., Myslinski, E., Schuster, C., Krol, A. and Carbon, P. (1997) Staf, a promiscuous activator for enhanced transcription by RNA polymerases II and III. *Embo J*, **16**, 173-181.
25. Tycowski, K.T., Aab, A. and Steitz, J.A. (2004) Guide RNAs with 5' caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoa. *Curr Biol*, **14**, 1985-1995.
26. Myslinski, E., Krol, A. and Carbon, P. (1998) ZNF76 and ZNF143 are two human homologs of the transcriptional activator Staf. *J Biol Chem*, **273**, 21998-22006.
27. Myslinski, E., Krol, A. and Carbon, P. (1992) Optimal tRNA((Ser)Sec) gene activity requires an upstream SPH motif. *Nucleic Acids Res*, **20**, 203-209.
28. Weinmann, A.S. and Farnham, P.J. (2002) Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods*, **26**, 37-47.
29. Myslinski, E., Gerard, M.A., Krol, A. and Carbon, P. (2006) A genome scale location analysis of human Staf/ZNF143-binding sites suggests a widespread role for human Staf/ZNF143 in mammalian promoters. *J Biol Chem*, **281**, 39953-39962.
30. Quandt, K., Frech, K., Karas, H., Wingender, E. and Werner, T. (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res*, **23**, 4878-4884.
31. Schaub, M., Krol, A. and Carbon, P. (1999) Flexible zinc finger requirement for binding of the transcriptional activator staf to U6 small nuclear RNA and tRNA(Sec) promoters. *J Biol Chem*, **274**, 24241-24249.
32. Gerard, M.A., Krol, A. and Carbon, P. (2007) Transcription factor hStaf/ZNF143 is required for expression of the human TFAM gene. *Gene*, **401**, 145-153.
33. Myslinski, E., Gerard, M.A., Krol, A. and Carbon, P. (2007) Transcription of the human cell cycle regulated BUB1B gene requires hStaf/ZNF143. *Nucleic Acids Res*.
34. Weiner, A.M. (2005) E Pluribus Unum: 3' end formation of polyadenylated mRNAs, histone mRNAs, and U snRNAs. *Mol Cell*, **20**, 168-170.
35. Smale, S.T. and Kadonaga, J.T. (2003) The RNA polymerase II core promoter. *Annu Rev Biochem*, **72**, 449-479.
36. Patikoglou, G.A., Kim, J.L., Sun, L., Yang, S.H., Kodadek, T. and Burley, S.K. (1999) TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev*, **13**, 3217-3230.
37. Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y. and Hume, D.A. (2007) Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet*, **8**, 424-436.
38. Zhao, J.Q., Hoare, S.F., McFarlane, R., Muir, S., Parkinson, E.K., Black, D.M. and Keith, W.N. (1998) Cloning and characterization of human and mouse telomerase RNA gene promoter sequences. *Oncogene*, **16**, 1345-1350.
39. Zhao, J., Bilsland, A., Hoare, S.F. and Keith, W.N. (2003) Involvement of NF-Y and Sp1 binding sequences in basal transcription of the human telomerase RNA gene. *FEBS Lett*, **536**, 111-119.
40. Reddy, R., Henning, D. and Busch, H. (1985) Primary and secondary structure of U8 small nuclear RNA. *J Biol Chem*, **260**, 10930-10935.

FIGURES

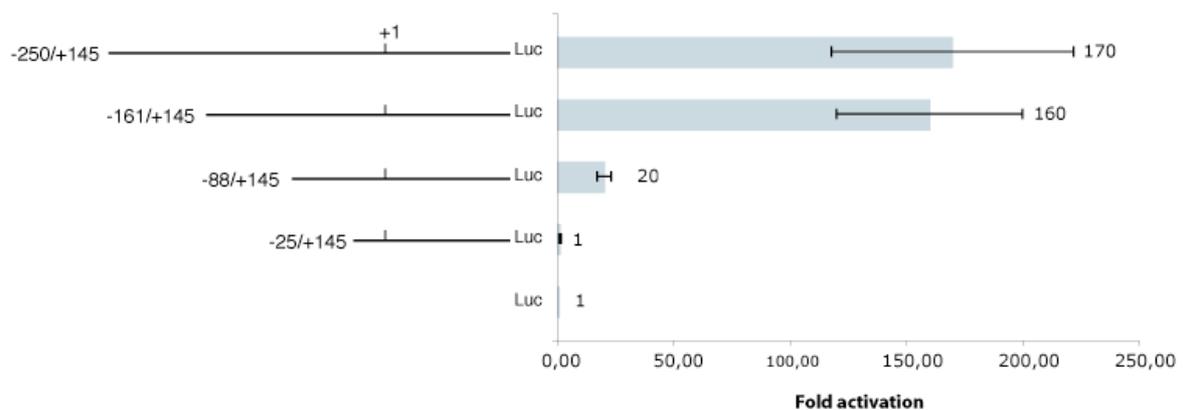


Figure 1. Detection of promoter activity in the 5' flanking sequence of the human SCARNA2 gene. Schematic representation of the 5'-deleted SCARNA2 constructs fused to a luciferase (Luc) reporter and their activities in COS-7 cells. For each construct, the length of the sequence upstream and downstream of the +1 is shown to the left and the luciferase activity to the right. Cells were transiently transfected with the constructs and assayed for luciferase activity. The firefly luciferase activity was normalized to that of the renilla luciferase provided by co-transfection of the control vector. Data for each construct was plotted as fold-activation which represents the ratio between the normalized luciferase activity of each reporter construct to that of the empty luciferase vector (pFLASH1). Data are presented as the mean \pm S.D. of three separate experiments.

A

```

-250
h -----CCTGGGCGACACAGCAAGAACGGGTCTCCAAAAAAAA
c -----CTGGGCGACACGGCGAGAACGGGTCTCCAAAAAAAA
d ---GTGTGTTGTGATGAATAAATAACTACAATCTTTAAAGAGAGGGGGGAAAAAAAAAGACA
m CCGGAGGCTGAAGTGGGAAGGATCTTTGAGTGGGATACCAGGTTGGATAACTTAGACTTTC
r ---GAGCCTGATGCAG-AAGGATCTTGAGTCGGTTACCAGGTTGGGCAACTTAGAATTTC
                                     **  **      *  *

-212
h AA-ACCA-----CGATTCTGTAAGGTGGAAATT-A---GTACTGCCTTTTCTGGGACCT
c AACACCA-----CGATTCTGTAAGGTGGAAATT-A---GTACCGCTTTTCTGGGACCT
d AGCACGATCTTGCGGTCTAGTCTTGCCGGAGTT-GTTTGTCCAAACCTTCCCAGGGACCC
m AA-ACAA-----AATGCAACAAAACCTTGA-----TTTATGGAAATCAATCCTGTGTCCC
r AA-ACAA-----AACACAACAAAACCTAAGTCCATCTGTGGAAATCAAGCTT-TCTGTC
  *  ** *                               *                               *

-162
                                     X
h ATGTTCCCC-TCCGAGCCACGCGAGGGCAGCAACGCGCGCACTCTGAGGAACACCCCTT
c ATGTTCCCC-TCCGAGCCACGCGAGGGCAGCAACGCGCGCACTCTGAGGAACACCTCGTT
d GGGTTCCCCCTCCGAGCCACGCGAGGGCAGCAGAGCGCAC---CGAGGGCGCACCCCG--
m CCGGGACCT--CCGAGCCACTAGGTGGCAGAAGAGAATGC---TTTGATAAACAAGGTA
r CTGGGACCT--CCCAGCCACAGGGTGGCAGAAGAGAACAC---TCTGGCTAAACAAGGTG
  *  **  ** ***** * ***** * * * * * * * * * * * * * * *

-103
                                     Y                               SBS
h CCTCAGGACTCGTCCGCCTGTGCTCGGTGGTTTCTCCCAGAATCCTCCAAGGCGCGGGAG
c CCTCAGGACTCGTCCGCCTGTGCTCGGTGGTTTCTCCCAGAATCCTCCAAGGCGCGGGAG
d -CCGAGGGCGT-----TGCCTCAGTGGCTCCTCCCAGAATCCTCCCGG-----
m TGCGGAAGCCG---CCTGATCGCTCTGTGGTTCTCCCAGAATCCTTTGAGGCA-----
r TACGGAAACCG---CCTGACTACTCTGTGGTTCTCCCAGAATCCTTTGAGACA-----
  *                               *** * * * * * * * * * * * * * * *

-43
          TATA element                               +1
h GGGGCGGAGAGCTTAAAAAGTCGCGGGGGGAGCGGACTTCCGCTGTTTATAGGGAGGGAGAG
c GGGGCGGAGGGCTTAAAAAGCCGCGGGGGGAGCGGACTTCCGCTGTTTATAGGGAGGGTGA
d ---GCGGAGGCCTTTAAAAA---GGGGGTGCCGACTTCCGCCCCTCCCGGAGGGAGAG
m -AGACCGAGTGTTTAAAAAG-----AGGGTGCCCCAGCCACTTCTCGCGAGTGGGAGTG
r -AGACCGAGTGTTTAAAAAG-----AGGGTGCAACTGCCCTTTCCTCGCGAGTGGGAGCG
  *  ***  ** ***                               *** * * * * * * * * * *

```

B

SBS	-70	CTCCCAGAATCCTCCAAG	-53
SBS cons.1		TACCCATAATGCATYGCG	
SBS cons.2		YWCCCRNMATSCMYRCR	
		***** * * * *	

Figure 2. Comparison of the promoter sequences of the human (h), chimpanzee (c), dog (d), mouse (m) and rat (r) SCARNA2 genes. (A) Sequences covering the -250/+17 human region and its orthologs were aligned with ClustalW. Asterisks represent identical nucleotides. The numbering is relative to the first nucleotide of the mature transcript (+1). The putative transcription factor binding sites, cis elements and transcription start site are highlighted in gray. (B) Alignment of the SBS sequence of the SCARNA2 promoter with the consensus SBS determined by binding site selection at highly (SBS cons.1) (31) or moderately stringent selection conditions (SBS cons.2) (24). Y, W, R, N, M and S stand for T/C, A/T, A/G, any nucleotide, A/C, and G/C, respectively. Identical nucleotides are indicated by asterisks.

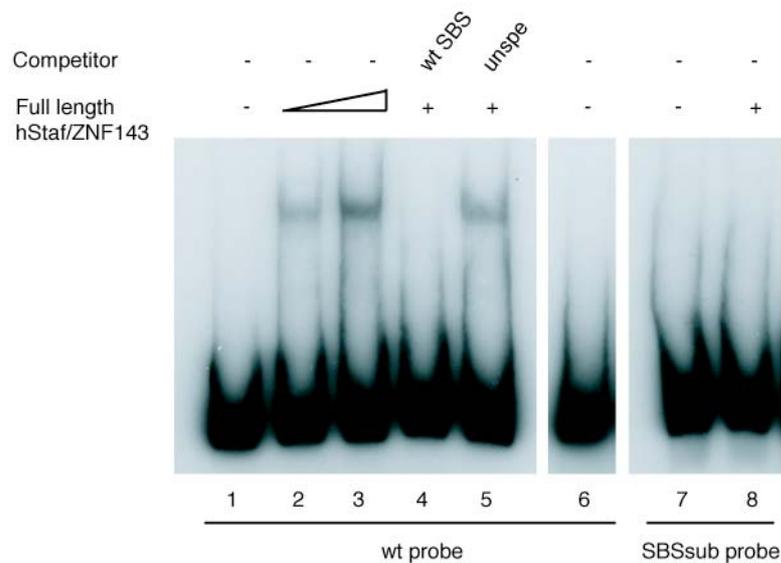


Figure 3. Identification of an hStaf/ZNF143 transcription factor binding site in the human SCARNA2 promoter. Gel retardation assays with ^{32}P -labeled promoter fragments encompassing positions -88/+145 of the SCARNA2 promoter and containing the wild-type SBS sequence (wt probe, lanes 1-6), mutant SBS (SBSsub probe, lanes 7,8). In the SBSsub probe, the -70 CTCCCA-65 sequence was replaced by GATATC. Probes were incubated in the absence (lanes 1 and 7) or presence of hStaf/ZNF143 (2.5 μl of programmed reticulocyte lysate in lanes 2, 4, 5 and 8; 5 μl in lane 3). Reactions in lanes 4 and 5 were performed in the presence of a 1000-fold molar excess of unlabeled specific competitor (wt SBS) and unspecific competitor (unspe). Reaction in lane 6 was performed in the presence of 5 μl of unprogrammed reticulocyte lysate. Lanes 1-6 and 7-8 are from separate experiments.

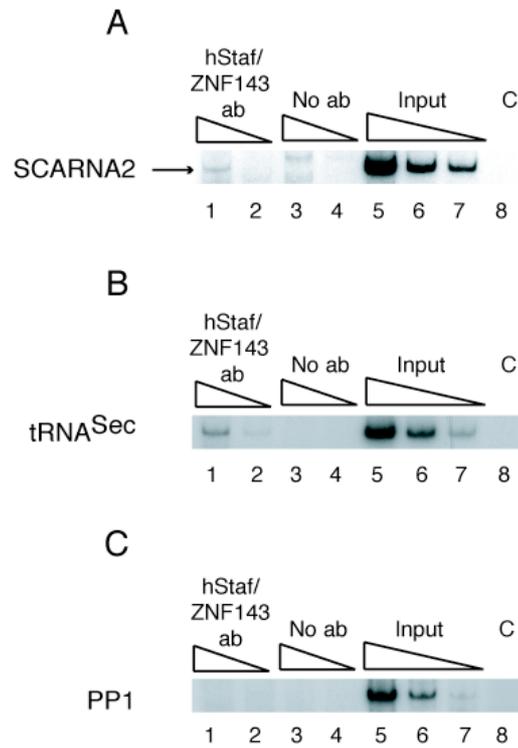


Figure 4. Immunoprecipitation of the SCARNA2 promoter from HeLa cells. Formaldehyde cross-linked chromatin was prepared from HeLa cells and immunoprecipitated with hStaf/ZNF143 antibody. DNA fragments recovered from input chromatin, or ChIP assays with anti-hStaf/ZNF143 or control experiment without antibody, were analyzed by semi-quantitative PCR with specific primer pairs in the presence of (α -³²P) dCTP. Lanes 1,2 and 3,4: serial dilutions of DNA immunoprecipitated with anti-hStaf/ZNF143 or no antibody, respectively. Lanes 5-7: serial dilutions of input material to demonstrate that the assays were within the linear range of PCR amplification. Lane 8 (C, PCR control): PCR lacking DNA template. (A) PCR was performed with specific primer pairs for the human SCARNA2 promoter or (B) for the tRNA^{Sec} promoter acting as the positive control. (C) Negative control with the PP1 primer pair targeting a region located upstream of the tRNA^{Sec} gene and lacking SBS.

