

50376
1993
336

50376
1993
336

N° D'ORDRE 1230

UNIVERSITE DES SCIENCES ET TECHNOLOGIES
DE LILLE FLANDRES-ARTOIS

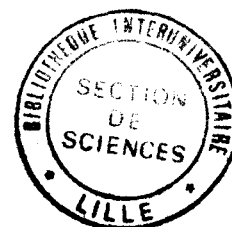
THESE

Présentée pour l'obtention du grade de :

DOCTEUR DE L'UNIVERSITE
DES SCIENCES ET TECHNOLOGIES DE LILLE
EN
SCIENCES DE LA VIE ET DE LA SANTE

PAR

Claude COCQUERELLE



FORMATION DE TRANSCRITS CIRCULAIRES
PAR EPISSAGE
POUR LE PROTO-ONCOGENE ETS-1

Soutenue le 13 Décembre 1993 devant la commission d'examen :

PRESIDENT : Professeur J. P. BOHIN
RAPORTEURS : Professeur M. FELLOUS
Docteur J. MARIE
EXAMINATEURS : Docteur B. BAILLEUL
Docteur M. H. LOUCHEUX-LEFEBVRE

Je tiens à remercier Monsieur J. P. Bohin qui me fait l'honneur de présider le jury de cette thèse.

Monsieur M. Fellous a bien voulu juger ce travail et en être le rapporteur. Je le remercie vivement d'avoir accepté cette tâche délicate.

Je remercie tout particulièrement Madame J. Marie d'avoir accepté d'être le rapporteur de ce travail, et de m'avoir accueilli chaleureusement dans son laboratoire pour effectuer une partie des travaux.

Mes remerciements s'adressent à Madame M. H. Loucheux-Lefebvre, qui siège à ce jury et nous a toujours encouragé dans la poursuite de ce sujet.

Je tiens à remercier Monsieur B. Bailleul de m'avoir accueilli dans son équipe, et de m'avoir dirigé vers et dans ce travail.

Mes vifs remerciements s'adressent à Mme D. Hétuin et Mlle M. Cavrois pour leur constant soutien moral et technique.

Je remercie Mme M. C. Duvieuxbourg pour le travail de photographie de qualité qu'elle a effectué avec patience et gentillesse depuis le début de cette thèse.

Je tiens à remercier tout le personnel de l'U.124 INSERM, qui m'a dispensé une aide technique tout au long de ce travail.

Mes remerciements s'adressent tout particulièrement au Docteur M. Collyn-d'Hooghe, auprès de laquelle il est toujours certain de trouver un réconfort culinaire et moral, en sus d'un savoir inestimable sur la culture de cellules.

Special thanks to K. MacLeod and J. B. Telliez.

Un Grand Merci au Docteur Luc Buée, et à toute l'équipe du laboratoire du Docteur A. Delacourte, pour les discussions scientifiques fort enrichissantes que nous avons eues durant ces dernières années.

Ce mémoire est pour moi l'occasion de remercier de tout mon coeur deux personnes qui ont su, de par leur amour et l'éducation qui m'a été donnée, me conduire jusqu'à cette étape : mes Parents.

Un grand merci à Martine qui supporte ma distraction depuis de nombreuses années, et me fait honneur en connaissant des termes aussi barbares que: ADN, ARN et ETS-1.

Le travail présenté dans ce mémoire a été effectué à l'Institut de Recherches sur le Cancer de Lille (Directeur: Professeur J. Vanlerenberghe), dans le laboratoire d' "Oncohématologie moléculaire" U.124 INSERM, dirigé par le Dr. M. H. Loucheux-Lefebvre, et a bénéficié d'un co-financement entre l'IRCL et la Région Nord-Pas-de-Calais.

TABLE DES MATIERES.

	pages
INTRODUCTION.	08
GENERALITES	10
<u>A. Le mécanisme d'épissage.</u>	11
I. Introduction.	11
II. Les introns et les mécanismes d'épissage.	12
III. L'épissage des ARN pré-messagers nucléaires.	15
<u>III.1 La réaction d'épissage.</u>	15
<u>III.2 Les sites d'épissage.</u>	16
<u>III.3 Le spliceosome.</u>	18
<i>Les snRNP.</i>	19
<i>Les facteurs non-snRNP.</i>	20
<i>Les motifs protéiques dans les facteurs d'épissage.</i>	21
<i>La voie d'assemblage du spliceosome.</i>	23
<u>III.4 La précision et l'ordre dans le mécanisme d'épissage.</u>	26
<i>Interactions entre snRNP, facteurs non-snRNP et sites d'épissage.</i>	26
<i>Rôle du contexte moléculaire.</i>	29
<i>Les modèles.</i>	31
<i>L'épissage alternatif.</i>	32
<i>L'épissage en trans.</i>	33
<i>L'épissage et l'organisation nucléaire.</i>	35
<u>B. De nouveaux rôles biologiques pour les ARN messagers.</u>	36
I. L'inactivation du chromosome X et le gène Xist.	37
II. Le gène H19.	38
III. Les séquences 3'UTR.	39
<u>C. Le proto-oncogène c-ets-1.</u>	41
I. Du rétrovirus E26 à la famille des gènes ets.	41
<u>I.1 Le rétrovirus E26.</u>	41
<u>I.2 Le proto-oncogène aviaire c-ets-1.</u>	44
<u>I.3 La famille des gènes ets.</u>	45

II. Le proto-oncogène humain c-ets-1 : ETS-1.	45
<u>II.1 Localisation chromosomique et expression.</u>	45
<u>II.2 Les protéines ets-1.</u>	47
<i>Les domaines fonctionnels des protéines ets.</i>	47
<i>Les protéines ets, facteurs de transcription.</i>	50
OBJECTIFS.	51
RESULTATS.	54
<u>A. La découverte de l'épissage inversé .</u>	55
Article 1: The EMBO Journal vol. 11 n°3. 1095-1098, 1992.	
Splicing with inverted order of exons occurs proximal to large introns.	
Cocquerelle C., Daubersies P., Majérus M., Kerckaert J. P. et Bailleul B.	
Observation citée dans le <i>Monitor de Trends in Genetics</i> , vol. 8 157, 1992.	60
<u>B. L'épissage inversé forme des transcrits circulaires.</u>	61
Article 2: The FASEB Journal vol. 7 n°1 155-160, 1993.	
Mis-splicing yields circular RNA molecules.	
Cocquerelle C., Mascrez B., Héтуin D. et Bailleul B.	
<u>C. La spécificité de formation des transcrits circulaires.</u>	68
Article 3: soumis à Molecular and Cellular Biology	
Circular RNA transcripts formation by splicing.	
Cavrois M., Cocquerelle C., Héтуin D. et Bailleul B.	
DISCUSSION.	94
CONCLUSIONS ET PERSPECTIVES.	112
BIBLIOGRAPHIE.	114

ABREVIATIONS

A	Adénine
C	Cytosine
G	Guanine
T	Thymidine
U	Uracile
ADN (<i>DNA</i>)	Acide désoxyribonucléique
ADNc (<i>cDNA</i>)	ADN complémentaire
ARN (<i>RNA</i>)	Acide ribonucléique
ARNm (<i>mRNA</i>)	ARN messenger
ARNr (<i>rRNA</i>)	ARN ribosomal
ARNt (<i>tRNA</i>)	ARN de transfert
ATP	Adénosine triphosphate
dNTP	désoxyribonucléotide triphosphate
hnRNP	<i>heterogeneous RiboNucleoProtein</i>
snRNA	<i>small nuclear RNA</i>
snRNP	<i>small nuclear RiboNucleoProtein</i>
ORF	<i>Open Reading Frame</i>
3'UTR	<i>3' Untranslated Region</i>
PCR	<i>Polymerase Chain Reaction</i>
RT-PCR	<i>Reverse Transcription-PCR</i>
b	base
cpm	coups par minute
ml	millilitre
µl	microlitre
µg	microgramme
pb (<i>bp</i>)	paire de bases
kb	kilo bases
kDa	kilo Dalton
DTT	Dithiothréitol
EDTA	Acide éthylènediaminetétraacétique
PBS	<i>Phosphate buffered saline</i>
SDS	Sodium dodécyl sulfate
TBE	Tris borate EDTA

LISTE DES FIGURES ET TABLEAU

	pages
Tableau 1: Molécules d'ARN et fonctions biologiques	37
Figure 1: Structure des introns de Groupes I et II.	13
Figure 2: Les quatre catégories d'intron et leurs mécanismes d'épissage respectifs.	14
Figure 3: Schéma général de la réaction d'épissage pour les ARN pré-messagers.	17
Figure 4: La voie d'assemblage des spliceosomes.	25
Figure 5: Appariements entre l'ARN pré-messager et les ARN U1, U2 et U5.	27
Figure 6: Modes d'épissage alternatifs.	34
Figure 7: Organisation du locus c-ets-1 aviaire et du génôme du virus E26.	43
Figure 8: Organisation du locus c-ets-1 humain et transcrits alternatifs exprimés.	46
Figure 9: Locus du gène SRY et modèle expliquant la circularisation de son ARN.	97
Figure 10: Epissage inversé pour les introns de groupe I et II.	100