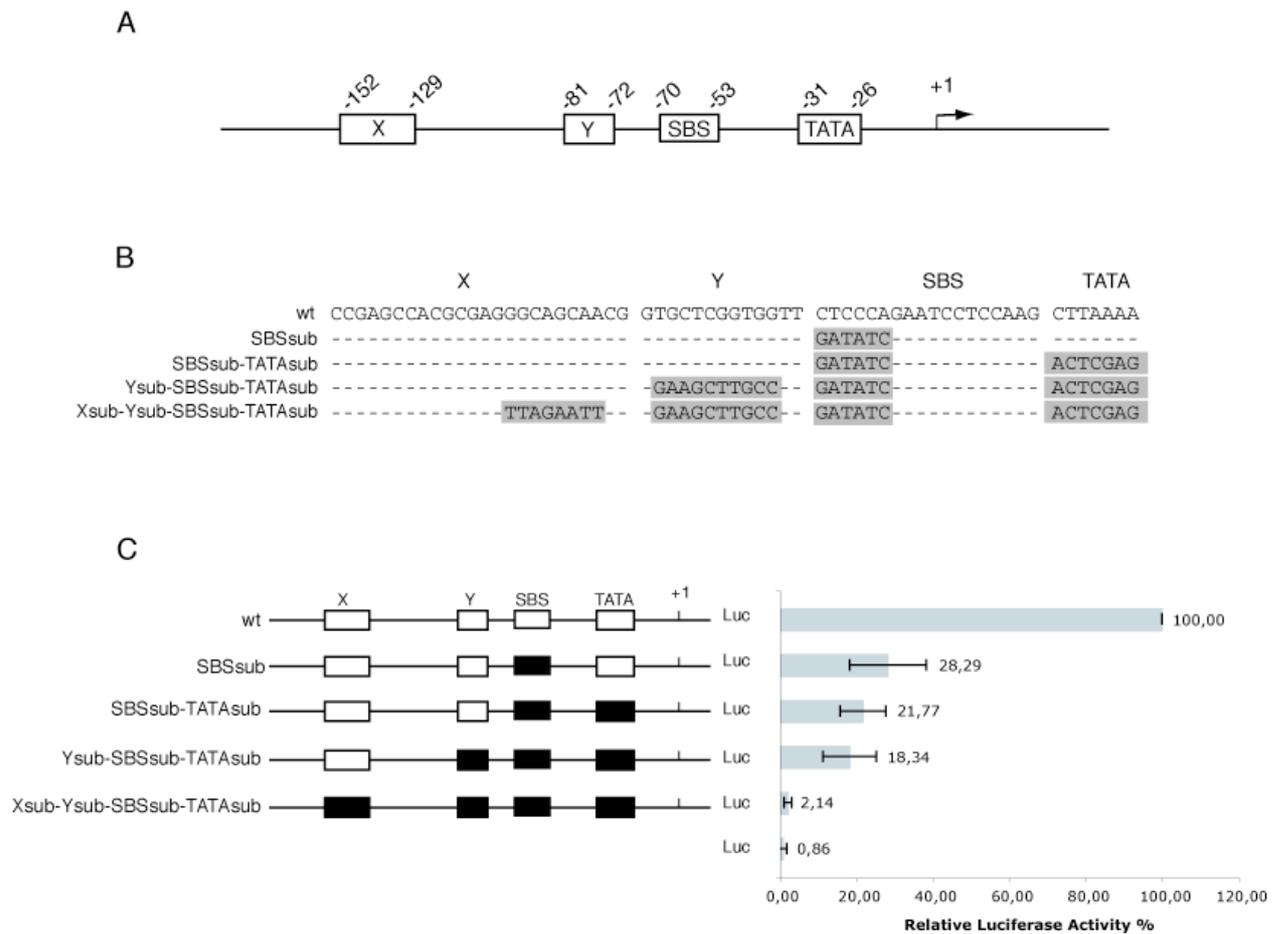


Figure 5. Functional analysis of the elements in the human SCARNA2 promoter. (A) Schematic representation of the SCARNA2 promoter. The four putative promoter elements that are cross-species conserved are boxed. (B) Sequences of the wild-type and mutant X, Y, SBS, and TATA elements of SCARNA2 promoter. Mutations are highlighted in gray. (C) Contribution of the X, Y, SBS, and TATA elements to SCARNA2 promoter activity. Left panel: schematic representation of the SCARNA2 promoter-luciferase (Luc) constructs substituted in the X, Y, SBS, and TATA elements. Open and solid boxes represent wild-type and mutant elements, respectively. Right panel: COS-7 cells were transfected with the reporter constructs and the luciferase activity was normalized to that of the renilla luciferase obtained from the co-transfected control vector. The normalized luciferase activity of the wild-type construct -250/+145 was arbitrarily set as 100%. Data are presented as the mean \pm S.D. of three separate experiments.

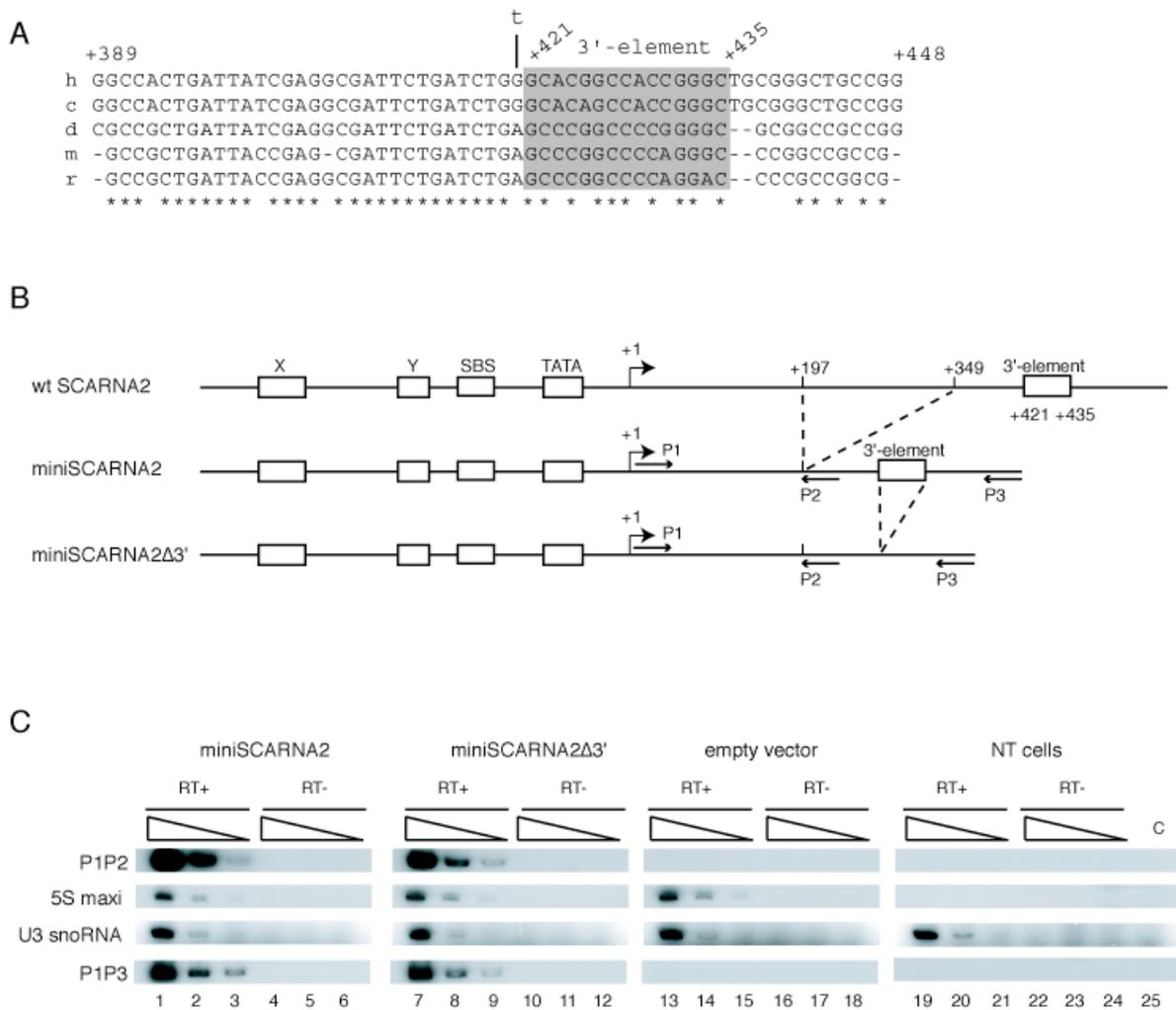


Figure 6. Transcription termination of the SCARNA2 gene occurs at least 112 nucleotides downstream of the region encoding the 3' end of the mature scaRNA2. (A) Alignment of about 30 bp upstream and downstream of the region encoding the 3'-end of the scaRNA2 (t). Asterisks represent identical nucleotides. The putative 3'-element is highlighted in gray. (B) Schematic representation of the structures of the wt SCARNA2, miniSCARNA2 and miniSCARNA2Δ3' genes used in the assays. The different elements are boxed and the location of P1 to P3 primers used in PCR is indicated. The numbering is relative to the transcription start site (+1). (C) COS-7 cells were either not transfected (NT) or transfected with the constructs indicated above the lanes in the presence of the 5S RNA maxigene. After RNA extraction, the scaRNA2, 5S maxi and U3 snoRNA contents were analyzed by RT-PCR. RT+ (lanes 1-3, 7-9, 13-15 and 19-21) and RT- (lanes 4-6, 10-12, 16-18 and 22-24) assays were performed in the presence and absence of reverse transcriptase, respectively. Lanes 1-3, 4-6, 7-9, 10-12, 13-15, 16-18, 19-21 and 22-24: PCR assays performed as indicated with decreasing amounts of DNA templates. 5S maxi and U3 snoRNA indicate PCR performed with PCR primer pairs for amplifying the reverse transcription products of the 5S maxi RNA and U3 snoRNA, respectively. P1P2 and P1P3: PCR primer pairs for amplification of the reverse transcription products of scaRNA2. Lane 25 (C, PCR control): PCR lacking DNA template.

3. Discussion

Chez les vertébrés, les gènes codant les ARNsca sont généralement localisés dans les introns de gènes d'ARNm. Après épissage du pré-ARNm, les ARNsca sont libérés par l'étape de débranchement de l'intron et sont maturés par des exonucléases (Bachellerie et al., 2002). Dans notre étude, nous avons montré que la transcription du gène SCARNA2 est assurée par une région minimale de 161 pb située en amont de ce gène. Ceci révèle pour la première fois qu'un ARNsca peut être produit à partir d'une unité transcriptionnelle indépendante. Nous avons également démontré que la transcription du gène SCARNA2 est dépendante de trois éléments conservés chez l'homme, le chimpanzé, le chien, la souris et le rat : un élément TATA, un site de liaison de Staf et un nouveau motif, nommé X, ne correspondant à aucun site connu de liaison de facteurs de transcription. Les éléments TATA et SBS sont également présents dans les promoteurs orthologues chez le xénope et le poisson zèbre et le fugu.

La position de l'élément TATA par rapport au TSS suggère que cet élément correspond à une boîte TATA, élément promoteur présent dans de nombreuses régions promotrices de gènes de protéines transcrits par la Pol II et de gènes d'ARNsn et de type ARNsca transcrits par la Pol III. La séquence CTTAAAAG de cet élément identifié dans la région promotrice du gène SCARNA2 est divergente en 5' par rapport à la séquence consensus de la boîte TATA (TATAWAAR) (Smale et Kadonaga, 2003). Cependant, plusieurs études ont montré qu'une grande variété de séquences riches en nucléotides A et T pouvait fonctionner *in vivo* et que la reconnaissance et la fixation de TBP sur la boîte TATA autorisaient un C et T respectivement en position 1 et 2 de cet élément (Singer et al., 1990; Patikoglou et al., 1999). La substitution de cet élément au sein de la région promotrice du gène SCARNA2 entraîne une légère diminution de l'efficacité de transcription. Il est probable que cet élément TATA permette l'initiation précise de la transcription au nucléotide correspondant à l'extrémité 5' de l'ARNsca2 mature comme cela a été observé pour les gènes de protéines transcrits par la Pol II (Sandelin et al., 2007). Il a été observé que l'initiation de la transcription se réalise au niveau de sites multiples lorsque les promoteurs de gènes de protéines sont dépourvus de boîte TATA (Sandelin et al., 2007). Ainsi, il se peut qu'une population d'ARNsca2 hétérogène au niveau de l'extrémité 5' soit synthétisée suite à la substitution de l'élément TATA réalisée dans le promoteur du gène SCARNA2.

Un motif de 24 pb, nommé X, associé à aucun facteur de transcription connu, a également été observé en comparant les régions promotrices du gène SCARNA2 orthologues chez l'homme, le chimpanzé, le chien, la souris et le rat. Ce motif X n'est pas présent dans le promoteur du gène SCARNA2 chez le xénope, le poisson zèbre et le fugu. Ceci suggère que l'utilisation de ce motif soit limitée aux mammifères.