INTRODUCTION

La découverte de l'ARN messenger, qui valut en 1965 le prix Nobel de Médecine à Jacques Monod, François Jacob et André Lwoff, parachevait l'énoncé du dogme central de la Biologie moléculaire, ADN-ARN-Protéine, proposé quelques années auparavant: l'ARN messenger est la molécule indispensable pour le transfert d'information du gène à la protéine. L'ARN pré-messenger transcrit à partir de l'ADN subit une série d'événements regroupés sous le terme de maturation, avant d'être transporté dans le cytoplasme puis traduit en protéine. Cette maturation du transcrit primaire consiste en l'addition d'une coiffe en son extrémité 5' et d'une queue polyadénylée en son extrémité 3'. La mise en évidence de la structure morcelée du gène dans les années 70 révéla la nécessité d'un mécanisme supplémentaire prenant part à la maturation de ces ARN pré-messagers: le mécanisme d'épissage. Ce mécanisme permet l'excision des régions non codantes ou introns de l'ARN pré-messenger. La lecture du message est ainsi rendue continue pour le processus de traduction. De plus, l'épissage alternatif de certains exons produit différents transcrits à partir d'un même gène. Ces différents transcrits peuvent même être à l'origine de protéines pourvues de fonctions antagonistes. Il apparaît donc que l'épissage est d'une grande importance dans la régulation de l'expression des gènes.

Bien que leur rôle de transfert d'information prévaut toujours au sein du dogme central, il a été montré récemment que certains ARN messagers ne sont pas traduits en protéine et pourraient avoir des fonctions intrinsèques. La mise en évidence d'une structure circulaire issue d'un mécanisme particulier d'épissage pour certains transcrits du proto-oncogène ETS-1 pourrait s'inscrire dans ce cadre.

Avant de décrire et de discuter les résultats montrant l'existence de ce nouveau type de transcrits, il nous semble important d'exposer les connaissances actuelles concernant l'épissage des ARN pré-messagers, les cas d'ARN messagers non traduits en protéines et le proto-oncogène c-ets-1.

GENERALITES

A. Le mécanisme d'épissage.

I. Introduction.

Philip Sharp décrivait dès 1974 la présence de séquences dans le gène d'une protéine d'enveloppe de l'Adénovirus qui n'étaient pas retrouvées au niveau de l'ARN messager (Sharp *et al.*, 1974). Mais c'est en 1977 que la structure morcelée du gène et le mécanisme d'épissage sont apparus comme une véritable révolution scientifique (Witkowski, 1988). Au cours de cette année, différents groupes de recherche ont mis en évidence le fait que la séquence codante de la plupart des gènes au niveau de l'ADN n'est pas continue mais interrompue par des séquences non codantes. Walter Gilbert proposa d'appeler introns (pour «intragenic regions») ces séquences excisées au cours du mécanisme d'épissage et exons les séquences codantes (pour «expressed regions») (Gilbert, 1978). En 1978, Breathnach et coll. dans le laboratoire de Pierre Chambon remarquaient que les frontières exon-intron dans le gène de l'ovalbumine possédaient de courtes séquences conservées (Breathnach *et al.*, 1978). Au niveau des ARN pré-messagers nucléaires, le consensus de ces séquences nommées sites donneur d'épissage (en 5' de l'intron) et accepteur d'épissage (en 3' de l'intron) a pu être défini d'après l'étude d'un grand nombre de gènes chez les organismes eucaryotes (Mount, 1982; Shapiro et Senepathy, 1987; Jackson, 1991). La découverte de la structure morcelée du gène au cours de l'année 1977 a été signalée dans différents systèmes: des gènes de virus tels que l'Adénovirus (Berget *et al.*, 1977; Chow *et al.*, 1977) et SV40 (Aloni *et al.*, 1977), des gènes transcrits en ARN messagers chez les eucaryotes supérieurs tels que les gènes de globine (Jeffreys et Flavell, 1977), de l'ovalbumine (Breathnach *et al.*, 1977), mais aussi des gènes codant pour des ARN de transfert chez la levure (Goodman *et al.*, 1977) et des ARN ribosomiaux chez la Drosophile (Glover et Hogness, 1977; Wellauer et Dawid, 1977; White et Hogness, 1977). Depuis, bon nombre de régions introniques ont été identifiées dans le génome des eucaryotes supérieurs et inférieurs, mais aussi dans les génomes d'organites cellulaires tels que les mitochondries et les chloroplastes et également dans les génomes de certains bactériophages et bactéries (pour revue Perlman *et al.*, 1990).

II. Les introns et les mécanismes d'épissage.

C'est en 1980 que l'existence de régions introniques dans l'ADN mitochondrial de levure a été mise en évidence (pour revue Perlman *et al.*, 1990). Ces introns ont été considérés comme atypiques, puisqu'ils ne montrent pas à leurs extrémités les séquences d'épissage comme dans les introns d'ARN pré-messagers nucléaires. Michel et coll. en 1982 ont révélé l'existence d'un autre type d'intron dans le génome mitochondrial (Michel *et al.*, 1982). Des études ont montré la présence de structures secondaires conservées dans ces introns nommés respectivement introns de groupes I et II (Michel *et al.*, 1982; Davies *et al.*, 1982; figure 1). La même année, l'équipe de Thomas Cech démontra que l'intron de groupe I d'un gène d'ARNr de *Tetrahymena thermophyla* était capable de s'auto-exciser (Kruger *et al.*, 1982). Cette caractéristique a été démontrée pour tous les introns de groupe I et étendue aux introns de groupe II (Perlman *et al.*, 1990). Les découvertes de l'activité d'auto-épissage de ces introns et de l'activité catalytique de la sous-unité ARN de la RNase P (Guerrier-Takada *et al.*, 1983) ont ouvert un nouveau domaine d'études, celui des ribozymes ou molécules d'ARN à activité catalytique. Quatre catégories d'introns et de mécanismes d'épissage sont définies selon la présence de séquences nucléotidiques conservées et/ou de structures secondaires. Ces quatre catégories et leurs mécanismes d'épissage respectifs sont schématisés sur la figure 2: les introns de groupe I et II pourvus d'une activité d'auto-épissage, les introns des ARN pré-messagers nucléaires excisés au sein d'un complexe ribonucléoprotéique, le spliceosome, et les introns des ARN de transfert excisés par processus enzymatique (Abelson, 1992).

Plusieurs similitudes apparaissent entre les mécanismes d'épissage des introns des groupes I et II et le mécanisme d'épissage des ARN pré-messagers nucléaires (figure 2). D'une part, ces introns sont excisés à la suite de deux étapes de trans-estérification (Weiner, 1993), et d'autre part il existe des molécules réactionnelles intermédiaires communes - la boucle de lariat - pour l'épissage des introns de groupe II et l'épissage des introns des ARN pré-messagers nucléaires. Par ailleurs, différents arguments sont en faveur d'un rôle catalytique pour certains petits ARN nucléaires constitutifs du spliceosome (Guthrie, 1991). A l'heure actuelle, les hypothèses de co-