Le troisième motif important dans la transcription du gène SCARNA2 est un site de liaison de hStaf/ZNF143. Les précédentes études ont montré que hStaf/ZNF143 est essentiel à la transcription des gènes d'ARNsn et de type ARNsn par la Pol II et la Pol III (Schuster et al., 1995; Schaub et al., 1997). Récemment, il a été mis en évidence que hStaf/ZNF143 est également impliqué dans la transcription de gènes de protéines chez les mammifères (Kubota et al., 2000; Mach et al., 2002; Saur et al., 2002; Barski et al., 2004; Di Leva et al., 2004; Grossman et al., 2004; Ishiguchi et al., 2004; Myslinski et al., 2006; Gerard et al., 2007b; Myslinski et al., 2007). L'étude du promoteur du gène SCARNA2 permet d'élargir encore le spectre de gènes cibles de ce facteur.

Le promoteur du gène SCARNA2 présente une architecture différente de celle des promoteurs des gènes d'ARNsn et de type ARNsn. Une des principales différences est l'absence de l'élément promoteur PSE caractéristique des promoteurs des gènes d'ARNsn et de type ARNsn (Hernandez, 2001). Les promoteurs des gènes d'ARNsn sont connus pour être des promoteurs forts produisant environ deux à trois millions de molécules d'ARN par cellule et par génération. Dans le cas des ARNsn U, cette quantité est nécessaire pour assurer l'épissage des nombreux pré-ARNm dans la cellule. Le gène du SCARNA2 ne nécessite pas un promoteur aussi fort. Ainsi, le fait que les gènes d'ARNsn et d'ARNsca soient sous le contrôle de promoteurs différents peut être dû aux besoins plus importants de la cellule en ARNsn par rapport aux ARNsca.

Pour finir, nous avons montré que la transcription du gène SCARNA2 se poursuit au moins 112 nts après la position correspondant à l'extrémité 3' de l'ARNsca2 mature. Ce phénomène a déjà été observé dans le cas des gènes des ARNsn pour lesquels les précurseurs des ARNsn sont étendus d'au moins 250 nts après l'extrémité 3' du transcrit mature (Cuello et al., 1999). Ces précurseurs sont une première fois maturés pour donner naissance à un précurseur intermédiaire présentant une courte extension à l'extrémité 3'. Ce clivage est couplé à la transcription et est dépendant de la présence d'une boîte 3' localisée juste en aval de la région

spécifiant l'extrémité 3' de l'ARNsn mature (Medlin et al., 2003; Uguen et Murphy, 2003; Baillat et al., 2005). L'élément conservé situé juste après la position spécifiant l'extrémité 3' de l'ARNsca2 mature ne correspond pas à la boîte 3' associée aux gènes d'ARNsn. Nos travaux ont montré que cet élément n'intervient pas dans la terminaison de la transcription mais son implication dans la maturation de l'ARNsca2 ne peut pas être exclue.

CONCLUSION GÉNÉRALE
ET PERSPECTIVES

Conclusion générale et perspectives

Les précédents travaux du laboratoire ainsi que ceux effectués au cours de ma thèse montrent que le facteur de transcription hStaf/ZNF143 régule l'activité de gènes impliqués dans des voies cellulaires diverses via des promoteurs présentant une architecture variée (Figure 27). En effet, Staf a été identifié au laboratoire comme l'activateur de la transcription du gène de l'ARNt^{Sec} chez le xénope (Schuster et al., 1995). Il est également essentiel à la transcription des gènes d'ARNnc transcrits par la Pol II comme les gènes des ARNsn U1-U5 ainsi qu'à la transcription d'autres gènes d'ARNnc transcrits par la Pol III comme les gènes de l'ARNsn U6 et de l'ARN de la MRP ou du 7SK (Figure 27) (Schuster et al., 1995; Schaub et al., 1997). Contrairement à certains gènes d'ARNsno et d'ARNsca localisés dans des introns de gènes codant des protéines, les gènes des ARNsn U3, U8 et U13 sont contrôlés par un promoteur reconnu par la Pol II et semblable dans son organisation à ceux des gènes des ARNsn U1-U5 (Tyc et Steitz, 1989; Ach et Weiner, 1991). Lors de mes travaux de thèse, nous avons mis en évidence que le gène SCARNA2, codant l'ARNsca2 prédit pour modifier l'ARNsn U2 au sein des corps de Cajal, est également transcrit à partir d'une unité transcriptionnelle indépendante (Gerard et al., 2007 soumis pour publication). Nous avons montré, par des expériences de transfections transitoires, que la région promotrice minimale, d'une taille de 161 pb, renferme trois éléments régulateurs fonctionnels responsables de la transcription du gène SCARNA2. Parmi ces trois éléments, nous avons identifié un site SBS dont l'intégrité est essentielle à l'activité de cette région promotrice. Ainsi, le gène SCARNA2 est un nouveau gène cible pour le facteur hStaf/ZNF143, l'orthologue humain de Staf (Figure 27).

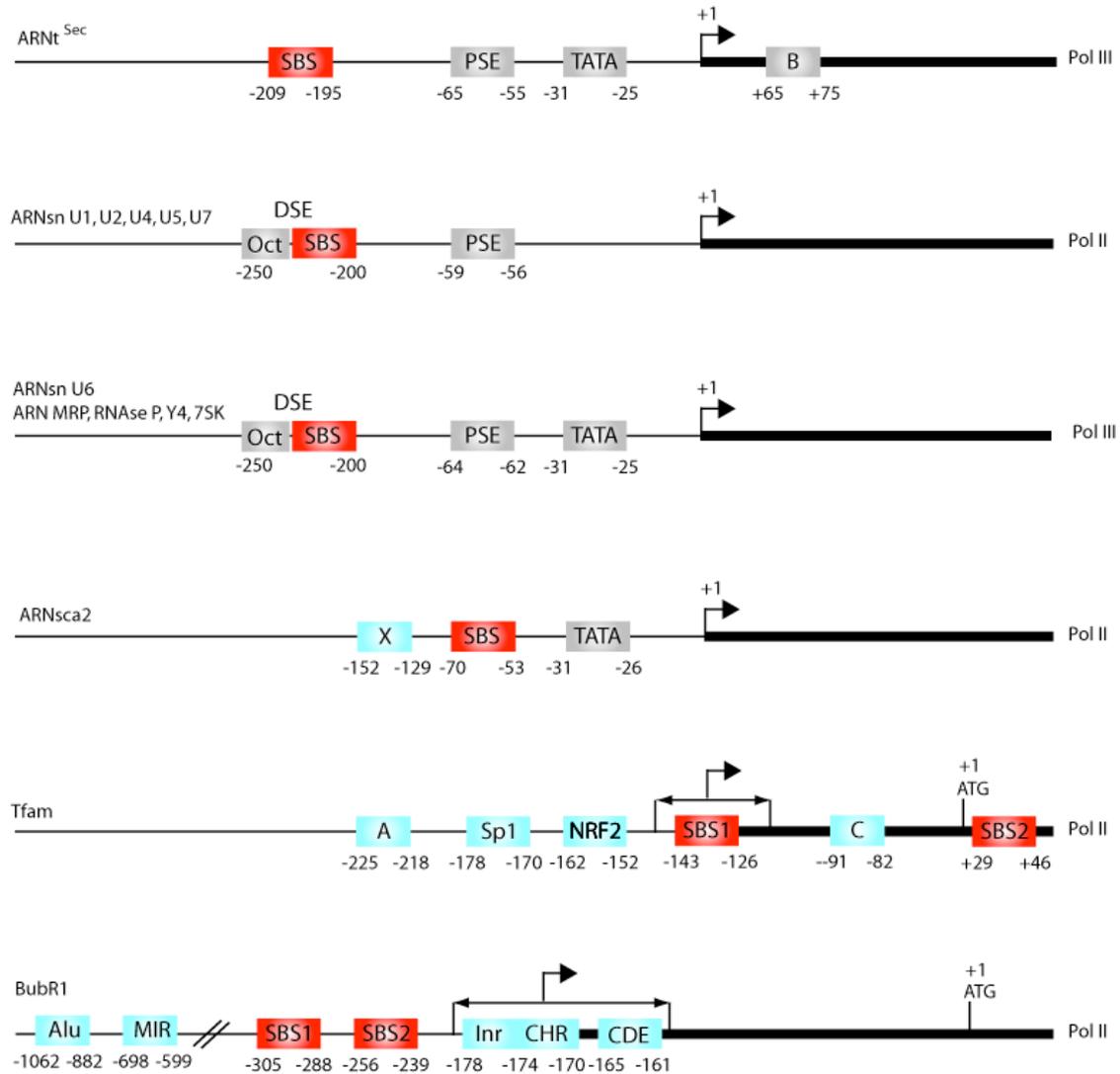


Figure 27 : Représentation schématique des promoteurs des gènes régulés par hStaf/ZNF143 chez les vertébrés caractérisés au laboratoire (Schuster et al., 1995; Schaub et al., 1997; Gerard et al., 2007b; Gerard et al., 2007a soumis pour publication; Myslinski et al., 2007).

Les sites de liaison de Staf sont représentés en rouge tandis que les autres éléments communs aux gènes spécifiant les ARNs et les protéines indiqués sont représentés en gris. Les éléments spécifiques à ces gènes sont colorés en bleu. Les ARN polymérase responsables de la transcription des différents gènes sont indiquées pour chaque promoteur. La région transcrite est représentée par un trait épais. La flèche indique l'unique TSS dans le cas des gènes d'ARN et la double flèche indique la région contenant les différents TSS identifiés pour les gènes codant les protéines Tfam et BubR1. La numérotation utilisée dans le cas des gènes codant les protéines Tfam et BubR1 est établie à partir du codon d'initiation de la traduction. Les positions indiquées pour le gène de l'ARNt^{Sec} sont celles du gène de *Xenopus laevis* et celles indiquées pour les gènes de l'ARNsca2 et des protéines Tfam et BubR1 sont spécifiques de l'homme.

Il a également été observé au laboratoire que Staf était capable d'activer la transcription d'un gène rapporteur d'ARNm sous le contrôle d'un promoteur chimère composé de plusieurs sites SBS clonés en amont du promoteur de base de la thymidine kinase (Schuster et al., 1995). De récentes études, réalisées par d'autres équipes, ont montré que hStaf/ZNF143 pouvait réguler la transcription de sept gènes de protéines (Kubota et al., 2000; Mach et al., 2002; Saur et al., 2002; Barski et al., 2004; Di Leva et al., 2004; Grossman et al., 2004; Ishiguchi et al., 2004). Les travaux effectués au cours de ma thèse ont permis de caractériser finement le promoteur de deux nouveaux gènes de protéines cibles du facteur de transcription hStaf/ZNF143 (Gerard et al., 2007b; Myslinski et al., 2007). Il s'agit du gène TFAM codant le facteur de transcription mitochondrial Tfam et du gène BUB1B codant la protéine BubR1 constituant majeur du complexe de contrôle mitotique (Figure 27). L'implication du facteur hStaf/ZNF143 dans la transcription de ces deux gènes a été analysée par des expériences de retard sur gel et d'immunoprécipitation de chromatine ainsi que par des études fonctionnelles en transfectant différentes constructions renfermant la région promotrice sauvage ou mutée au niveau des sites SBS identifiés. Il existe également d'autres éléments régulateurs conservés dans ces deux régions promotrices. Ces éléments sont de nature variable selon le promoteur étudié. On peut supposer qu'il existe des facteurs de transcription particuliers impliqués dans la régulation de la transcription de gènes cibles de hStaf/ZNF143. L'analyse d'un plus grand nombre de promoteurs de gènes de protéines pourrait mettre en évidence un réseau de facteurs de transcription agissant en collaboration avec hStaf/ZNF143.

La capacité de hStaf/ZNF143 à activer la transcription de gènes d'ARNm et d'ARNnc est liée à la présence de deux domaines fonctionnels distincts au sein de ce facteur. Le domaine protéique impliqué dans l'activation de la transcription des gènes d'ARNm est composé de 4 répétitions d'un motif de 15 acides aminés tandis qu'une région de 18 acides aminés est responsable de l'activation de la transcription des gènes d'ARNnc (Schuster et al., 1998). hStaf/ZNF143 reconnaît une séquence dégénérée de 18 pb grâce à son domaine de liaison à l'ADN constitué de 7 doigts de zinc de type C₂H₂ (Schaub et al., 1997; Schaub et al., 1999a; Schaub et al., 1999b; Schaub et al., 2000).

Le nombre croissant de gènes de protéines régulés par hStaf/ZNF143 nous a amené à nous intéresser à l'importance de ce facteur dans la transcription de ce type de gène à l'échelle du génome humain. Pour cela, nous avons combiné une approche bioinformatique et une approche biochimique qui nous ont permis de mettre en évidence un nombre inattendu de

SBS en amont de gènes d'ARNm humains (Myslinski et al., 2006). En utilisant la banque de données DBTSS, nous avons identifié 1175 SBS repartis dans 938 promoteurs de gènes de protéines chez l'homme et conservés dans les promoteurs des gènes orthologues chez la souris, le rat et le chien. 295 des régions promotrices identifiées par l'approche bioinformatique ont été recherchées par PCR dans un échantillon d'ADN immunoprécipité par des anticorps spécifiques de hStaf/ZNF143. Ces expériences de ChIP-PCR ont permis de montrer que hStaf/ZNF143 est lié à 90% des régions analysées. La banque DBTSS contenait au moment de cette étude 8793 séquences promotrices humaines. L'extrapolation des résultats de cette étude au génome humain laisse supposer qu'il existe 2500 SBS répartis parmi 2000 promoteurs de gènes de protéines. Ceci suggère que hStaf/ZNF143 est un facteur majeur de la transcription chez les mammifères.