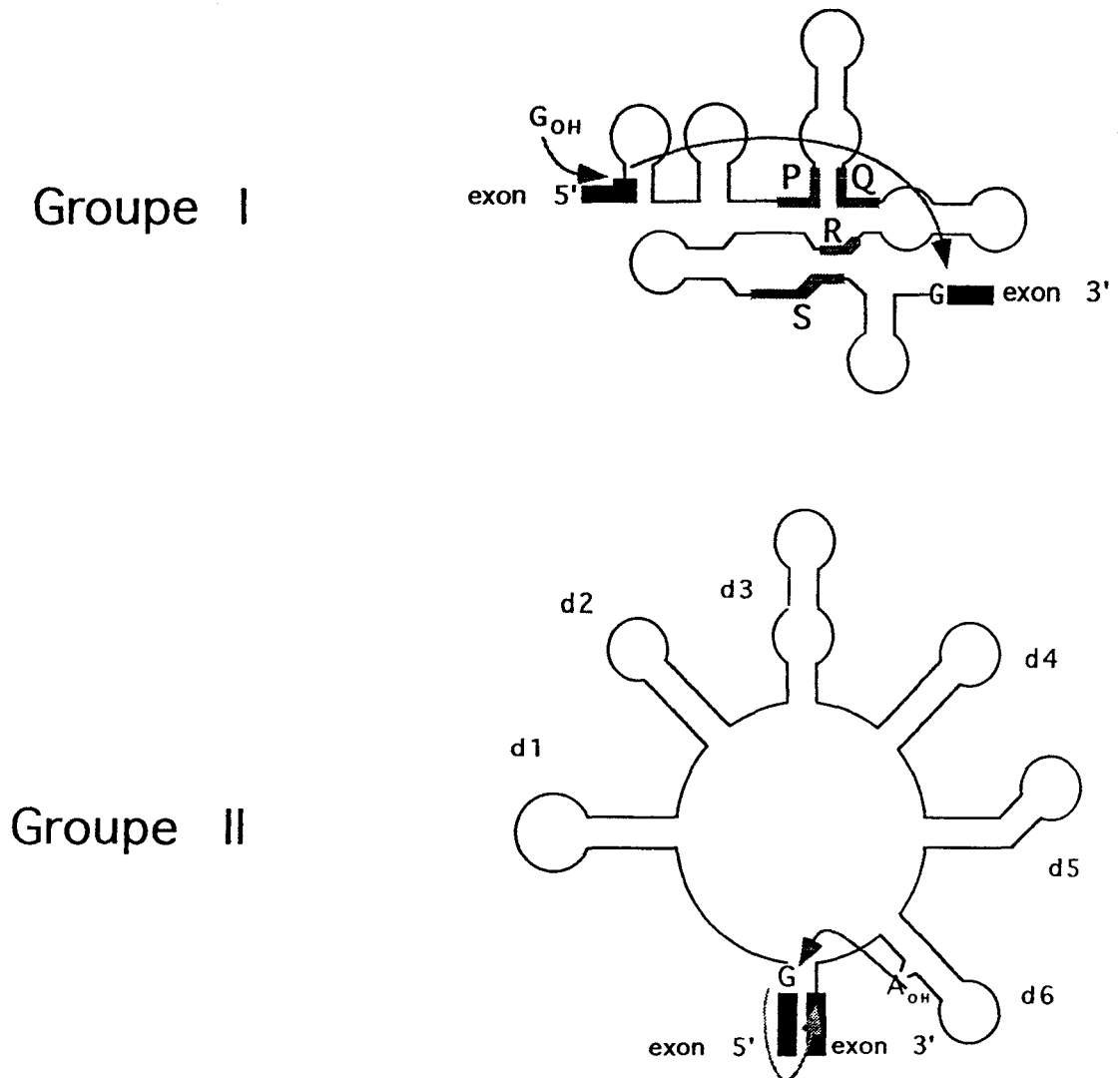
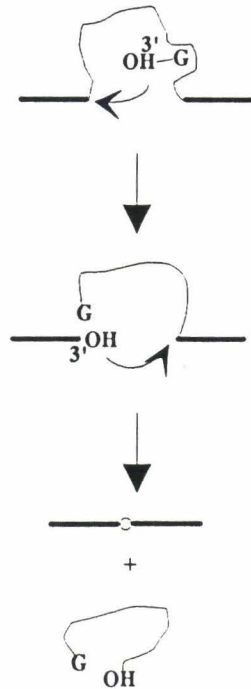


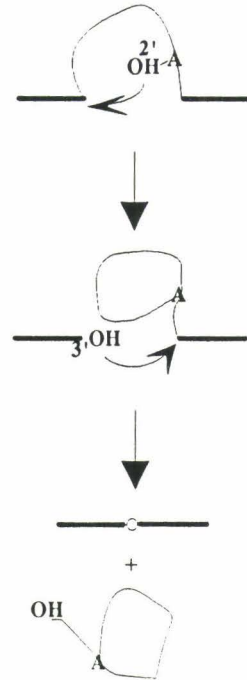
Figure 1: Structure des introns de groupes I et II.

La structure secondaire des introns de groupe I est schématisée selon le modèle proposé par Burke et al. (1987). Les séquences conservées sont nommées P, Q, R, S suivant la nomenclature de Davies et al. (1982). La structure secondaire centrale des introns de groupe II est schématisée selon le modèle proposé par Michel et Dujon (1983). Les domaines divisant la structure secondaire en sections sont nommés d1 à d6. Pour les deux types d'introns les exons sont figurés par des boîtes noires et le groupement OH qui initie les deux étapes de transestérification est indiqué.

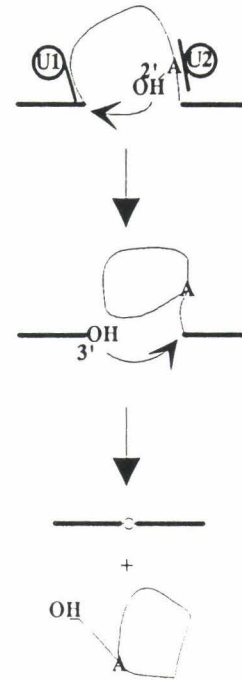
Autoépissage des introns de groupe I



Autoépissage des introns de groupe II



Épissage des ARNs pré-messagers nucléaires (Spliceosome)



Épissage enzymatique des ARNt nucléaires

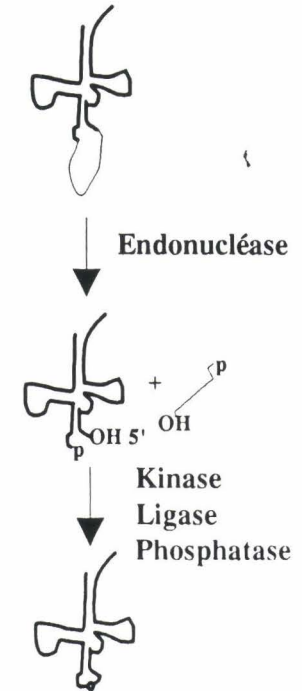


Figure 2. Les quatre catégories d'introns et leurs mécanismes d'épissage respectifs.

Les lignes fines représentent les introns et les lignes épaisses représentent les séquences exoniques adjacentes à ces derniers. Les groupements OH (2' ou 3') impliqués dans les réactions de transestérification sont indiqués. Pour l'épissage des ARNs pré-messagers nucléaires, deux snRNPs (U1 et U2) compris dans le spliceosome et s'appariant à l'ARN pré-messager ont été figurés. Les différentes enzymes intervenant lors de l'épissage des ARNt sont indiquées.

émergence ou de filiation de ces mécanismes au cours de l'évolution n'ont pas été confirmées (Weiner, 1993).

III. L'épissage des ARN pré-messagers nucléaires.

Nous allons voir dans ce chapitre que différents éléments en cis et des facteurs en trans participent à la réaction d'épissage. Au niveau de l'ARN pré-messager, des séquences importantes permettent de définir les exons et les introns. Ces séquences sont reconnues par un complexe, le spliceosome, constitué d'une variété de protéines et de petits éléments nucléaires de nature ribonucléoprotéique.

La structure morcelée des gènes implique que le mécanisme d'épissage soit capable d'assurer l'excision des introns sans modifier le sens de l'information portée par les séquences codantes. D'une part, la réaction d'épissage doit être précise à la base près dans l'excision des séquences introniques et être capable de sélectionner les sites d'épissage adéquats parmi la multitude de séquences d'épissage potentielles. D'autre part, la jonction des exons doit s'effectuer suivant l'ordre existant au niveau du génome. Pour cela, nous décrirons différents aspects de l'épissage qui montrent toute la diversité et la complexité du mécanisme d'épissage engendrées par la cellule eucaryote.

III.1 La réaction d'épissage.

L'épissage est réalisé par deux réactions de trans-estérification menant à l'excision de la séquence intronique et à la jonction des deux séquences exoniques délimitant l'intron (pour revue voir Green, 1991; Moore et al., 1993). La première réaction de trans-estérification implique le transfert du phosphate 5' du premier nucléotide intronique (une Guanosine) vers le groupement 2"OH du ribose du résidu Adénine caractérisant le point de branchement. Les produits formés au cours de cette première étape sont un exon 5' libre, pourvu d'une extrémité 3' OH, et l'intron toujours lié au deuxième exon. Cette dernière molécule en forme de lasso est

appelée boucle de Lariat (figure 3).

Au cours de la seconde réaction de trans-estérification, le phosphate 5' du premier nucléotide du second exon est transféré au groupement 3' OH du premier exon afin de former la séquence exonique épissée (figure 3). A la suite de cette seconde étape, les produits de la réaction d'épissage se séparent, la séquence intronique excisée est dégradée et le spliceosome est désassemblé.

Les études réalisées *in vitro* ont montré que la réaction d'épissage exige, outre l'ARN pré-messager pourvu de séquences consensus fonctionnelles, la présence d'ATP, de cations monovalents et divalents. L'ATP est nécessaire à différentes étapes. Il n'est pas requis pour activer des séquences en vue de la jonction exonique, mais serait plutôt nécessaire à la production d'énergie pour des hélicases à ARN intervenant dans la mise en forme du spliceosome et/ou de l'ARN pré-messager, ou pour tout autre phénomène mécanique affectant la structure du spliceosome durant le processus d'épissage (Perlman *et al.*, 1990). Par ailleurs, des phosphatases spécifiques des résidus sérine et thréonine sont requises *in vitro* pour les deux réactions catalytiques (Mermoud *et al.*, 1992).

III.2 Les sites d'épissage.