La majorité des sites SBS identifiés lors de cette étude est localisée dans une région de 200pb en amont du +1. Les régions promotrices identifiées sont principalement localisées dans des îlots CpG et ne possèdent pas de boîte TATA. On peut observer que plusieurs SBS peuvent être présents au sein d'une même région promotrice. 15% des régions identifiées partagent un même SBS et sont organisées selon une orientation divergente. Ceci laisse supposer que hStaf/ZNF143 pourrait intervenir dans la régulation de la transcription initiée à partir de promoteurs bidirectionnels. Afin de déterminer l'implication de hStaf/ZNF143 dans ce type de mécanisme, différentes régions avec des TSS séparés par une distance maximale de 1 kpb et renfermant des SBS communs seront clonées dans un vecteur d'expression entre les gènes de la luciférase firefly et renilla placés en orientation inverse. La capacité de la région promotrice à activer la transcription dans les deux orientations sera analysée par la mesure de l'expression des deux luciférases après transfections dans des cellules COS7. Le rôle de hStaf/ZNF143 sera déterminé en mutant chacun des sites SBS de ces régions. Ces expériences sont en cours de réalisation au laboratoire.

De manière surprenante, nous avons identifié la présence d'une séquence ACTACAN localisée juste en amont de l'extrémité 5' des sites SBS pour environ la moitié des sites identifiés par l'approche bioinformatique. Les études réalisées par Xie *et al* montrent que ce motif est fréquemment retrouvé dans les promoteurs de gènes de protéines de quatre mammifères (Xie et al., 2005; Xie et al., 2007). La fonction de cette séquence ACTACAN n'a pas encore été établie. Il est possible qu'elle permette la fixation d'un facteur de transcription non caractérisé à ce jour ou qu'elle participe à une structuration de l'ADN qui favoriserait la

liaison optimale de hStaf/ZNF143 sur son site. La fixation d'un facteur particulier sur cette séquence pourra être montrée par des expériences de retard sur gel en ajoutant un extrait nucléaire à un fragment d'ADN renfermant les deux motifs successivement mutés. Cette séquence ACTACAN peut également être uniquement fonctionnelle dans certains tissus ou à certains stades du développement. L'identification d'un plus grand nombre de promoteurs renfermant un site SBS associé à cette séquence pourra permettre de vérifier cette hypothèse en mettant en évidence certains types de gènes spécifiquement associés à cette combinaison de motifs.

Il a également été montré que la simple présence de SBS en amont du gène de la luciférase est suffisante pour permettre l'expression de ce gène. Ainsi, hStaf/ZNF143 est capable de permettre à lui seul le recrutement de la machinerie de transcription. L'identification des facteurs interagissant avec hStaf/ZNF143 permettrait de mieux comprendre le mécanisme de la transcription des gènes d'ARNm dépendant de hStaf/ZNF143. Pour cela, on peut imaginer utiliser un fragment d'ADN biotinylé renfermant une région promotrice contenant des sites SBS afin de sélectionner le complexe se fixant à ce promoteur. Un extrait nucléaire sera ajouté au fragment d'ADN immobilisé sur des billes couplées à la streptavidine et les protéines retenues seront séparées par électrophorèse sur gel. La détermination de la nature des protéines pourra être réalisée par spectrométrie de masse ou par western-blot en utilisant des anticorps dirigés contre des facteurs généraux de la transcription.

L'analyse bioinformatique a révélé que 938 gènes de protéines contiennent des SBS dans leur promoteur. Parmi ces gènes, moins d'un tiers n'avait pas de fonction attribuée lors de notre analyse. Le classement des autres gènes en fonction du rôle de la protéine codée montre que hStaf/ZNF143 ne semble pas être impliqué spécifiquement dans la transcription d'une classe de gènes particuliers. Cependant, on estime qu'il existe environ 25000 gènes de protéines dans le génome. L'analyse d'un plus grand nombre de promoteurs de gènes de protéines pourrait peut-être mettre en évidence un enrichissement de promoteurs renfermant des SBS en amont de gènes dont le produit serait impliqué dans des fonctions particulières.

Notre analyse de la distribution des sites SBS dans le génome humain s'est limitée aux régions promotrices de gènes de protéines contenues dans la banque de données DBTSS. L'obtention du répertoire complet des sites SBS dans le génome humain permettrait d'identifier tous les gènes de protéines et d'ARNnc dont l'expression est sous la dépendance

du facteur hStaf/ZNF143. Ceci a été initié dans l'équipe en collaboration avec le Laboratoire de Bioinformatique et de Génomique Intégratives de l'IGBMC (Strasbourg) en recherchant des SBS dans la dernière version de la séquence du génome humain. Cette recherche bioinformatique a été réalisée en utilisant une approche de recherche par profil sur une version masquée du génome qui ne contient pas les séquences répétées. La matrice de score position spécifique utilisée a été construite à partir des 400 SBS validés expérimentalement lors de notre précédente étude (Myslinski et al., 2006). Cette étude a permis de mettre en évidence 230 000 SBS potentiels répartis sur les différents chromosomes humains.

Cette analyse bioinformatique se poursuit actuellement afin d'associer les SBS identifiés à des gènes de protéines ou d'ARNnc connus et à de nouveaux gènes. La corrélation entre la localisation des SBS et les promoteurs de gènes de protéines a déjà été réalisée. La cartographie obtenue montre que la répartition des SBS identifiés remplit les caractéristiques des sites de liaison de facteurs de transcription avec un enrichissement entre les positions -1000 et +100 par rapport au TSS. Comme cela a été observé pour les facteurs de transcription p53, cMyc ou Sp1, de nombreux sites SBS sont localisés à des positions non canoniques dans des gènes connus notamment dans leur premier intron (Cawley et al., 2004). Ces sites pourraient être impliqués dans les mécanismes de transcription sens-antisens importants par exemple dans la régulation de l'expression des gènes par ARN interférence.

L'analyse de la conservation des divers SBS identifiés lors de l'étude bioinformatique parmi les régions orthologues de plusieurs mammifères permettra de révéler les SBS qui ont de fortes probabilités d'être des éléments régulateurs de la transcription de gènes de protéines ou d'ARNnc. La fixation de hStaf/ZNF143 au niveau de ces régions pourra être validée expérimentalement par une approche de ChIP on chip (présentée dans la partie 2.2.2 de mon introduction). Les échantillons d'ADN immunoprécipité par des anticorps dirigés contre hStaf/ZNF143 ainsi que par des anticorps pré-immuns seront alors hybridés sur des puces qui porteront des oligonucléotides correspondant aux SBS identifiés. L'analyse des signaux obtenus indiquera les régions fixant effectivement hStaf/ZNF143 *in vivo*.

L'approche de CHIP-PET (présentée dans la partie 2.2.2 de mon introduction) en cours d'application au laboratoire devrait également permettre de constituer un répertoire global des SBS au sein du génome humain. Les parties 5' et 3' de l'ADN immunoprécipité par des anticorps spécifiques de hStaf/ZNF143 seront associées en une double étiquette spécifique de

chaque fragment. Ces étiquettes seront ensuite associées entre elles, clonées puis séquencées. L'analyse de données brutes de séquençage au sein du Laboratoire de Bioinformatique et de Génomique Intégratives de Strasbourg permettra de replacer les différentes étiquettes sur le génome humain et d'identifier les fragments renfermant un SBS.

Les analyses à haut débit du niveau de transcription des génomes de mammifères ont révélé un nombre inattendu de transcrits par rapport au nombre de gènes connus (Kapranov et al., 2002; Okazaki et al., 2002; Bertone et al., 2004; Carninci et al., 2005; Cheng et al., 2005). Ceci suggère l'existence de nombreux promoteurs de gènes d'ARNnc à identifier. Pour cela, nous pourrions rechercher la présence d'un élément promoteur PSE associé ou non à une boîte TATA, éléments promoteurs caractéristiques des gènes d'ARNnc, à proximité des SBS identifiés lors de l'analyse bioinformatique et non attribués à un gène de protéines ou d'ARNnc. Nous pourrions également mettre en relation la position des SBS et les résultats des études de transcriptomique afin de corréler la présence d'un SBS avec la synthèse d'un transcrit.

Les travaux effectués au cours de ma thèse montrent que hStaf/ZNF143 régule l'activité de gènes impliqués dans des voies cellulaires diverses via des promoteurs présentant une architecture variée (Figure 27). Nous avons également identifié un nombre inattendu de sites de liaison de ce facteur dans le génome de quatre mammifères. Ceci suggère que hStaf/ZNF143 est un facteur majeur dans le mécanisme de transcription des gènes de protéines chez les mammifères. On peut penser que ce facteur joue un rôle également essentiel dans la transcription d'un grand nombre de gènes d'ARNnc dont les promoteurs ne sont pas encore caractérisés à l'heure actuelle. De ce fait, les études à venir se focaliseront principalement sur l'identification et la caractérisation des promoteurs de ces gènes d'ARNnc dont la transcription est potentiellement sous le contrôle du facteur de transcription hStaf/ZNF143.

BIBLIOGRAPHIE

- Ach, R.A. and A.M. Weiner. 1991. Cooperation between CCAAT and octamer motifs in the distal sequence element of the rat U3 small nucleolar RNA promoter. *Nucleic Acids Res* **19**: 4209-18.
- Adachi, K., H. Saito, T. Tanaka, and T. Oka. 1998. Molecular cloning and characterization of the murine staf cDNA encoding a transcription activating factor for the selenocysteine tRNA gene in mouse mammary gland. *J Biol Chem* **273**: 8598-606.
- Adachi, N. and M.R. Lieber. 2002. Bidirectional gene organization: a common architectural feature of the human genome. *Cell* **109**: 807-9.
- Alam, T.I., T. Kanki, T. Muta, K. Ukaji, Y. Abe, H. Nakayama, K. Takio, N. Hamasaki, and D. Kang. 2003. Human mitochondrial DNA is packaged with TFAM. *Nucleic Acids Res* **31**: 1640-5.
- Ame, J.C., V. Schreiber, V. Fraulob, P. Dolle, G. de Murcia, and C.P. Niedergang. 2001. A bidirectional promoter connects the poly(ADP-ribose) polymerase 2 (PARP-2) gene to the gene for RNase P RNA. structure and expression of the mouse PARP-2 gene. *J Biol Chem* **276**: 11092-9.
- Angrand, P.O. 1993. Les domaines de liaison à l'ADN des facteurs de transcription eucaryotes. *Médecines/Sciences* **9**: 725-36.
- Antequera, F. 2003. Structure, function and evolution of CpG island promoters. *Cell Mol Life Sci* **60**: 1647-58.
- Antequera, F., D. Macleod, and A.P. Bird. 1989. Specific protection of methylated CpGs in mammalian nuclei. *Cell* **58**: 509-17.
- Ares, M., Jr., M. Mangin, and A.M. Weiner. 1985. Orientation-dependent transcriptional activator upstream of a human U2 snRNA gene. *Mol Cell Biol* **5**: 1560-70.
- B**achellerie, J.P., J. Cavaille, and A. Huttenhofer. 2002. The expanding snoRNA world. *Biochimie* **84**: 775-90.
- Baek, D., C. Davis, B. Ewing, D. Gordon, and P. Green. 2007. Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. *Genome Res* **17**: 145-55.
- Baer, M., T.W. Nilsen, C. Costigan, and S. Altman. 1990. Structure and transcription of a human gene for H1 RNA, the RNA component of human RNase P. *Nucleic Acids Res* **18**: 97-103.
- Baillat, D., M.A. Hakimi, A.M. Naar, A. Shilatifard, N. Cooch, and R. Shiekhattar. 2005. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**: 265-76.
- Bajic, V.B., S.L. Tan, A. Christoffels, C. Schonbach, L. Lipovich, L. Yang, O. Hofmann, A. Kruger, W. Hide, C. Kai et al. 2006. Mice and men: their promoter properties. *PLoS Genet* **2**: e54.
- Balakin, A.G., L. Smith, and M.J. Fournier. 1996. The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* **86**: 823-34.
- Baldi, A., V. Boccia, P.P. Claudio, A. De Luca, and A. Giordano. 1996. Genomic structure of the human retinoblastoma-related Rb2/p130 gene. *Proc Natl Acad Sci U S A* **93**: 4629-32.
- Bark, C., P. Weller, J. Zabielski, L. Janson, and U. Pettersson. 1987. A distant enhancer element is required for polymerase III transcription of a U6 RNA gene. *Nature* **328**: 356-9.

- Barski, O.A., V.Z. Papusha, G.R. Kunkel, and K.H. Gabbay. 2004. Regulation of aldehyde reductase expression by STAF and CHOP. *Genomics* **83**: 119-29.
- Bellizzi, D., S. Dato, P. Cavalcante, G. Covello, F. Di Cianni, G. Passarino, G. Rose, and G. De Benedictis. 2007. Characterization of a bidirectional promoter shared between two human genes related to aging: SIRT3 and PSMD13. *Genomics* **89**: 143-50.
- Bennett, J.L., S. Jeong-Yu, and D.A. Clayton. 1992. Characterization of a *Xenopus laevis* ribonucleoprotein endoribonuclease. Isolation of the RNA component and its expression during development. *J Biol Chem* **267**: 21765-72.
- Bernstein, B.E., C.L. Liu, E.L. Humphrey, E.O. Perlstein, and S.L. Schreiber. 2004. Global nucleosome occupancy in yeast. *Genome Biol* **5**: R62.
- Bertone, P., V. Stolc, T.E. Royce, J.S. Rozowsky, A.E. Urban, X. Zhu, J.L. Rinn, W. Tongprasit, M. Samanta, S. Weissman et al. 2004. Global identification of human transcribed sequences with genome tiling arrays. *Science* **306**: 2242-6.
- Bharadwaj, R. and H. Yu. 2004. The spindle checkpoint, aneuploidy, and cancer. *Oncogene* **23**: 2016-27.
- Birney, E. J.A. Stamatoyannopoulos A. Dutta R. Guigo T.R. Gingeras E.H. Margulies Z. Weng M. Snyder E.T. Dermitzakis R.E. Thurman et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**: 799-816.
- Blackwood, E.M. and J.T. Kadonaga. 1998. Going the distance: a current view of enhancer action. *Science* **281**: 60-3.
- Blake, M.C., R.C. Jambou, A.G. Swick, J.W. Kahn, and J.C. Azizkhan. 1990. Transcriptional initiation is controlled by upstream GC-box interactions in a TATAA-less promoter. *Mol Cell Biol* **10**: 6632-41.
- Borchert, G.M., W. Lanier, and B.L. Davidson. 2006. RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol* **13**: 1097-101.
- Boyer, L.A., T.I. Lee, M.F. Cole, S.E. Johnstone, S.S. Levine, J.P. Zucker, M.G. Guenther, R.M. Kumar, H.L. Murray, R.G. Jenner et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**: 947-56.
- Brandeis, M., D. Frank, I. Keshet, Z. Siegfried, M. Mendelsohn, A. Nemes, V. Temper, A. Razin, and H. Cedar. 1994. Sp1 elements protect a CpG island from de novo methylation. *Nature* **371**: 435-8.
- Branden, C. and J. Tooze. 1991. Introduction to protein structure *Graland Ed.*: 117.
- Breathnach, R. and P. Chambon. 1981. Organization and expression of eucaryotic split genes coding for proteins. *Annu Rev Biochem* **50**: 349-83.
- Brent, R. and M. Ptashne. 1985. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* **43**: 729-36.
- Brodsky, A.S., C.A. Meyer, I.A. Swinburne, G. Hall, B.J. Keenan, X.S. Liu, E.A. Fox, and P.A. Silver. 2005. Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. *Genome Biol* **6**: R64.
- Brosius, J. 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* **238**: 115-34.
- Burke, T.W. and J.T. Kadonaga. 1996. Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev* **10**: 711-24.
- Burke, T.W. and J.T. Kadonaga. 1997. The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes Dev* **11**: 3020-31.
- Burley, S.K. and R.G. Roeder. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu Rev Biochem* **65**: 769-99.

- Camon, E., D. Barrell, V. Lee, E. Dimmer, and R. Apweiler. 2004a. The Gene Ontology Annotation (GOA) Database--an integrated resource of GO annotations to the UniProt Knowledgebase. *In Silico Biol* **4**: 5-6.
- Camon, E., M. Magrane, D. Barrell, D. Binns, W. Fleischmann, P. Kersey, N. Mulder, T. Oinn, J. Maslen, A. Cox et al. 2003. The Gene Ontology Annotation (GOA) project: implementation of GO in SWISS-PROT, TrEMBL, and InterPro. *Genome Res* **13**: 662-72.
- Camon, E., M. Magrane, D. Barrell, V. Lee, E. Dimmer, J. Maslen, D. Binns, N. Harte, R. Lopez, and R. Apweiler. 2004b. The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res* **32**: D262-6.
- Carbon, P. and A. Krol. 1991. Transcription of the *Xenopus laevis* selenocysteine tRNA(Ser)Sec gene: a system that combines an internal B box and upstream elements also found in U6 snRNA genes. *Embo J* **10**: 599-606.
- Carbon, P., S. Murgo, J.P. Ebel, A. Krol, G. Tebb, and L.W. Mattaj. 1987. A common octamer motif binding protein is involved in the transcription of U6 snRNA by RNA polymerase III and U2 snRNA by RNA polymerase II. *Cell* **51**: 71-9.
- Carcamo, J., L. Buckbinder, and D. Reinberg. 1991. The initiator directs the assembly of a transcription factor IID-dependent transcription complex. *Proc Natl Acad Sci U S A* **88**: 8052-6.
- Carninci, P. T. Kasukawa S. Katayama J. Gough M.C. Frith N. Maeda R. Oyama T. Ravasi B. Lenhard C. Wells et al. 2005. The transcriptional landscape of the mammalian genome. *Science* **309**: 1559-63.
- Carninci, P., C. Kvam, A. Kitamura, T. Ohsumi, Y. Okazaki, M. Itoh, M. Kamiya, K. Shibata, N. Sasaki, M. Izawa et al. 1996. High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics* **37**: 327-36.
- Carninci, P., A. Sandelin, B. Lenhard, S. Katayama, K. Shimokawa, J. Ponjavic, C.A. Semple, M.S. Taylor, P.G. Engstrom, M.C. Frith et al. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38**: 626-35.
- Carninci, P., A. Westover, Y. Nishiyama, T. Ohsumi, M. Itoh, S. Nagaoka, N. Sasaki, Y. Okazaki, M. Muramatsu, C. Schneider et al. 1997. High efficiency selection of full-length cDNA by improved biotinylated cap trapper. *DNA Res* **4**: 61-6.
- Cawley, S., S. Bekiranov, H.H. Ng, P. Kapranov, E.A. Sekinger, D. Kampa, A. Piccolboni, V. Sementchenko, J. Cheng, A.J. Williams et al. 2004. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**: 499-509.
- Chalkley, G.E. and C.P. Verrijzer. 1999. DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *Embo J* **18**: 4835-45.
- Chang, D.D. and D.A. Clayton. 1985. Priming of human mitochondrial DNA replication occurs at the light-strand promoter. *Proc Natl Acad Sci U S A* **82**: 351-5.
- Chang, D.D., W.W. Hauswirth, and D.A. Clayton. 1985. Replication priming and transcription initiate from precisely the same site in mouse mitochondrial DNA. *Embo J* **4**: 1559-67.
- Cheng, J., P. Kapranov, J. Drenkow, S. Dike, S. Brubaker, S. Patel, J. Long, D. Stern, H. Tammana, G. Helt et al. 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**: 1149-54.
- Ciliberto, G., R. Buckland, R. Cortese, and L. Philipson. 1985. Transcription signals in embryonic *Xenopus laevis* U1 RNA genes. *Embo J* **4**: 1537-43.
- Ciliberto, G., F. Palla, G. Tebb, I.W. Mattaj, and L. Philipson. 1987. Properties of a U1 RNA enhancer-like sequence. *Nucleic Acids Res* **15**: 2403-16.

- Coisy, M., V. Roure, M. Ribot, A. Philips, C. Muchardt, J.M. Blanchard, and J.C. Dantonel. 2004. Cyclin A repression in quiescent cells is associated with chromatin remodeling of its promoter and requires Brahma/SNF2alpha. *Mol Cell* **15**: 43-56.
- Consortium, I.H.G.S. 2004. Finishing the euchromatic sequence of the human genome. *Nature* **431**: 931-45.
- Cooper, S.J., N.D. Trinklein, E.D. Anton, L. Nguyen, and R.M. Myers. 2006. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res* **16**: 1-10.
- Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**: 10881-90.
- Crooks, G.E., G. Hon, J.M. Chandonia, and S.E. Brenner. 2004. WebLogo: a sequence logo generator. *Genome Res* **14**: 1188-90.
- Cuello, P., D.C. Boyd, M.J. Dye, N.J. Proudfoot, and S. Murphy. 1999. Transcription of the human U2 snRNA genes continues beyond the 3' box in vivo. *Embo J* **18**: 2867-77.
- Cullen, B.R. 2004. Transcription and processing of human microRNA precursors. *Mol Cell* **16**: 861-5.

- D**ahlberg, J.E. and E.T. Schenborn. 1988. The human U1 snRNA promoter and enhancer do not direct synthesis of messenger RNA. *Nucleic Acids Res* **16**: 5827-40.
- Dai, W., Q. Wang, T. Liu, M. Swamy, Y. Fang, S. Xie, R. Mahmood, Y.M. Yang, M. Xu, and C.V. Rao. 2004. Slippage of mitotic arrest and enhanced tumor development in mice with BubR1 haploinsufficiency. *Cancer Res* **64**: 440-5.
- Darzacq, X., B.E. Jady, C. Verheggen, A.M. Kiss, E. Bertrand, and T. Kiss. 2002. Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *Embo J* **21**: 2746-56.
- Davenport, J.W., E.R. Fernandes, L.D. Harris, G.A. Neale, and R. Goorha. 1999. The mouse mitotic checkpoint gene bub1b, a novel bub1 family member, is expressed in a cell cycle-dependent manner. *Genomics* **55**: 113-7.
- Deng, W. and S.G. Roberts. 2005. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev* **19**: 2418-23.
- Desjarlais, J.R. and J.M. Berg. 1993. Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. *Proc Natl Acad Sci U S A* **90**: 2256-60.
- Di Leva, F., M.I. Ferrante, F. Demarchi, A. Caravelli, M.R. Matarazzo, M. Giacca, M. D'Urso, M. D'Esposito, and A. Franze. 2004. Human synaptobrevin-like 1 gene basal transcription is regulated through the interaction of selenocysteine tRNA gene transcription activating factor-zinc finger 143 factors with evolutionary conserved cis-elements. *J Biol Chem* **279**: 7734-9.
- Diakun, G.P., L. Fairall, and A. Klug. 1986. EXAFS study of the zinc-binding sites in the protein transcription factor IIIA. *Nature* **324**: 698-9.

- E**kstrand, M.I., M. Falkenberg, A. Rantanen, C.B. Park, M. Gaspari, K. Hultenby, P. Rustin, C.M. Gustafsson, and N.G. Larsson. 2004. Mitochondrial transcription factor A regulates mtDNA copy number in mammals. *Hum Mol Genet* **13**: 935-44.
- Euskirchen, G., T.E. Royce, P. Bertone, R. Martone, J.L. Rinn, F.K. Nelson, F. Sayward, N.M. Luscombe, P. Miller, M. Gerstein et al. 2004. CREB binds to multiple loci on human chromosome 22. *Mol Cell Biol* **24**: 3804-14.

- Euskirchen, G.M., J.S. Rozowsky, C.L. Wei, W.H. Lee, Z.D. Zhang, S. Hartman, O. Emanuelsson, V. Stolc, S. Weissman, M.B. Gerstein et al. 2007. Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies. *Genome Res* **17**: 898-909.
- Evans, R., J.A. Fairley, and S.G. Roberts. 2001. Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. *Genes Dev* **15**: 2945-9.
- F**ajás, L., L. Le Cam, J. Polanowska, E. Fabbrizio, N. Servant, A. Philips, G. Carnac, and C. Sardet. 2000. A CDE/CHR-like element mediates repression of transcription of the mouse RB2 (p130) gene. *FEBS Lett* **471**: 29-33.
- Falkenberg, M., M. Gaspari, A. Rantanen, A. Trifunovic, N.G. Larsson, and C.M. Gustafsson. 2002. Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. *Nat Genet* **31**: 289-94.
- Falkenberg, M., N.G. Larsson, and C.M. Gustafsson. 2007. DNA replication and transcription in mammalian mitochondria. *Annu Rev Biochem* **76**: 679-99.
- Fisher, R.P. and D.A. Clayton. 1985. A transcription factor required for promoter recognition by human mitochondrial RNA polymerase. Accurate initiation at the heavy- and light-strand promoters dissected and reconstituted in vitro. *J Biol Chem* **260**: 11330-8.
- Fisher, R.P. and D.A. Clayton. 1988. Purification and characterization of human mitochondrial transcription factor 1. *Mol Cell Biol* **8**: 3496-509.
- Fisher, R.P., T. Lisowsky, M.A. Parisi, and D.A. Clayton. 1992. DNA wrapping and bending by a mitochondrial high mobility group-like transcriptional activator protein. *J Biol Chem* **267**: 3358-67.
- Fisher, R.P., M.A. Parisi, and D.A. Clayton. 1989. Flexible recognition of rapidly evolving promoter sequences by mitochondrial transcription factor 1. *Genes Dev* **3**: 2202-17.
- FitzGerald, P.C., A. Shlyakhtenko, A.A. Mir, and C. Vinson. 2004. Clustering of DNA sequences in human promoters. *Genome Res* **14**: 1562-74.
- Fletcher, C., N. Heintz, and R.G. Roeder. 1987. Purification and characterization of OTF-1, a transcription factor regulating cell cycle expression of a human histone H2b gene. *Cell* **51**: 773-81.
- Ford, E. and N. Hernandez. 1997. Characterization of a trimeric complex containing Oct-1, SNAPc, and DNA. *J Biol Chem* **272**: 16048-55.
- Ford, E., M. Strubin, and N. Hernandez. 1998. The Oct-1 POU domain activates snRNA gene transcription by contacting a region in the SNAPc largest subunit that bears sequence similarities to the Oct-1 coactivator OBF-1. *Genes Dev* **12**: 3528-40.
- Fry, C.J. and C.L. Peterson. 2001. Chromatin remodeling enzymes: who's on first? *Curr Biol* **11**: R185-97.
- G**allagher, P.G., D.G. Nilson, C. Wong, J.L. Weisbein, L.J. Garrett-Beal, S.W. Eber, and D.M. Bodine. 2005. A dinucleotide deletion in the ankyrin promoter alters gene expression, transcription initiation and TFIID complex formation in hereditary spherocytosis. *Hum Mol Genet* **14**: 2501-9.
- Ganot, P., M. Caizergues-Ferrer, and T. Kiss. 1997. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev* **11**: 941-56.