Les réactions de trans-estérification du processus d'épissage vont avoir lieu à des sites exacts, définis par des séquences délimitant les frontières exon-intron. Ces séquences sont très conservées chez la levure, mais beaucoup moins chez les mammifères (figure 3). Si le site donneur d'épissage (encore appelé site d'épissage 5', puisqu'il définit l'extrémité 5' de l'intron) a une séquence consensus relativement courte chez tous les eucaryotes, le site d'épissage à l'extrémité 3' de l'intron est défini par trois séquences distinctes: le site de branchement, la séquence polypyrimidique et le site accepteur d'épissage. Ces trois éléments peuvent couvrir une région d'une centaine de nucléotides en amont du site accepteur d'épissage. La taille de la séquence polypyrimidique est plus prononcée chez les mammifères que chez la levure,

Site donneur d'épissage :Site accepteur :

chez les Mammifères

(C/A)AG'GURAGU

(Y)₁₁NCAG'G

chez la Levure

AG'GUAUGU

CAG'G



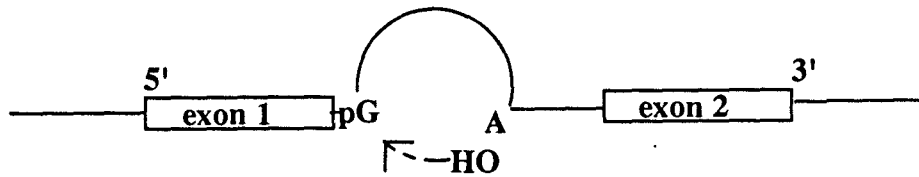
chez les mammifères

YNYUR⊙C

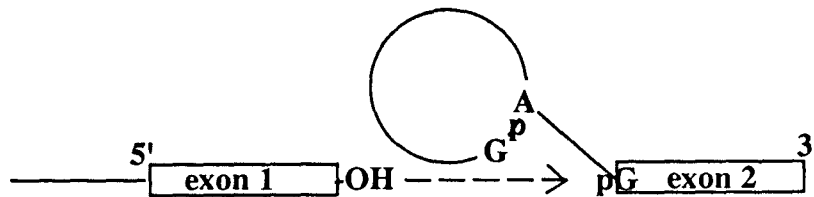
Site de branchement

chez la levure

UACUA⊙C



1ère étape



2ème étape

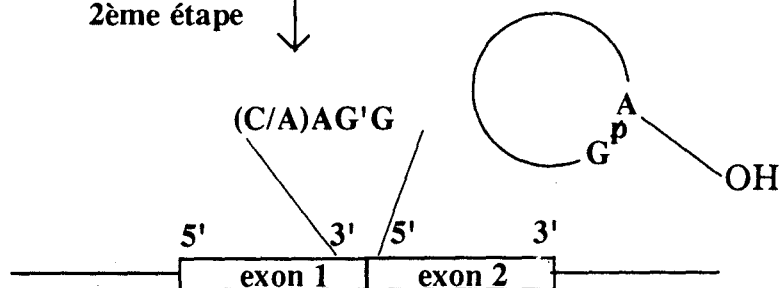


Figure 3 : Schéma général de la réaction d'épissage pour les ARN pré-messagers nucléaires chez les Eucaryotes.

Les séquences consensus d'épissage pour les sites donneur, de branchement et accepteur sont indiquées pour les Mammifères et la Levure (où R= Purine, Y= Pyrimidine, N= n'importe quelle base). Les jonction exon-intron sont marquées par le signe " ' ". Le point de branchement est figuré par le cercle. Au niveau de ce point de branchement, l'adénosine réactive est engagée dans trois liaisons phosphoesters (5', 3', 2'), ce qui donne une structure intronique en "lasso", ou boucle de Lariat.

compensant peut être le fait que la séquence de branchement soit moins conservée (Green, 1991).

Les sites d'épissage sont d'une importance primordiale, puisqu'ils seront reconnus par les petits éléments ribonucléoprotéiques du spliceosome. Comme la reconnaissance de ces séquences implique des appariements de bases nucléotidiques, des mutations au niveau des sites d'épissage peuvent nuire à l'expression correcte du gène de manière qualitative et/ou quantitative. La séquence codante finale peut être modifiée par exemple parce qu'un exon est anormalement excisé. Cela peut intervenir à la suite d'une reconnaissance insuffisante, ou dans le cas de l'utilisation d'un site d'épissage potentiel (site cryptique). D'autres mutations auront pour effet de diminuer l'efficacité de la réaction d'épissage. De nombreuses maladies génétiques, au rang desquelles certaines thalassémies et hémophilies, des cas de phénylcétonurie, certaines formes de mucoviscidose et différentes onco-pathologies, sont imputables à de telles mutations (Krawczak *et al.*, 1992). Par exemple, une mutation G en A de la première base intronique [A C / (g-a) t a a g c] du site donneur d'épissage du vingt-septième exon du gène de susceptibilité au rétinoblastome entraîne l'absence de cet exon dans le transcrit mature, cela dans 100% des cas (Dunn *et al.*, 1989). La prise en compte de cet exon peut être également affectée par une mutation G en A [c c t c a (g-a) / A C] du site accepteur d'épissage (Dunn *et al.*, 1989). Les mutations observées dans ces différentes pathologies impliquent dans 60 % des cas les bases aux positions +1 et +2 (A G / g t a a g t) du site donneur d'épissage, et dans 87% des cas les bases aux positions -2 et -1 (y y n c a g / G) du site accepteur d'épissage (Krawczak *et al.*, 1992). Ces données confirment l'importance des séquences de reconnaissance des frontières exon-intron.

III.3 Le spliceosome.

La découverte du spliceosome est due à l'observation de complexes, formés entre les ARN pré-messagers et des facteurs présents dans les extraits cellulaires de la lignée HeLa ou de levure (Grabowski *et al.*, 1985; Brody et Abelson, 1985). Les expériences de centrifugation en gradient de saccharose montrent que l'ARN pré-messager est associé à un complexe de constante de

sédimentation élevée (40-60S). Les techniques plus récentes comme l'électrophorèse en gel natif (Konarska et Sharp, 1986), la filtration en gel de haute résolution et des méthodes chromatographiques apparentées (Konarska et Sharp, 1986; Perkins *et al.*, 1986; Reed et Maniatis, 1988) ont permis d'identifier les facteurs présents dans le spliceosome et de définir le mode d'assemblage de ce dernier.

Les snRNP.

Les snRNP (*small nuclear RiboNucleoProtein*) sont des complexes contenant un ARN de petite taille (100-200 nucléotides), riche en résidus uracile et associé à des protéines (pour revue voir Zieve et Sauterer, 1990). Des études réalisées avec des extraits de cellules de mammifères ou de levure ont mis en évidence l'existence de cinq snRNP (U1, U2, U4, U5 et U6). Ces snRNP s'associent spécifiquement avec l'ARN pré-messager au cours de la réaction d'épissage (pour revue voir Green, 1991; Moore *et al.*, 1993). Les petits ARN (ou snRNA pour *small nuclear RNA*, U1 à U5) sont transcrits par l'ARN polymérase II, alors que le snRNA U6 est transcrit par l'ARN polymérase III. Ces cinq snRNA sont indispensables à l'excision des introns (Patterson et Guthrie, 1987; Lamm *et al.*, 1991; Lygerou *et al.*, 1991).

Chaque type de snRNP est constituée, en plus de ces petits ARN, de plusieurs protéines. Certaines protéines se trouvent dans chaque snRNP, d'autres sont associées spécifiquement à un seul type de snRNP. Chez les mammifères, huit protéines présentes dans chaque snRNP sont reconnues par un antisérum autoimmun (appelé anti-Sm) produit par des patients souffrant de Lupus érythémateux disséminé (Zieve et Sauterer, 1990). Ces anticorps sont d'un grand intérêt pour l'étude des structures et des fonctions des snRNP.

Les snRNP impliquées dans le mécanisme d'épissage des ARN pré-messagers ne représentent qu'une catégorie de l'ensemble des snRNP identifiées à ce jour. En effet, des complexes ribonucléoprotéiques ont pu être localisés dans différents compartiments cellulaires comme le nucléoplasme, le nucléole, le cytoplasme et la mitochondrie et ils jouent des rôles au niveau de la maturation des ARN, de la traduction ou encore de la réplication de l'ADN (Baserga et Steitz, 1993).

Les facteurs non-snRNP.