- Garrido, N., L. Griparic, E. Jokitalo, J. Wartiovaara, A.M. van der Blik, and J.N. Spelbrink. 2003. Composition and dynamics of human mitochondrial nucleoids. *Mol Biol Cell* **14**: 1583-96.
- Gaszner, M. and G. Felsenfeld. 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* **7**: 703-13.
- Gerard, M.A., A. Krol, and P. Carbon. 2007a. The scaRNA2 gene is encoded by an independent unit and its transcription is regulated by hStaf/ZNF143. *Soumis pour publication dans Nucleic Acids Research*.
- Gerard, M.A., A. Krol, and P. Carbon. 2007b. Transcription factor hStaf/ZNF143 is required for expression of the human TFAM gene. *Gene* **401**: 145-53.
- Gilbert, S.L. and P.A. Sharp. 1999. Promoter-specific hypoacetylation of X-inactivated genes. *Proc Natl Acad Sci U S A* **96**: 13825-30.
- Goldberg, M.L. 1979. Sequence analysis of *Drosophila* histone genes. *Ph.D dissertation, Stanford University, California*.
- Gross, P. and T. Oelgeschlager. 2006. Core promoter-selective RNA polymerase II transcription. *Biochem Soc Symp*: 225-36.
- Grosschedl, R. and M.L. Birnstiel. 1980. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. *Proc Natl Acad Sci U S A* **77**: 1432-6.
- Grossman, C.E., Y. Qian, K. Banki, and A. Perl. 2004. ZNF143 mediates basal and tissue-specific expression of human transaldolase. *J Biol Chem* **279**: 12190-205.
- Grosveld, G.C., C.K. Shewmaker, P. Jat, and R.A. Flavell. 1981. Localization of DNA sequences necessary for transcription of the rabbit beta-globin gene in vitro. *Cell* **25**: 215-26.
- Gu, J., Y. Chen, and R. Reddy. 1998. Small RNA database. *Nucleic Acids Res* **26**: 160-2.
- Guo, Z., G.H. Boekhoudt, and J.M. Boss. 2003. Role of the intronic enhancer in tumor necrosis factor-mediated induction of manganese superoxide dismutase. *J Biol Chem* **278**: 23570-8.
- H**arbers, M. and P. Carninci. 2005. Tag-based approaches for transcriptome research and genome annotation. *Nat Methods* **2**: 495-502.
- Hardwick, K.G., R.C. Johnston, D.L. Smith, and A.W. Murray. 2000. MAD3 encodes a novel component of the spindle checkpoint which interacts with Bub3p, Cdc20p, and Mad2p. *J Cell Biol* **148**: 871-82.
- Hashimoto, S., Y. Suzuki, Y. Kasai, K. Morohoshi, T. Yamada, J. Sese, S. Morishita, S. Sugano, and K. Matsushima. 2004. 5'-end SAGE for the analysis of transcriptional start sites. *Nat Biotechnol* **22**: 1146-9.
- Hawkins, R.D. and B. Ren. 2006. Genome-wide location analysis: insights on transcriptional regulation. *Hum Mol Genet* **15 Spec No 1**: R1-7.
- Hernandez, N. 1992. Transcription of vertebrate snRNA genes and related genes. Dans MacKnight, S.L. and Yamamoto, K.R. (Ed.), *Transcriptional regulation, Cold Spring Harbor Press, New York* **1**: 281-313.
- Hernandez, N. 2001. Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J Biol Chem* **276**: 26733-6.
- Hernandez, N. and R. Lucito. 1988. Elements required for transcription initiation of the human U2 snRNA gene coincide with elements required for snRNA 3' end formation. *Embo J* **7**: 3125-34.
- Hu, S.L. and J.L. Manley. 1981. DNA sequence required for initiation of transcription in vitro from the major late promoter of adenovirus 2. *Proc Natl Acad Sci U S A* **78**: 820-4.

- Huet, X., J. Rech, A. Plet, A. Vie, and J.M. Blanchard. 1996. Cyclin A expression is under negative transcriptional control during the cell cycle. *Mol Cell Biol* **16**: 3789-98.
- Huttenhofer, A., M. Kiefmann, S. Meier-Ewert, J. O'Brien, H. Lehrach, J.P. Bachellerie, and J. Brosius. 2001. RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *Embo J* **20**: 2943-53.
- I**mpey, S., S.R. McCorkle, H. Cha-Molstad, J.M. Dwyer, G.S. Yochum, J.M. Boss, S. McWeeney, J.J. Dunn, G. Mandel, and R.H. Goodman. 2004. Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell* **119**: 1041-54.
- Ishiguchi, H., H. Izumi, T. Torigoe, Y. Yoshida, H. Kubota, S. Tsuji, and K. Kohno. 2004. ZNF143 activates gene expression in response to DNA damage and binds to cisplatin-modified DNA. *Int J Cancer* **111**: 900-9.
- J**avahery, R., A. Khachi, K. Lo, B. Zenzie-Gregory, and S.T. Smale. 1994. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol Cell Biol* **14**: 116-27.
- Ji, H., S.A. Vokes, and W.H. Wong. 2006. A comparative analysis of genome-wide chromatin immunoprecipitation data for mammalian transcription factors. *Nucleic Acids Res* **34**: e146.
- Johnson, D.S., A. Mortazavi, R.M. Myers, and B. Wold. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**: 1497-502.
- Juven-Gershon, T., J.Y. Hsu, and J.T. Kadonaga. 2006. Perspectives on the RNA polymerase II core promoter. *Biochem Soc Trans* **34**: 1047-50.
- K**ampa, D., J. Cheng, P. Kapranov, M. Yamanaka, S. Brubaker, S. Cawley, J. Drenkow, A. Piccolboni, S. Bekiranov, G. Helt et al. 2004. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res* **14**: 331-42.
- Kapranov, P., S.E. Cawley, J. Drenkow, S. Bekiranov, R.L. Strausberg, S.P. Fodor, and T.R. Gingeras. 2002. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**: 916-9.
- Kapranov, P., J. Cheng, S. Dike, D.A. Nix, R. Duttgupta, A.T. Willingham, P.F. Stadler, J. Hertel, J. Hackermuller, I.L. Hofacker et al. 2007a. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**: 1484-8.
- Kapranov, P., J. Drenkow, J. Cheng, J. Long, G. Helt, S. Dike, and T.R. Gingeras. 2005. Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res* **15**: 987-97.
- Kapranov, P., A.T. Willingham, and T.R. Gingeras. 2007b. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet* **8**: 413-23.
- Katayama, S., Y. Tomaru, T. Kasukawa, K. Waki, M. Nakanishi, M. Nakamura, H. Nishida, C.C. Yap, M. Suzuki, J. Kawai et al. 2005. Antisense transcription in the mammalian transcriptome. *Science* **309**: 1564-6.
- Kim, J., A.A. Bhinge, X.C. Morgan, and V.R. Iyer. 2005a. Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. *Nat Methods* **2**: 47-53.

- Kim, T.H., L.O. Barrera, M. Zheng, C. Qu, M.A. Singer, T.A. Richmond, Y. Wu, R.D. Green, and B. Ren. 2005b. A high-resolution map of active promoters in the human genome. *Nature* **436**: 876-80.
- Kimura, K., A. Wakamatsu, Y. Suzuki, T. Ota, T. Nishikawa, R. Yamashita, J. Yamamoto, M. Sekine, K. Tsuritani, H. Wakaguri et al. 2006. Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res* **16**: 55-65.
- Kiss, T. 2001. Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *Embo J* **20**: 3617-22.
- Kiss-Laszlo, Z., Y. Henry, and T. Kiss. 1998. Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *Embo J* **17**: 797-807.
- Kouzarides, T. 2007. Chromatin modifications and their function. *Cell* **128**: 693-705.
- Koyanagi, K.O., M. Hagiwara, T. Itoh, T. Gojobori, and T. Imanishi. 2005. Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system. *Gene* **353**: 169-76.
- Krol, A., P. Carbon, J.P. Ebel, and B. Appel. 1987. Xenopus tropicalis U6 snRNA genes transcribed by Pol III contain the upstream promoter elements used by Pol II dependent U snRNA genes. *Nucleic Acids Res* **15**: 2463-78.
- Kubota, H., S. Yokota, H. Yanagi, and T. Yura. 2000. Transcriptional regulation of the mouse cytosolic chaperonin subunit gene Cct4/t-complex polypeptide 1 by selenocysteine tRNA gene transcription activating factor family zinc finger proteins. *J Biol Chem* **275**: 28641-8.
- Kunkel, G.R. and T. Pederson. 1988. Upstream elements required for efficient transcription of a human U6 RNA gene resemble those of U1 and U2 genes even though a different polymerase is used. *Genes Dev* **2**: 196-204.
- Kutach, A.K. and J.T. Kadonaga. 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Mol Cell Biol* **20**: 4754-64.
- L**agrange, T., A.N. Kapanidis, H. Tang, D. Reinberg, and R.H. Ebricht. 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev* **12**: 34-44.
- Landry, J.R., D.L. Mager, and B.T. Wilhelm. 2003. Complex controls: the role of alternative promoters in mammalian genomes. *Trends Genet* **19**: 640-8.
- Lange-zu Dohna, C., M. Brandeis, F. Berr, J. Mossner, and K. Engeland. 2000. A CDE/CHR tandem element regulates cell cycle-dependent repression of cyclin B2 transcription. *FEBS Lett* **484**: 77-81.
- Lavorgna, G., D. Dahary, B. Lehner, R. Sorek, C.M. Sanderson, and G. Casari. 2004. In search of antisense. *Trends Biochem Sci* **29**: 88-94.
- Lee, C.K., Y. Shibata, B. Rao, B.D. Strahl, and J.D. Lieb. 2004a. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat Genet* **36**: 900-5.
- Lee, M.P., K. Howcroft, A. Kotekar, H.H. Yang, K.H. Buetow, and D.S. Singer. 2005. ATG deserts define a novel core promoter subclass. *Genome Res* **15**: 1189-97.
- Lee, M.S., G.P. Gippert, K.V. Soman, D.A. Case, and P.E. Wright. 1989. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* **245**: 635-7.
- Lee, T.I., R.G. Jenner, L.A. Boyer, M.G. Guenther, S.S. Levine, R.M. Kumar, B. Chevalier, S.E. Johnstone, M.F. Cole, K. Isono et al. 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**: 301-13.

- Lee, Y., M. Kim, J. Han, K.H. Yeom, S. Lee, S.H. Baek, and V.N. Kim. 2004b. MicroRNA genes are transcribed by RNA polymerase II. *Embo J* **23**: 4051-60.
- Lemon, B. and R. Tjian. 2000. Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev* **14**: 2551-69.
- Lengauer, C., K.W. Kinzler, and B. Vogelstein. 1998. Genetic instabilities in human cancers. *Nature* **396**: 643-9.
- Lescure, A., P. Carbon, and A. Krol. 1991. The different positioning of the proximal sequence element in the *Xenopus* RNA polymerase II and III snRNA promoters is a key determinant which confers RNA polymerase III specificity. *Nucleic Acids Res* **19**: 435-41.
- Lescure, A., S. Murgo, P. Carbon, and A. Krol. 1992. The proximal promoter and the start site cooperate to specify correct U1 snRNA transcription initiation by RNA polymerase II. *Nucleic Acids Res* **20**: 1573-8.
- Li, B., M. Carey, and J.L. Workman. 2007. The role of chromatin during transcription. *Cell* **128**: 707-19.
- Li, F. and D.C. Altieri. 1999. The cancer antiapoptosis mouse survivin gene: characterization of locus and transcriptional requirements of basal and cell cycle-dependent expression. *Cancer Res* **59**: 3143-51.
- Li, L., S. He, J.M. Sun, and J.R. Davie. 2004. Gene regulation by Sp1 and Sp3. *Biochem Cell Biol* **82**: 460-71.
- Li, W., Z. Lan, H. Wu, S. Wu, J. Meadows, J. Chen, V. Zhu, and W. Dai. 1999. BUBR1 phosphorylation is regulated during mitotic checkpoint activation. *Cell Growth Differ* **10**: 769-75.
- Lim, C.Y., B. Santoso, T. Boulay, E. Dong, U. Ohler, and J.T. Kadonaga. 2004. The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev* **18**: 1606-17.
- Lin, J.M., P.J. Collins, N.D. Trinklein, Y. Fu, H. Xi, R.M. Myers, and Z. Weng. 2007. Transcription factor binding and modified histones in human bidirectional promoters. *Genome Res* **17**: 818-27.
- Litt, M.D., I.K. Hornstra, and T.P. Yang. 1996. In vivo footprinting and high-resolution methylation analysis of the mouse hypoxanthine phosphoribosyltransferase gene 5' region on the active and inactive X chromosomes. *Mol Cell Biol* **16**: 6190-9.
- Littlefield, O., Y. Korkhin, and P.B. Sigler. 1999. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc Natl Acad Sci U S A* **96**: 13668-73.
- Liu, N., F.C. Lucibello, K. Engeland, and R. Muller. 1998. A new model of cell cycle-regulated transcription: repression of the cyclin A promoter by CDF-1 and anti-repression by E2F. *Oncogene* **16**: 2957-63.
- Lo, K. and S.T. Smale. 1996. Generality of a functional initiator consensus sequence. *Gene* **182**: 13-22.
- Lobo, S.M. and N. Hernandez. 1989. A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter. *Cell* **58**: 55-67.
- Lobo, S.M. and N. Hernandez. 1994. Transcription of snRNA genes by RNA polymerase II and III. *Dans Conaway, R.C. and Conaway, J.W. (Ed.), Transcription, mechanisms and regulation, Raven Press, New York*: 127-59.
- Lobo, S.M., S. Ifill, and N. Hernandez. 1990. cis-acting elements required for RNA polymerase II and III transcription in the human U2 and U6 snRNA promoters. *Nucleic Acids Res* **18**: 2891-9.
- Loh, Y.H., Q. Wu, J.L. Chew, V.B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu et al. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**: 431-40.

- Lucibello, F.C., M. Truss, J. Zwicker, F. Ehlert, M. Beato, and R. Muller. 1995. Periodic cdc25C transcription is mediated by a novel cell cycle-regulated repressor element (CDE). *Embo J* **14**: 132-42.
- Lund, E. and J.E. Dahlberg. 1988. The genes and transcription of the major small nuclear RNAs. Dans M.L. Birnstiel (Ed.), *Structure and function of major and minor small nuclear ribonucleoprotein particles*. Springer-Verlag, Berlin: 38-70.
- Luo, Y., H. Fujii, T. Gerster, and R.G. Roeder. 1992. A novel B cell-derived coactivator potentiates the activation of immunoglobulin promoters by octamer-binding transcription factors. *Cell* **71**: 231-41.
- M**ach, C.M., B.W. Hargrove, and G.R. Kunkel. 2002. The Small RNA gene activator protein, SphI postoctamer homology-binding factor/selenocysteine tRNA gene transcription activating factor, stimulates transcription of the human interferon regulatory factor-3 gene. *J Biol Chem* **277**: 4853-8.
- Macleod, D., J. Charlton, J. Mullins, and A.P. Bird. 1994. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes Dev* **8**: 2282-92.
- Maraia, R.J., N. Sasaki-Tozawa, C.T. Driscoll, E.D. Green, and G.J. Darlington. 1994. The human Y4 small cytoplasmic RNA gene is controlled by upstream elements and resides on chromosome 7 with all other hY scRNA genes. *Nucleic Acids Res* **22**: 3045-52.
- Martens, J.A., L. Laprade, and F. Winston. 2004. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* **429**: 571-4.
- Martone, R., G. Euskirchen, P. Bertone, S. Hartman, T.E. Royce, N.M. Luscombe, J.L. Rinn, F.K. Nelson, P. Miller, M. Gerstein et al. 2003. Distribution of NF-kappaB-binding sites across human chromosome 22. *Proc Natl Acad Sci U S A* **100**: 12247-52.
- Maruyama, K. and S. Sugano. 1994. Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* **138**: 171-4.
- Mattaj, I.W., N.A. Dathan, H.D. Parry, P. Carbon, and A. Krol. 1988. Changing the RNA polymerase specificity of U snRNA gene promoters. *Cell* **55**: 435-42.
- Mattaj, I.W., S. Lienhard, J. Jiricny, and E.M. De Robertis. 1985. An enhancer-like sequence within the *Xenopus* U2 gene promoter facilitates the formation of stable transcription complexes. *Nature* **316**: 163-7.
- McCulloch, V. and G.S. Shadel. 2003. Human mitochondrial transcription factor B1 interacts with the C-terminal activation region of h-mtTFA and stimulates transcription independently of its RNA methyltransferase activity. *Mol Cell Biol* **23**: 5816-24.
- McNamara, K.J. and W.E. Stumph. 1990. Site-directed mutational analysis of a U4 small nuclear RNA gene proximal sequence element. Localization and identification of functional nucleotides. *J Biol Chem* **265**: 9728-31.
- McNamara, K.J., R.J. Walker, K.A. Roebuck, and W.E. Stumph. 1987. Transcriptional signals of a U4 small nuclear RNA gene. *Nucleic Acids Res* **15**: 9239-54.
- Medlin, J.E., P. Uguen, A. Taylor, D.L. Bentley, and S. Murphy. 2003. The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *Embo J* **22**: 925-34.
- Millband, D.N. and K.G. Hardwick. 2002. Fission yeast Mad3p is required for Mad2p to inhibit the anaphase-promoting complex and localizes to kinetochores in a Bub1p-, Bub3p-, and Mph1p-dependent manner. *Mol Cell Biol* **22**: 2728-42.
- Miller, J., A.D. McLachlan, and A. Klug. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *Embo J* **4**: 1609-14.

- Mittal, V., M.A. Cleary, W. Herr, and N. Hernandez. 1996. The Oct-1 POU-specific domain can stimulate small nuclear RNA gene transcription by stabilizing the basal transcription complex SNAPc. *Mol Cell Biol* **16**: 1955-65.
- Mittal, V., B. Ma, and N. Hernandez. 1999. SNAP(c): a core promoter factor with a built-in DNA-binding damper that is deactivated by the Oct-1 POU domain. *Genes Dev* **13**: 1807-21.
- Muller, F., M.A. Demeny, and L. Tora. 2007. New problems in RNA polymerase II transcription initiation: matching the diversity of core promoters with a variety of promoter recognition factors. *J Biol Chem* **282**: 14685-9.
- Murphy, S. 1997. Differential in vivo activation of the class II and class III snRNA genes by the POU-specific domain of Oct-1. *Nucleic Acids Res* **25**: 2068-76.
- Murphy, S., C. Di Liegro, and M. Melli. 1987. The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. *Cell* **51**: 81-7.
- Murphy, S., M. Tripodi, and M. Melli. 1986. A sequence upstream from the coding region is required for the transcription of the 7SK RNA genes. *Nucleic Acids Res* **14**: 9243-60.
- Murphy, S., J.B. Yoon, T. Gerster, and R.G. Roeder. 1992. Oct-1 and Oct-2 potentiate functional interactions of a transcription factor with the proximal sequence element of small nuclear RNA genes. *Mol Cell Biol* **12**: 3247-61.
- Musacchio, A. and E.D. Salmon. 2007. The spindle-assembly checkpoint in space and time. *Nat Rev Mol Cell Biol* **8**: 379-93.
- Myslinski, E., J.C. Ame, A. Krol, and P. Carbon. 2001. An unusually compact external promoter for RNA polymerase III transcription of the human H1RNA gene. *Nucleic Acids Res* **29**: 2502-9.
- Myslinski, E., M.A. Gerard, A. Krol, and P. Carbon. 2006. A genome scale location analysis of human Staf/ZNF143-binding sites suggests a widespread role for human Staf/ZNF143 in mammalian promoters. *J Biol Chem* **281**: 39953-62.
- Myslinski, E., M.A. Gerard, A. Krol, and P. Carbon. 2007. Transcription of the human cell cycle regulated BUB1B gene requires hStaf/ZNF143. *Nucleic Acids Res* **35**: 3453-64.
- Myslinski, E., A. Krol, and P. Carbon. 1992. Optimal tRNA((Ser)Sec) gene activity requires an upstream SPH motif. *Nucleic Acids Res* **20**: 203-9.
- Myslinski, E., A. Krol, and P. Carbon. 1998. ZNF76 and ZNF143 are two human homologs of the transcriptional activator Staf. *J Biol Chem* **273**: 21998-2006.
- Myslinski, E., A. Krol, and P. Carbon. 2004. Characterization of snRNA and snRNA-type genes in the pufferfish *Fugu rubripes*. *Gene* **330**: 149-58.
- Myslinski, E., C. Schuster, J. Huet, A. Sentenac, A. Krol, and P. Carbon. 1993. Point mutations 5' to the tRNA selenocysteine TATA box alter RNA polymerase III transcription by affecting the binding of TBP. *Nucleic Acids Res* **21**: 5852-8.
- Ng, P., C.L. Wei, W.K. Sung, K.P. Chiu, L. Lipovich, C.C. Ang, S. Gupta, A. Shahab, A. Ridwan, C.H. Wong et al. 2005. Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. *Nat Methods* **2**: 105-11.
- Nicklas, R.B., S.C. Ward, and G.J. Gorbsky. 1995. Kinetochore chemistry is sensitive to tension and may link mitotic forces to a cell cycle checkpoint. *J Cell Biol* **130**: 929-39.
- Nikolov, D.B., H. Chen, E.D. Halay, A.A. Usheva, K. Hisatake, D.K. Lee, R.G. Roeder, and S.K. Burley. 1995. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**: 119-28.

- Nishida, H., T. Suzuki, S. Kondo, H. Miura, Y. Fujimura, and Y. Hayashizaki. 2006. Histone H3 acetylated at lysine 9 in promoter is associated with low nucleosome density in the vicinity of transcription start site in human cell. *Chromosome Res* **14**: 203-11.
- Nolte, R.T., R.M. Conlin, S.C. Harrison, and R.S. Brown. 1998. Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proc Natl Acad Sci U S A* **95**: 2938-43.
- O**'Shea-Greenfield, A. and S.T. Smale. 1992. Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. *J Biol Chem* **267**: 1391-402.
- Ohler, U., G.C. Liao, H. Niemann, and G.M. Rubin. 2002. Computational analysis of core promoters in the Drosophila genome. *Genome Biol* **3**: RESEARCH0087.
- Okazaki, Y. M. Furuno T. Kasukawa J. Adachi H. Bono S. Kondo I. Nikaido N. Osato R. Saito H. Suzuki et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**: 563-73.
- Orphanides, G., T. Lagrange, and D. Reinberg. 1996. The general transcription factors of RNA polymerase II. *Genes Dev* **10**: 2657-83.
- P**arisi, M.A. and D.A. Clayton. 1991. Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. *Science* **252**: 965-9.
- Parry, H.D., G. Tebb, and I.W. Mattaj. 1989. The Xenopus U2 gene PSE is a single, compact, element required for transcription initiation and 3' end formation. *Nucleic Acids Res* **17**: 3633-44.
- Parslow, T.G., D.L. Blair, W.J. Murphy, and D.K. Granner. 1984. Structure of the 5' ends of immunoglobulin genes: a novel conserved sequence. *Proc Natl Acad Sci U S A* **81**: 2650-4.
- Patikoglou, G.A., J.L. Kim, L. Sun, S.H. Yang, T. Kodadek, and S.K. Burley. 1999. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev* **13**: 3217-30.
- Paule, M.R. and R.J. White. 2000. Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res* **28**: 1283-98.
- Pavletich, N.P. and C.O. Pabo. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**: 809-17.
- Pavletich, N.P. and C.O. Pabo. 1993. Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* **261**: 1701-7.
- Peters, J.M. 2006. The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nat Rev Mol Cell Biol* **7**: 644-56.
- Philips, A., S. Chambeyron, N. Lamb, A. Vie, and J.M. Blanchard. 1999. CHF: a novel factor binding to cyclin A CHR corepressor element. *Oncogene* **18**: 6222-32.
- Piantadosi, C.A. and H.B. Suliman. 2006. Mitochondrial transcription factor A induction by redox activation of nuclear respiratory factor 1. *J Biol Chem* **281**: 324-33.
- Pokholok, D.K., C.T. Harbison, S. Levine, M. Cole, N.M. Hannett, T.I. Lee, G.W. Bell, K. Walker, P.A. Rolfe, E. Herbolsheimer et al. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**: 517-27.
- Ponjavic, J., B. Lenhard, C. Kai, J. Kawai, P. Carninci, Y. Hayashizaki, and A. Sandelin. 2006. Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biol* **7**: R78.

- Powell, J. 2000. SAGE. The serial analysis of gene expression. *Methods Mol Biol* **99**: 297-319.
- Prakash, A. and M. Tompa. 2005. Discovery of regulatory elements in vertebrates through comparative genomics. *Nat Biotechnol* **23**: 1249-56.
- Ptashne, M. and A. Gann. 1997. Transcriptional activation by recruitment. *Nature* **386**: 569-77.

Quandt, K., K. Frech, H. Karas, E. Wingender, and T. Werner. 1995. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res* **23**: 4878-84.

Rao, C.V., Y.M. Yang, M.V. Swamy, T. Liu, Y. Fang, R. Mahmood, M. Jhanwar-Uniyal, and W. Dai. 2005. Colonic tumorigenesis in BubR1^{+/-}-ApcMin⁺ compound mutant mice is linked to premature separation of sister chromatids and enhanced genomic instability. *Proc Natl Acad Sci U S A* **102**: 4365-70.

Reddy, S.K., M. Rape, W.A. Margansky, and M.W. Kirschner. 2007. Ubiquitination by the anaphase-promoting complex drives spindle checkpoint inactivation. *Nature* **446**: 921-5.

Ren, B., H. Cam, Y. Takahashi, T. Volkert, J. Terragni, R.A. Young, and B.D. Dynlacht. 2002. E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev* **16**: 245-56.

Richard, P., X. Darzacq, E. Bertrand, B.E. Judy, C. Verheggen, and T. Kiss. 2003. A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *Embo J* **22**: 4283-93.

Rinn, J.L., G. Euskirchen, P. Bertone, R. Martone, N.M. Luscombe, S. Hartman, P.M. Harrison, F.K. Nelson, P. Miller, M. Gerstein et al. 2003. The transcriptional activity of human Chromosome 22. *Genes Dev* **17**: 529-40.

Robertson, G., M. Hirst, M. Bainbridge, M. Bilenky, Y. Zhao, T. Zeng, G. Euskirchen, B. Bernier, R. Varhol, A. Delaney et al. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods*.

Roebuck, K.A., D.P. Szeto, K.P. Green, Q.N. Fan, and W.E. Stumph. 1990. Octamer and SPH motifs in the U1 enhancer cooperate to activate U1 RNA gene expression. *Mol Cell Biol* **10**: 341-52.

Roeder, R.G. 1996. Nuclear RNA polymerases: role of general initiation factors and cofactors in eukaryotic transcription. *Methods Enzymol* **273**: 165-71.

Sandelin, A., P. Carninci, B. Lenhard, J. Ponjavic, Y. Hayashizaki, and D.A. Hume. 2007. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet* **8**: 424-36.

Saur, D., B. Seidler, H. Paehge, V. Schusdziarra, and H.D. Allescher. 2002. Complex regulation of human neuronal nitric-oxide synthase exon 1c gene transcription. Essential role of Sp and ZNF family members of transcription factors. *J Biol Chem* **277**: 25798-814.