D'autres protéines qui ne sont pas associées de manière constitutive aux snRNP sont appelées facteurs non-snRNP. Ces protéines sont impliquées dans les différentes étapes de la réaction d'épissage. Les techniques de complémentation biochimique, de liaison à l'ARN, d'immuno-inhibition ou déprivation et l'utilisation de la génétique chez la levure ont permis d'identifier et de purifier plusieurs de ces facteurs (pour revue Green, 1991; Brown J. D. *et al.*, 1992; Moore *et al.*, 1993). Parmi ces protéines, plusieurs interagissent avec la séquence polypyrimidique et le site d'épissage accepteur. Chez les mammifères, U2AF (*U2 Auxiliary Factor*) a été démontré être un facteur essentiel pour la fixation de la snRNP U2 au site accepteur d'épissage (Ruskin *et al.*, 1988; Fu et Maniatis, 1990; Zamore et Green, 1991; Zamore *et al.*, 1992; Zhang *et al.*, 1992). Un autre facteur appelé *intron-binding protein* (IBP) est associé à la snRNP U5 (Gerke et Steitz, 1986; Tazi *et al.*, 1986). Le facteur *polypyrimidin tract binding protein* (PTB) se lie à la séquence polypyrimidique du site accepteur d'épissage (Garcia-Blanco *et al.*, 1989; Patton *et al.*, 1991). D'autres facteurs, tels que la protéine hnRNP C (*heterogeneous nuclear RiboNucleoProtein C*) et le facteur SC-35 ont été identifiés comme facteurs non-snRNP dans des expériences d'immuno-déprivation (Choi *et al.*, 1986; Fu et Maniatis, 1990).

Chez la levure, l'identification de facteurs non-snRNP a été facilitée par l'isolement de mutants thermosensibles pour l'épissage des ARN pré-messagers (pour revue Brown J.D. *et al.*, 1992). Ces mutants ont été nommés prp pour *pre-RNA processing*. Michael Rosbash et coll. ont mis en évidence l'inhibition du mécanisme d'épissage lorsque ces souches mutantes sont incubées à des températures non permissives et ont proposé que cette inhibition doit être la base du phénotype mutant (Rosbash *et al.*, 1981). Des études biochimiques *in vitro* ont confirmé cette hypothèse (Lustig *et al.*, 1986; Lin *et al.*, 1987). Une vingtaine de ces facteurs PRP ont pu être clonés à ce jour (pour revue Green, 1991; Brown J. D. *et al.*, 1992; Moore *et al.*, 1993). Des facteurs non-snRNP chez les mammifères qui sont apparentés aux facteurs PRP de levure ont pu être identifiés. Il semble que les facteurs PRP2 et PRP8 aient respectivement pour homologues chez les mammifères le facteur SF4, agissant durant la première étape de l'épissage

(Utans et Kramer, 1990) et une protéine de 200 kDa associée au snRNP U5 (Anderson *et al.*, 1989; Bach *et al.*, 1989; Pinto et Steitz, 1989).

De plus, l'ARN pré-messager est dès sa transcription lié à une multitude de protéines, les hnRNP (*heterogeneous nuclear RiboNucleoProtein*), dont la fonction générale doit être de rendre l'ARN pré-messager accessible à la machinerie d'épissage. En effet, les ARN pré-messagers sont généralement de très grande taille, et l'intervention de ces protéines pourrait garantir la fidélité de la réaction d'épissage (Dreyfuss *et al.*, 1993).

Les motifs protéiques dans les facteurs d'épissage.

Les protéines du spliceosome, facteurs snRNP et non-snRNP, sont impliquées dans des interactions ARN-protéine et/ou protéine/protéine. L'analyse de la séquence de ces protéines révèle l'existence de plusieurs motifs qui expliquent partiellement de telles interactions.

Deux motifs protéiques présents dans une grande variété de facteurs d'épissage semble avoir une fonction importante pour leur interaction avec l'ARN. En effet, le facteur PRP24 chez la levure, les protéines non-snRNP U2AF, ASF/SF2 et SC35, plusieurs hnRNP ainsi que plusieurs protéines participant à la structure des snRNP (Green, 1991; Mattaj, 1993) contiennent un domaine nommé RNP (*RiboNucleoProtein*; Dreyfuss *et al.*, 1988), RRM (*RNA Recognition Motif*; Query *et al.*, 1989) ou encore RNP80 (Scherly *et al.*, 1989). Ce domaine comprend deux séquences consensus (un octapeptide Lys/Arg-Gly-Phe/Tyr-Gly/Ala-Phe-Val-X-Phe/tyr (RNP1) et une séquence hexapeptidique (RNP2) moins conservée et riche en résidus aromatiques et aliphatiques) localisées à une trentaine d'acides aminés l'une de l'autre. Plusieurs travaux ont démontré que ce motif était un domaine de liaison à l'ARN (Query *et al.*, 1989; Scherly *et al.*, 1989; Lutz-Freyermuth *et al.*, 1990). Le deuxième motif se caractérise par une région riche en résidus sérine et arginine (nommé région SR; pour revue Green, 1991; Moore *et al.*, 1993). Des régions SR sont observées dans les facteurs d'épissage d'eucaryotes supérieurs tels que la protéine U1 snRNP de 70 kDa, les facteurs non-snRNP: U2AF et ASF. Une région SR est observée dans les protéines tra, tra-2 et su(w⁺) de *Drosophila melanogaster*, protéines qui

régulent un épissage alternatif au cours de la détermination sexuelle (Green, 1991). Le grand nombre de facteurs d'épissage eucaryotes possédant ce motif SR prouve indirectement qu'il doit jouer un rôle important dans le processus d'épissage. Ces facteurs pourraient influencer la liaison à l'ARN pré-messager d'autres protéines et des snRNP. Il est surprenant de constater que ce motif SR n'a pas été identifié dans les facteurs PRP, connus à ce jour, chez la levure. Il se pourrait que la flexibilité du processus d'épissage chez les mammifères rende celui-ci dépendant d'un tel motif (Moore et al., 1993).

Chez la levure, les motifs peptidiques: DEAD (asp-glu-ala-asp) et DEAH (asp-glu-ala-his) apparentent certains facteurs d'épissage à des hélicases à ARN dépendantes de l'ATP (Wassermann et Steitz, 1991; Schmid et Linder, 1992). Plusieurs facteurs d'épissage contenant un de ces motifs ont pu être caractérisés. Cependant, à ce jour aucune activité d'hélicase à ARN n'a été mise en évidence pour ces protéines. Elles interviennent à différents stades de la réaction d'épissage, comme l'assemblage du spliceosome (PRP5), la première étape (PRP2) et la seconde étape (PRP16) du processus d'épissage ou encore le désassemblage du spliceosome (PRP22) (Green, 1991). Une activité d'ATPase dépendante de l'ARN a été démontrée pour la protéine PRP16 (Schwer et Guthrie, 1991).

Il existe également dans les protéines impliquées dans l'épissage, des motifs présents dans certains facteurs de transcription. En effet, un motif protéique apparenté au doigt de zinc des facteurs de transcription a été identifié dans PRP6 et PRP9 et pourrait être le site de fixation pour des ions métalliques. Legrain et Choulika ont montré qu'un de ces motifs riche en résidus cystéine et histidine était essentiel pour l'activité de PRP9 (Legrain et Choulika, 1990). Ces deux facteurs contiennent également des motifs à leucine, ressemblant au motif de fermeture à glissière à Leucines permettant la dimérisation de certains facteurs de transcription, comme les facteurs FOS et JUN (Landschulz, 1988).

A ces motifs s'ajoute un autre domaine protéique très probablement impliqué dans des interactions protéine-protéine. En effet, Dalrymple et coll. ont montré que PRP4 comportait cinq motifs répétés, déjà identifiés dans les sous-unités β de la famille des protéines G (Dalrymple *et al.*, 1989). Il a été proposé que cette région de PRP4 joue un rôle au niveau de l'assemblage des trois snRNP U4/U6/U5 (Bordonne *et al.*, 1990).

La voie d'assemblage du spliceosome.

Outre des interactions ARN-protéine et protéine-protéine, l'assemblage du spliceosome implique des appariements de bases nucléotidiques, d'une part entre snRNA, et d'autre part entre les snRNA et l'ARN pré-messager.

L'assemblage du spliceosome a pu être étudié grâce à la cinétique de formation de complexes observés sur gel natif (pour revue Green, 1991; Moore *et al.*, 1993). Des études reposant sur l'inactivation spécifique de chaque snRNP ont montré que l'association de ces snRNP formant le spliceosome devait être réalisée de manière ordonnée. Ces études ont montré que le premier appariement important, qui ne nécessite pas d'ATP, dans la formation du spliceosome est l'appariement du snRNA U1 avec le site donneur d'épissage (Séraphin *et al.*, 1988). Le complexe comprenant la snRNP U1 et l'ARN pré-messager, appelé «commitment complex» ou complexe C chez la levure, engage l'ARN dans la voie d'épissage (Legrain *et al.*, 1988). Deux complexes C (CC1 et CC2) ont été observés chez la levure (Séraphin et Rosbash, 1989; figure 4). Il a pu être montré que le complexe CC2 nécessite la présence du site de branchement (Séraphin et Rosbash, 1991). La reconnaissance du site accepteur est réalisée par son appariement avec l'extrémité 5' du snRNA U1. Cet appariement est conservé dans le complexe CC2. Dans celui-ci, la reconnaissance du site de branchement est indépendante de la snRNP U2 et doit être médiée par un facteur d'épissage non-snRNP (Ruby et Abelson, 1988; Rosbash et Séraphin 1991). L'étape suivante de l'assemblage est plus rapide que la précédente et fait intervenir la snRNP U2, qui va se lier au complexe C en présence d'ATP pour donner le pré-spliceosome (Liao *et al.*, 1992).

Michaud et Reed ont mis en évidence chez les mammifères un complexe similaire au complexe C, qu'ils ont nommé E, lequel contient les deux snRNP U1 et U2 (Michaud et Reed, 1991).

Enfin, la formation complète du spliceosome est caractérisée par l'entrée des snRNP U4, U5, et U6 (figure 4). Il a été constaté que les snRNP U4 et U6 étaient fortement associées dans les extraits cellulaires (Shannon et Guthrie, 1991). Ce complexe U4/U6 va interagir par ailleurs

avec la snRNP U5 pour permettre l'entrée de ces trois snRNP ensemble dans le spliceosome (Lamm *et al.*, 1991; Séraphin *et al.*, 1991a; Utans *et al.*, 1992).

Les deux réactions de trans-estérification, menant à l'excision de l'intron et à l'appariement des séquences exoniques, auront lieu après que la snRNP U4 subisse un changement de conformation et se dissocie de la snRNP U6 (Blencowe *et al.*, 1989). Les exons fusionnés sont relâchés du spliceosome au cours d'un processus actif. Des mutations dans le gène PRP22 chez la levure bloque ce relâchement (Company *et al.*, 1991). L'intron encore associé aux snRNP sera relâché en présence d'ATP (Ruskin et Green, 1985; Arenas et Hurwitz, 1987; Chapman et Boeke, 1991) et les snRNP seront régénérés pour la formation de nouveaux spliceosomes (figure 4).

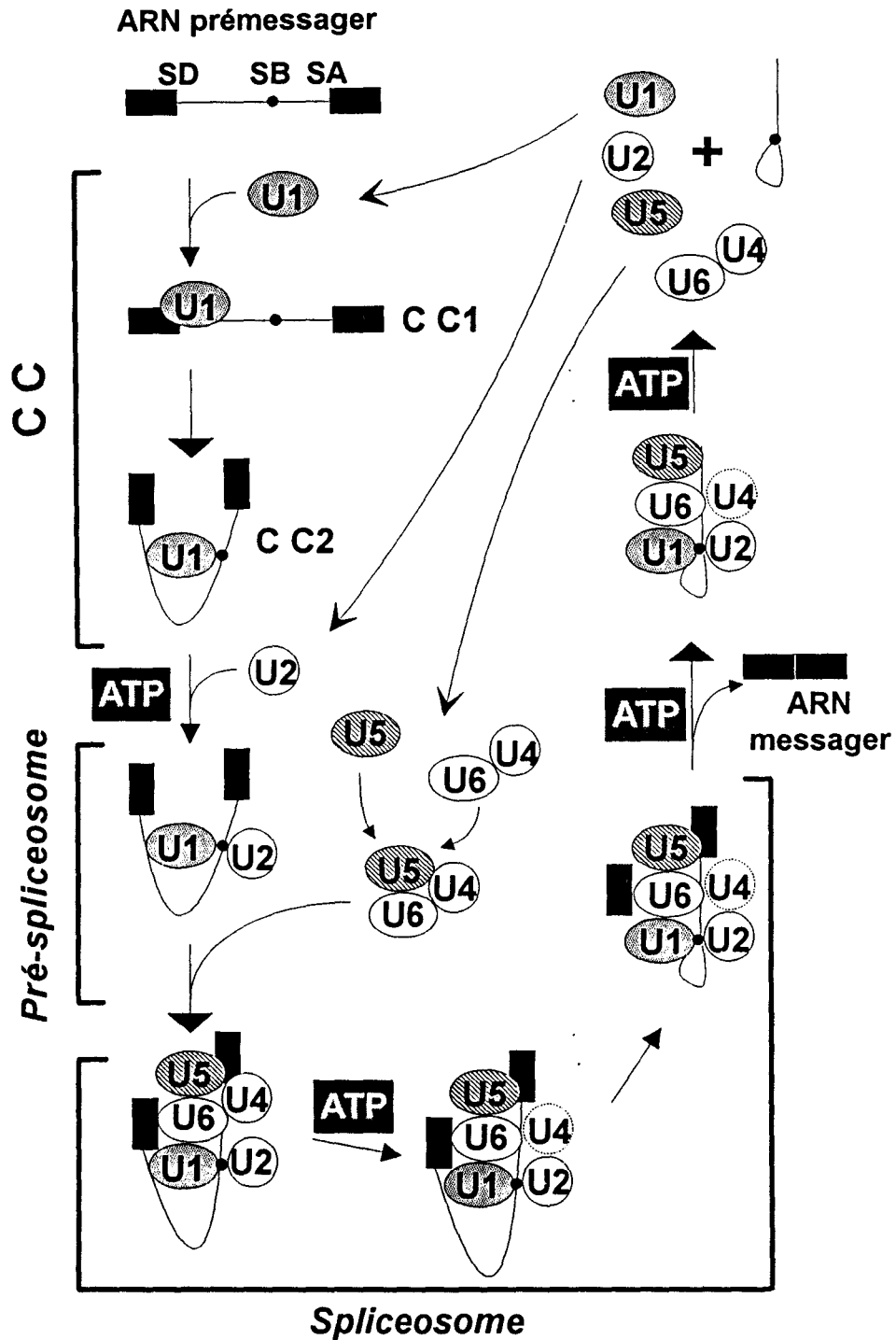


Figure 4. La voie d'assemblage des spliceosomes.

L'assemblage a été schématisé selon le modèle établi chez la levure. Les complexes C (CC), pré-spliceosome et spliceosome sont indiqués. Les snRNPs sont symbolisées par des disques portant leur nom. Les sites donneur (SD) et accepteur (SA) d'épissage, et le site de branchement (SB) sont positionnés sur l'ARN pré-messager. Les étapes nécessitant de l'ATP sont marquées. La snRNP U4 représentée en pointillés doit se dissocier de la snRNP U6 pour que la réaction d'épissage ait lieu (modifié de Lygerou *et al.*, 1993).

III.4 La précision et l'ordre dans le mécanisme d'épissage.

Pour une synthèse d'ARN messagers fonctionnels, le spliceosome doit exciser les régions non codantes de manière extrêmement précise. Il faut que l'excision des régions introniques se fasse à la base près à la suite d'une sélection adéquate des sites d'épissage parmi de nombreux sites potentiels. Nous allons voir que plusieurs éléments peuvent intervenir pour assurer cette précision: le consensus des sites d'épissage reconnu par les snRNP, l'intervention de facteurs protéiques non-snRNP, l'influence des régions exoniques et introniques adjacentes aux sites d'épissage et la chronologie d'épissage dans le cas d'un ARN pré-messager multi-intronique. Nous verrons également les modèles qui ont été proposés en vue d'assurer la précision de la sélection d'épissage et l'appariement des sites d'épissage.

Au delà de la sélection des sites d'épissage se pose également le problème de la conservation de l'ordre génomique des exons. En effet, il doit exister un mécanisme assurant l'appariement correct des sites d'épissage, du site donneur d'épissage d'un exon en 5', vers le site accepteur d'épissage de l'exon en 3'. Un tel mécanisme prend son importance pour expliquer l'excision de grands introns, ou pour assurer la conservation de l'ordre des exons au cours de l'épissage de transcrit comportant de nombreux introns.

Interactions entre snRNP, facteurs non-snRNP et sites d'épissage.

Les séquences d'épissage sont de la plus grande importance pour la précision. En effet, cette précision va être assurée tout au long de la réaction par de nombreuses interactions entre les différents facteurs participant à l'épissage et les sites d'épissage.

Des expériences réalisées dans des lignées de cellules humaines et chez les levures *Saccharomyces cerevisiae* et *Schizosaccharomyces pombe* ont démontré que l'extrémité 5' du snRNA U1 pouvait s'apparier aux extrémités 5' et 3' des introns (Zhuang et Weiner, 1986; Rosbash et Séraphin, 1991; figure 5). La reconnaissance précoce du site d'épissage 3' est due à un appariement entre deux nucléotides du snRNA U1 et le dinucléotide AG du même site (Reich