- Saxonov, S., P. Berg, and D.L. Brutlag. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* **103**: 1412-7.
- Schaub, M., A. Krol, and P. Carbon. 1999a. Flexible zinc finger requirement for binding of the transcriptional activator Staf to U6 small nuclear RNA and tRNA(Sec) promoters. *J Biol Chem* **274**: 24241-9.
- Schaub, M., A. Krol, and P. Carbon. 2000. Structural organization of Staf-DNA complexes. *Nucleic Acids Res* **28**: 2114-21.
- Schaub, M., E. Myslinski, A. Krol, and P. Carbon. 1999b. Maximization of selenocysteine tRNA and U6 small nuclear RNA transcriptional activation achieved by flexible utilization of a Staf zinc finger. *J Biol Chem* **274**: 25042-50.
- Schaub, M., E. Myslinski, C. Schuster, A. Krol, and P. Carbon. 1997. Staf, a promiscuous activator for enhanced transcription by RNA polymerases II and III. *Embo J* **16**: 173-81.
- Schjerven, H., P. Brandtzaeg, and F.E. Johansen. 2001. A novel NF-kappa B/Rel site in intron 1 cooperates with proximal promoter elements to mediate TNF-alpha-induced transcription of the human polymeric Ig receptor. *J Immunol* **167**: 6412-20.
- Schmitt, S. and R. Paro. 2004. Gene regulation: a reason for reading nonsense. *Nature* **429**: 510-1.
- Schramm, L. and N. Hernandez. 2002. Recruitment of RNA polymerase III to its target promoters. *Genes Dev* **16**: 2593-620.
- Schug, J., W.P. Schuller, C. Kappen, J.M. Salbaum, M. Bucan, and C.J. Stoeckert, Jr. 2005. Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol* **6**: R33.
- Schuster, C., A. Krol, and P. Carbon. 1998. Two distinct domains in Staf to selectively activate small nuclear RNA-type and mRNA promoters. *Mol Cell Biol* **18**: 2650-8.
- Schuster, C., E. Myslinski, A. Krol, and P. Carbon. 1995. Staf, a novel zinc finger protein that activates the RNA polymerase III promoter of the selenocysteine tRNA gene. *Embo J* **14**: 3777-87.
- Segal, E., Y. Fondufe-Mittendorf, L. Chen, A. Thastrom, Y. Field, I.K. Moore, J.P. Wang, and J. Widom. 2006. A genomic code for nucleosome positioning. *Nature* **442**: 772-8.
- Seike, M., A. Gemma, Y. Hosoya, Y. Hosomi, T. Okano, F. Kurimoto, K. Uematsu, K. Takenaka, A. Yoshimura, M. Shibuya et al. 2002. The promoter region of the human BUBR1 gene and its expression analysis in lung cancer. *Lung Cancer* **38**: 229-34.
- Sekinger, E.A., Z. Moqtaderi, and K. Struhl. 2005. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol Cell* **18**: 735-48.
- Shimada, T., H. Fujii, and H. Lin. 1989. A 165-base pair sequence between the dihydrofolate reductase gene and the divergently transcribed upstream gene is sufficient for bidirectional transcriptional activity. *J Biol Chem* **264**: 20171-4.
- Shiraki, T., S. Kondo, S. Katayama, K. Waki, T. Kasukawa, H. Kawaji, R. Kodzius, A. Watahiki, M. Nakamura, T. Arakawa et al. 2003. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A* **100**: 15776-81.
- Singer, V.L., C.R. Wobbe, and K. Struhl. 1990. A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. *Genes Dev* **4**: 636-45.
- Skoufias, D.A., P.R. Andreassen, F.B. Lacroix, L. Wilson, and R.L. Margolis. 2001. Mammalian mad2 and bub1/bubR1 recognize distinct spindle-attachment and kinetochore-tension checkpoints. *Proc Natl Acad Sci U S A* **98**: 4492-7.

- Skuzeski, J.M., E. Lund, J.T. Murphy, T.H. Steinberg, R.R. Burgess, and J.E. Dahlberg. 1984. Synthesis of human U1 RNA. II. Identification of two regions of the promoter essential for transcription initiation at position +1. *J Biol Chem* **259**: 8345-52.
- Smale, S.T. 1994. Dans Conaway, R.C. and Conaway, J.W. (Ed.), Transcription, mechanisms and regulation, Raven Press, New York: 63-81.
- Smale, S.T. and J.T. Kadonaga. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem* **72**: 449-79.
- Smith, C.L. and C.L. Peterson. 2005. ATP-dependent chromatin remodeling. *Curr Top Dev Biol* **65**: 115-48.
- Somma, M.P., C. Pisano, and P. Lavia. 1991. The housekeeping promoter from the mouse CpG island HTF9 contains multiple protein-binding elements that are functionally redundant. *Nucleic Acids Res* **19**: 2817-24.
- Stegmeier, F., M. Rape, V.M. Draviam, G. Nalepa, M.E. Sowa, X.L. Ang, E.R. McDonald, 3rd, M.Z. Li, G.J. Hannon, P.K. Sorger et al. 2007. Anaphase initiation is regulated by antagonistic ubiquitination and deubiquitination activities. *Nature* **446**: 876-81.
- Struhl, K. 2007. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* **14**: 103-5.
- Sudakin, V., G.K. Chan, and T.J. Yen. 2001. Checkpoint inhibition of the APC/C in HeLa cells is mediated by a complex of BUBR1, BUB3, CDC20, and MAD2. *J Cell Biol* **154**: 925-36.
- Suzuki, Y., H. Taira, T. Tsunoda, J. Mizushima-Sugano, J. Sese, H. Hata, T. Ota, T. Isogai, T. Tanaka, S. Morishita et al. 2001. Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites. *EMBO Rep* **2**: 388-93.
- Suzuki, Y., R. Yamashita, K. Nakai, and S. Sugano. 2002. DBTSS: DataBase of human Transcriptional Start Sites and full-length cDNAs. *Nucleic Acids Res* **30**: 328-31.
- Suzuki, Y., K. Yoshitomo-Nakagawa, K. Maruyama, A. Suyama, and S. Sugano. 1997. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. *Gene* **200**: 149-56.
- T**akai, D. and P.A. Jones. 2004. Origins of bidirectional promoters: computational analyses of intergenic distance in the human genome. *Mol Biol Evol* **21**: 463-7.
- Tanaka, M., J.S. Lai, and W. Herr. 1992. Promoter-selective activation domains in Oct-1 and Oct-2 direct differential activation of an snRNA and mRNA promoter. *Cell* **68**: 755-67.
- Tanaka, M., A. Ueda, H. Kanamori, H. Ideguchi, J. Yang, S. Kitajima, and Y. Ishigatsubo. 2002. Cell-cycle-dependent regulation of human aurora A transcription is mediated by periodic repression of E4TF1. *J Biol Chem* **277**: 10719-26.
- Tebb, G. and I.W. Mattaj. 1989. The *Xenopus laevis* U2 gene distal sequence element (enhancer) is composed of four subdomains that can act independently and are partly functionally redundant. *Mol Cell Biol* **9**: 1682-90.
- Testa, A., G. Donati, P. Yan, F. Romani, T.H. Huang, M.A. Vigano, and R. Mantovani. 2005. Chromatin immunoprecipitation (ChIP) on chip experiments uncover a widespread distribution of NF-Y binding CCAAT sites outside of core promoters. *J Biol Chem* **280**: 13606-15.
- Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673-80.

- Tommasi, S. and G.P. Pfeifer. 1995. In vivo structure of the human cdc2 promoter: release of a p130-E2F-4 complex from sequences immediately upstream of the transcription initiation site coincides with induction of cdc2 expression. *Mol Cell Biol* **15**: 6901-13.
- Tommasi, S. and G.P. Pfeifer. 1999. In vivo structure of two divergent promoters at the human PCNA locus. Synthesis of antisense RNA and S phase-dependent binding of E2F complexes in intron 1. *J Biol Chem* **274**: 27829-38.
- Trinklein, N.D., S.F. Aldred, S.J. Hartman, D.I. Schroeder, R.P. Otilar, and R.M. Myers. 2004. An abundance of bidirectional promoters in the human genome. *Genome Res* **14**: 62-6.
- Trinklein, N.D., S.J. Aldred, A.J. Saldanha, and R.M. Myers. 2003. Identification and functional analysis of human transcriptional promoters. *Genome Res* **13**: 308-12.
- Tsai, F.T. and P.B. Sigler. 2000. Structural basis of preinitiation complex assembly on human pol II promoters. *Embo J* **19**: 25-36.
- Tsuritani, K., T. Irie, R. Yamashita, Y. Sakakibara, H. Wakaguri, A. Kanai, J. Mizushima-Sugano, S. Sugano, K. Nakai, and Y. Suzuki. 2007. Distinct class of putative "non-conserved" promoters in humans: comparative studies of alternative promoters of human and mouse genes. *Genome Res* **17**: 1005-14.
- Tyc, K. and J.A. Steitz. 1989. U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *Embo J* **8**: 3113-9.
- Tycowski, K.T., A. Aab, and J.A. Steitz. 2004. Guide RNAs with 5' caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoa. *Curr Biol* **14**: 1985-95.
- Tycowski, K.T., M.D. Shu, and J.A. Steitz. 1996. A mammalian gene with introns instead of exons generating stable RNA products. *Nature* **379**: 464-6.
- U**chiumi, T., D.L. Longo, and D.K. Ferris. 1997. Cell cycle regulation of the human polo-like kinase (PLK) promoter. *J Biol Chem* **272**: 9166-74.
- Uguen, P. and S. Murphy. 2003. The 3' ends of human pre-snRNAs are produced by RNA polymerase II CTD-dependent RNA processing. *Embo J* **22**: 4544-54.
- V**elculescu, V.E., L. Zhang, B. Vogelstein, and K.W. Kinzler. 1995. Serial analysis of gene expression. *Science* **270**: 484-7.
- Virbasius, J.V. and R.C. Scarpulla. 1994. Activation of the human mitochondrial transcription factor A gene by nuclear respiratory factors: a potential regulatory link between nuclear and mitochondrial gene expression in organelle biogenesis. *Proc Natl Acad Sci U S A* **91**: 1309-13.
- W**akasugi, T., H. Izumi, T. Uchiumi, H. Suzuki, T. Arao, K. Nishio, and K. Kohno. 2007. ZNF143 interacts with p73 and is involved in cisplatin resistance through the transcriptional regulation of DNA repair genes. *Oncogene*.
- Wang, Q., W. Li, X.S. Liu, J.S. Carroll, O.A. Janne, E.K. Keeton, A.M. Chinnaiyan, K.J. Pienta, and M. Brown. 2007. A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. *Mol Cell* **27**: 380-92.
- Wang, S.M. 2005. SAGE: Current Technologies and Applications. *Horizon Bioscience, Norwich*.

- Wang, Y. and D.F. Bogenhagen. 2006. Human mitochondrial DNA nucleoids are linked to protein folding machinery and metabolic enzymes at the mitochondrial inner membrane. *J Biol Chem* **281**: 25791-802.
- Wasserman, W.W. and A. Sandelin. 2004. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet* **5**: 276-87.
- Wasylyk, B., R. Derbyshire, A. Guy, D. Molko, A. Roget, R. Teoule, and P. Chambon. 1980. Specific in vitro transcription of conalbumin gene is drastically decreased by single-point mutation in T-A-T-A box homology sequence. *Proc Natl Acad Sci U S A* **77**: 7024-8.
- Watkins, N.J., S.C. Phillips, and P.C. Turner. 1992. The *Xenopus* U7 snRNA-encoding gene has an unusually compact structure. *Gene* **120**: 271-6.
- Wei, C.L., Q. Wu, V.B. Vega, K.P. Chiu, P. Ng, T. Zhang, A. Shahab, H.C. Yong, Y. Fu, Z. Weng et al. 2006. A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**: 207-19.
- Weinmann, A.S., S.M. Bartley, T. Zhang, M.Q. Zhang, and P.J. Farnham. 2001. Use of chromatin immunoprecipitation to clone novel E2F target promoters. *Mol Cell Biol* **21**: 6820-32.
- Weinmann, A.S., P.S. Yan, M.J. Oberley, T.H. Huang, and P.J. Farnham. 2002. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev* **16**: 235-44.
- Whitfield, M.L., G. Sherlock, A.J. Saldanha, J.I. Murray, C.A. Ball, K.E. Alexander, J.C. Matese, C.M. Perou, M.M. Hurt, P.O. Brown et al. 2002. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* **13**: 1977-2000.
- Wingender, E., X. Chen, R. Hehl, H. Karas, I. Liebich, V. Matys, T. Meinhardt, M. Pruss, I. Reuter, and F. Schacherer. 2000. TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res* **28**: 316-9.
- Workman, J.L. and R.E. Kingston. 1998. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu Rev Biochem* **67**: 545-79.
- X**ie, X., J. Lu, E.J. Kulbokas, T.R. Golub, V. Mootha, K. Lindblad-Toh, E.S. Lander, and M. Kellis. 2005. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**: 338-45.
- Xie, X., T.S. Mikkelsen, A. Gnirke, K. Lindblad-Toh, M. Kellis, and E.S. Lander. 2007. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proc Natl Acad Sci U S A* **104**: 7145-50.
- Xu, L.C., M. Thali, and W. Schaffner. 1991. Upstream box/TATA box order is the major determinant of the direction of transcription. *Nucleic Acids Res* **19**: 6699-704.
- Y**amashita, R., Y. Suzuki, H. Wakaguri, K. Tsuritani, K. Nakai, and S. Sugano. 2006. DBTSS: DataBase of Human Transcription Start Sites, progress report 2006. *Nucleic Acids Res* **34**: D86-9.
- Yelin, R., D. Dahary, R. Sorek, E.Y. Levanon, O. Goldstein, A. Shoshan, A. Diber, S. Biton, Y. Tamir, R. Khosravi et al. 2003. Widespread occurrence of antisense transcription in the human genome. *Nat Biotechnol* **21**: 379-86.

- Yuan, G.C., Y.J. Liu, M.F. Dion, M.D. Slack, L.F. Wu, S.J. Altschuler, and O.J. Rando. 2005. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**: 626-30.
- Zhang, Y., X.S. Liu, Q.R. Liu, and L. Wei. 2006. Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. *Nucleic Acids Res* **34**: 3465-75.
- Z**heng, G. and Y.C. Yang. 2004. ZNF76, a novel transcriptional repressor targeting TATA-binding protein, is modulated by sumoylation. *J Biol Chem* **279**: 42410-21.
- Zheng, G. and Y.C. Yang. 2006. Acetylation and alternative splicing regulate ZNF76-mediated transcription. *Biochem Biophys Res Commun* **339**: 1069-75.
- Zwicker, J., C. Gross, F.C. Lucibello, M. Truss, F. Ehlert, K. Engeland, and R. Muller. 1995a. Cell cycle regulation of *cdc25C* transcription is mediated by the periodic repression of the glutamine-rich activators NF-Y and Sp1. *Nucleic Acids Res* **23**: 3822-30.
- Zwicker, J., F.C. Lucibello, L.A. Wolfrain, C. Gross, M. Truss, K. Engeland, and R. Muller. 1995b. Cell cycle regulation of the cyclin A, *cdc25C* and *cdc2* genes is based on a common mechanism of transcriptional repression. *Embo J* **14**: 4514-22.