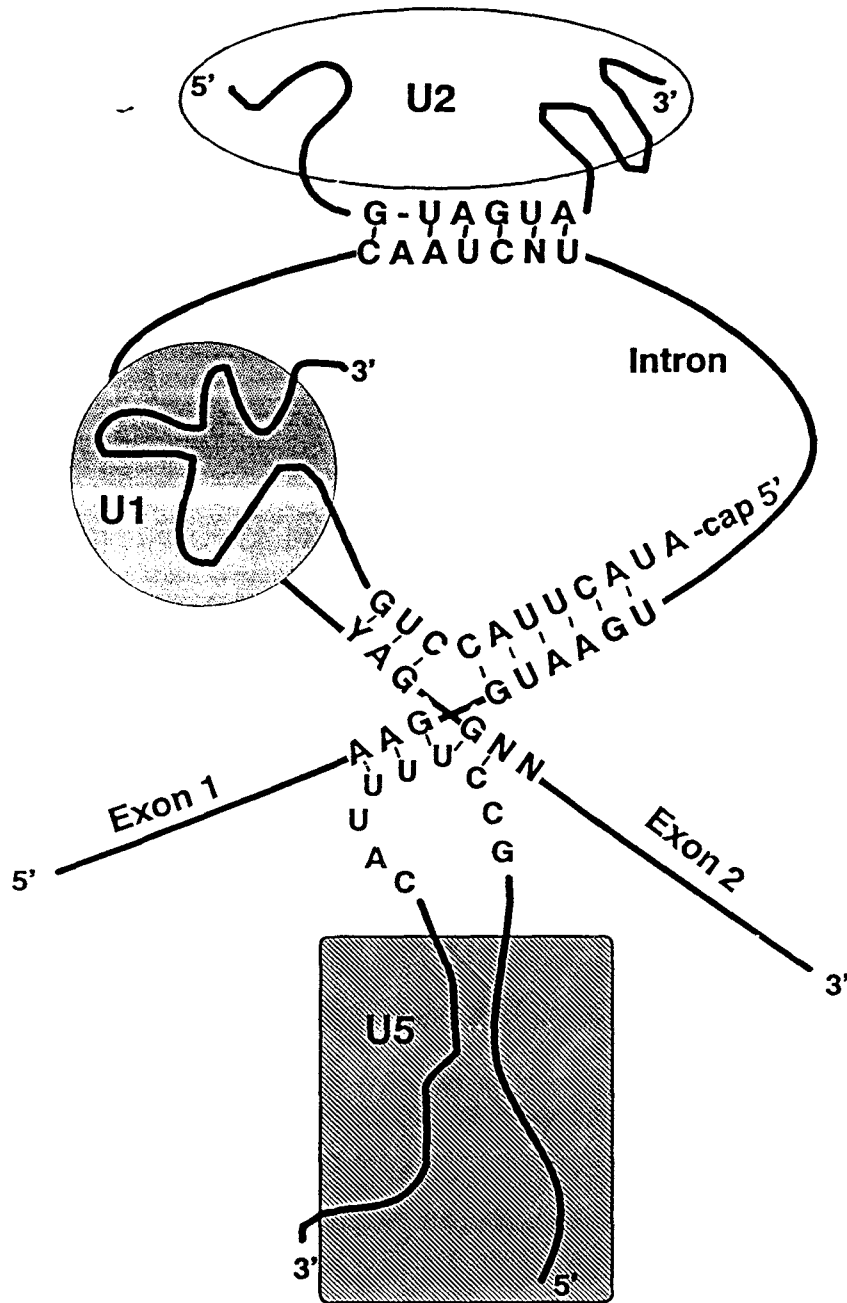


Figure 5. Appariements entre l'ARN pré-messager et les ARNs des snRNPs U1, U2 et U5. Les séquences exoniques et introniques sont les nucléotides consensus pour les eucaryotes supérieurs (Mount, 1982). L'appariement entre l'ARN U1 et le site d'épissage en 3' a été observé chez *Schizosaccharomyces pombe* (Reich *et al.*, 1992). (modifié d'après Lygerou *et al.*, 1993).

et al., 1992). Cependant, la snRNP U1 ne peut être le facteur unique déterminant le clivage au niveau des sites d'épissage. En effet, des mutants chez la levure montrent que le site de coupure à la bordure exon-intron peut être situé à plusieurs nucléotides de la séquence d'appariement entre l'ARN pré-messager et l'ARN U1 (Séraphin et Rosbach, 1990). La snRNP U5 participe également à la sélection des sites 5' et 3' en s'appariant aux séquences exoniques flanquant l'intron (Newman et Norman, 1992). Steitz a proposé que les snRNA U1 et U5 pourraient s'apparier simultanément aux sites d'épissage 5' et 3' en formant une structure à quatre branches similaire à celle d'Holliday pour le mécanisme de Crossing-Over (Steitz, 1992; figure 5). Des appariements non Watson-Crick sont envisagés dans les cas où la complémentarité entre les séquences n'existe pas, mais ils n'ont pas été démontrés.

Un appariement entre l'ARN U2 et le point de branchement de l'ARN pré-messager a été également démontré chez la levure et les eucaryotes supérieurs (Parker *et al.*, 1987; figure 5). Chez la levure, la séquence de branchement très conservée est primordiale pour l'appariement avec le snRNA U2. Chez les mammifères, c'est la séquence polypyrimidique qui est plus importante. Cependant, comme pour les sites d'épissage, la sélection du point de branchement doit certainement impliquer d'autres facteurs (Lygerou *et al.*, 1993). Les facteurs, U2AF, PTB/hnRNP I et hnRNP C ont une affinité similaire pour la séquence polypyrimidique (Choi *et al.*, 1986; Garcia-Blanco *et al.*, 1989; Gil *et al.*, 1991; Patton *et al.*, 1991; Ghetti *et al.*, 1992). Il a été démontré que U2AF est nécessaire pour que la snRNP U2 puisse se lier au point de branchement (Ruskin *et al.*, 1988; Fu et Maniatis, 1990; Zamore et Green, 1991). Il est ainsi possible que les protéines se fixant sur la séquence pyrimidique entrent en compétition ou se lient de façon successive à ce site durant l'assemblage du spliceosome. Bennet et coll. ont montré *in vitro* que les ARN pré-messager lient différemment les facteurs PTB et hnRNP C, avant l'assemblage du spliceosome (Bennett *et al.*, 1992).

Récemment, il a été montré *in vitro* que les protéines hnRNP A1 et ASF/SF2 participent à la sélection des sites d'épissage. En présence de plusieurs sites d'épissage 5', la protéine hnRNP A1 induit l'utilisation du site d'épissage 5', le plus distal, alors que la protéine ASF/SF2 facilite l'utilisation d'un site d'épissage 5' proximal (Krainer *et al.*, 1990; Mayeda et Krainer, 1992).

Il a été démontré que le rapport entre ces deux protéines détermine le choix de sites d'épissage 5' (Mayeda et Krainer, 1992). Cependant, le rôle du rapport de ces deux facteurs reste à déterminer *in vivo*.

Un autre facteur, la protéine IBP serait associé à l'ARN U5 et se lierait spécifiquement au site d'épissage accepteur lors de l'assemblage du spliceosome (Gerke et Steitz, 1986; Tazi *et al.*, 1986). A l'heure actuelle, aucune interaction précise entre l'ARN pré-messager et les snRNP U4 et U6 n'a pu être mise en évidence. Néanmoins, Sawa et coll. ont réalisé des expériences de pontage montrant une interaction entre l'ARN U6 et la région du site donneur d'épissage, suggérant ainsi un possible rôle dans la sélection des sites d'épissage (Sawa et Shimura, 1992).

Dernièrement, Fu a démontré *in vitro* qu'un facteur non-snRNP, la protéine SC35, est suffisante pour sélectionner certains sites d'épissage. En effet, il a montré que l'ARN pré-messager du gène de la β -globine humaine était épissé préférentiellement *in vitro* en présence d'un ARN compétiteur, cela à condition que l'extrait cellulaire nécessaire à l'épissage soit ajouté ultérieurement à SC35 (Fu, 1993). La protéine ASF/SF2 n'a pas d'action sur l'ARN de β -globine, mais sur l'ARN de la protéine tat du virus VIH (Fu, 1993). Ces deux protéines qui possèdent un motif SR, pourraient sélectionner précisément les sites d'épissage, en fonction de leur affinité pour une séquence d'ARN donnée (Manley, 1993). Cependant, bon nombre de questions sont encore sans réponse: quelles séquences d'ARN sont reconnues par les protéines SR et est-ce que différentes protéines reconnaissent différentes séquences? Que font les protéines SR après s'être fixées sur l'ARN (Manley, 1993)?

Rôle du contexte moléculaire.

Les séquences d'épissage seules ne peuvent suffire au mécanisme assurant la précision. En effet, il arrive que des séquences d'épissage non totalement conformes au consensus soient sélectionnées aux dépens de séquences plus proches du consensus (Ohshima et Gotoh, 1987). A ce niveau, il est nécessaire de considérer l'importance de ce qui est appelé «effet de contexte».

Cet effet de contexte repose sur les séquences plus ou moins voisines des sites d'épissage, séquences de nature intronique et/ou exonique. Goguel et coll. ont démontré, par des expériences *in vitro* chez la levure, que des appariements entre deux séquences introniques respectivement proches du site donneur d'épissage et du point de branchement affectaient de façon positive la sélection des sites d'épissage, ainsi que l'efficacité de réaction (Goguel *et al.*, 1993). Watakabe et coll. ont mis en évidence l'existence d'une séquence exonique activatrice d'épissage (ERS pour *exon recognition sequence*). Cette séquence, riche en bases puriques et située en 5' de l'exon M2 du gène murin codant pour la μ (IgM), augmente l'efficacité d'épissage de toute séquence intronique se situant en amont. Watakabe et coll. ont montré par des expériences de compétition *in vitro* que cette séquence fixait des facteurs d'épissage. La snRNP U1 semble être un de ces facteurs (Watakabe *et al.*, 1993). Ces études montrent que les séquences exoniques ou introniques peuvent être hautement structurées et influencer la sélection des sites d'épissage.

Un autre mécanisme qui pourrait contribuer à la sélection des sites d'épissage est l'ordre d'épissage des exons. Un ordre préférentiel d'épissage où le premier intron est l'un des derniers à être excisé a été observé pour plusieurs ARN pré-messagers tels que celui du gène de l'ovomucoïde chez la poule (Tsai *et al.*, 1980), le gène de la thymidine kinase chez la souris et le hamster chinois (Gudas *et al.*, 1990), et dernièrement celui du gène codant pour la dihydrofolate réductase (Kessler *et al.*, 1993). A l'opposé, d'autres gènes ne connaissent pas d'ordre préférentiel pour l'excision de leurs introns (Nelson *et al.*, 1983; Zeitlin et Efstratiadis, 1984; Hatzoglou *et al.*, 1985; Weil *et al.*, 1990). L'ordre d'épissage pourrait favoriser la sélection de sites d'épissage, en créant par exemple des structures secondaires lors de l'excision préférentielle d'un intron. Ceci a été proposé pour expliquer l'épissage alternatif du gène de la préprotachykinine (Nasim *et al.*, 1990).

Les différents éléments que nous venons d'aborder comme le consensus de la séquence d'épissage, l'affinité des facteurs d'épissage pour une séquence spécifique d'ARN, l'effet de contexte, doivent intervenir au niveau de la compétition existant entre les sites d'épissage potentiels dans la séquence de l'ARN pré-messager (Green, 1991). Le résultat de cette compétition aboutira à la sélection des sites d'épissage et à l'excision, à la base près, des introns.

Plusieurs modèles à ce jour expliquent également de manière partielle les mécanismes de sélection des sites d'épissage et le contrôle de l'épissage 5'-3'.

Les modèles.

Le modèle décrit sous le nom de «scanning model» propose que la machinerie d'épissage s'attache sur un site 5' d'épissage de l'ARN pré-messager puis progresse de 5' vers 3' sur la molécule d'ARN, afin de trouver le premier site 3' d'épissage, ou vice versa après l'attachement sur un site 3' d'épissage (Lewin, 1980). Cependant ce modèle n'est pas possible ou est incomplet. En effet, des expériences *in vivo* montrent qu'à l'aide de constructions dans lesquelles les sites d'épissage ont été dupliqués, l'épissage ne s'effectue pas toujours avec les sites 5' et 3' les plus proches l'un de l'autre (Kuhne *et al.*, 1983). De plus, un tel modèle ne permet pas d'expliquer le mécanisme d'épissage en trans entre deux molécules d'ARN distinctes (Aebi et Weissmann, 1987).

Un autre modèle qui prend en compte la synthèse progressive du transcrit a été avancé par Aebi et Weissmann sous le terme: «Premier arrivé, premier servi» (Aebi et Weissmann, 1987). Les auteurs proposent que les complexes du spliceosome s'assemblent sur les sites d'épissage de l'ARN pré-messager, dès l'initiation de la transcription. Un tel mécanisme assurerait qu'aucun exon ne soit oublié au cours de l'épissage. Bien que des auteurs ont montré que la transcription et l'assemblage du spliceosome au niveau d'un même transcrit pourraient être concomitantes (Sass et Pederson, 1984; Fakan *et al.*, 1986; Beyer et Osheim, 1988), ce modèle n'explique cependant ni l'appariement des sites au delà de grandes régions introniques contenant des sites potentiels d'épissage ni les cas d'épissage alternatif.

Le modèle énoncé par Robberson et coll. propose que la détermination des sites accepteur et donneur d'épissage s'effectue au niveau de l'exon par la snRNP U1. Celle-ci n'interagirait plus initialement avec le site d'épissage 5', mais avec le site 3' et chercherait dans une direction 5'-3' un site donneur d'épissage (Robberson *et al.*, 1990). Des expériences ont montré qu'il existe bien des interactions spécifiques entre les sites d'épissage de l'exon. Hoffman et

Grabowski ont identifié une protéine recrutée pour la reconnaissance d'un site 3' faible d'un exon alternatif, dans le cas où le site 5' a un fort potentiel d'épissage (Hoffman et Grabowski, 1992). Cette interaction nécessite la présence de la snRNP U1. Les auteurs ont identifié la protéine se liant au site 3' faible comme étant U2AF. Ils ont également montré que l'interaction entre U2AF et la snRNP U1 nécessitait au moins une autre protéine (Hoffman et Grabowski, 1992). Le modèle appelé «définition de l'exon» présente certains avantages comme celui d'expliquer la limite généralement observée de 300 bases pour les exons internes (Hawkins, 1988) et d'éviter les sites cryptiques dans l'intron (Moore *et al.*, 1993). Il est certain que des mécanismes supplémentaires doivent assurer la définition des premier et dernier exons d'un transcrit (Inoue *et al.*, 1989; Niwa et Berget, 1991)

A l'heure actuelle, aucun modèle n'est réellement complètement élaboré pour comprendre la remarquable fidélité de l'épissage en ce qui concerne la sélection des sites d'épissage et leur appariement adéquat dans une polarité 5'-3'. Il est probable que des interactions au travers des introns et des exons se combinent pour spécifier les sites d'épissage adéquats au niveau d'ARN de grande taille et les appairer dans une polarité 5'-3' (Moore *et al.*, 1993). Des mécanismes assurant de telles interactions prendraient également toute leur importance lorsque la polarité 5'-3' doit être assurée au cours d'épissages alternatifs. L'existence d'épissage en trans, pour lequel la réaction a lieu entre des sites d'épissage de deux ARN physiquement distincts, montre bien qu'il doit exister des mécanismes particuliers intervenant dans l'appariement des exons.

L'épissage alternatif.

L'épissage alternatif est utilisé chez tous les métazoaires et constitue un niveau supplémentaire de régulation de l'expression génétique de nombreux gènes au cours de la différenciation cellulaire et de l'embryogénèse, permettant d'obtenir une variété de protéines à partir d'un même gène (Smith *et al.*, 1989; Maniatis, 1991; McKeown, 1992). Ce mécanisme peut également empêcher l'expression d'un gène en interrompant le cadre de lecture ouvert (Bingham *et al.*, 1988). Il existe plusieurs modes d'épissage alternatifs comme le montre la

figure 6. Ainsi, parfois, certains exons internes ne sont pas inclus dans le transcrit mature; c'est le cas d'un exon cassette et de l'exclusion mutuelle d'exon. Un exon interne peut également contenir plusieurs sites accepteur ou donneur d'épissage permettant ainsi de définir des exons différents. Ce mécanisme touche parfois les exons 5' ou 3' d'un gène lorsque que le gène est transcrit à partir de deux promoteurs ou lorsqu'il existe deux sites de polyadénylation en 3' du gène.

Il ne semble pas que les sites d'épissage au niveau des exons alternatifs soient impliqués directement dans l'épissage alternatif. En effet, dans la plupart des cas les sites 5' et 3' sont conformes aux séquences consensus d'épissage. D'autres éléments en cis et en trans doivent intervenir dans le mécanisme d'épissage alternatif. C'est le cas des facteurs ASF/SF2 et hnRNP A1, dont le rapport détermine la sélection d'un site 5' proximal ou distal (Ge et Manley, 1990; Mayeda et Krainer, 1992). Chez la drosophile, des protéines spécialisées peuvent également intervenir dans le mécanisme d'épissage. Les gènes *Sex lethal (sxl)*, *transformer (tra)* et *transformer2 (tra2)* qui participent à la cascade de gènes impliqués dans la mise en place du phénotype sexuel codent pour des facteurs ayant un rôle activateur ou répresseur dans l'épissage d'exons. Certains cas d'épissage alternatif impliquent des séquences au niveau de l'ARN pré-messager (Moore *et al.*, 1993). C'est le cas des exons 6a et 6b de l'ARN pré-messager du gène de la β -tropomyosine qui sont exclus mutuellement (figure 6b) où des structures secondaires dans l'intron sont impliquées dans le choix de l'exon à épisser (Clouet d'Orval *et al.*, 1991).

L'épissage en trans.

L'épissage en trans a été initialement décrit chez les trypanosomes et les nématodes (De Lange *et al.*, 1984; Milhausen *et al.*, 1984; Krause et Hirsh, 1987). Chez les trypanosomes, cet épissage s'effectue entre un site accepteur d'épissage localisé immédiatement en amont de tous les ARN pré-messagers et un site donneur d'épissage d'un ARN précurseur de 137 nucléotides. Tous les ARNm des trypanosomes sont ainsi pourvus d'une séquence 5'-proximale non codante, appelée séquence *leader* (Guyaux *et al.*, 1985; Campbell *et al.*, 1984; Kooter *et*

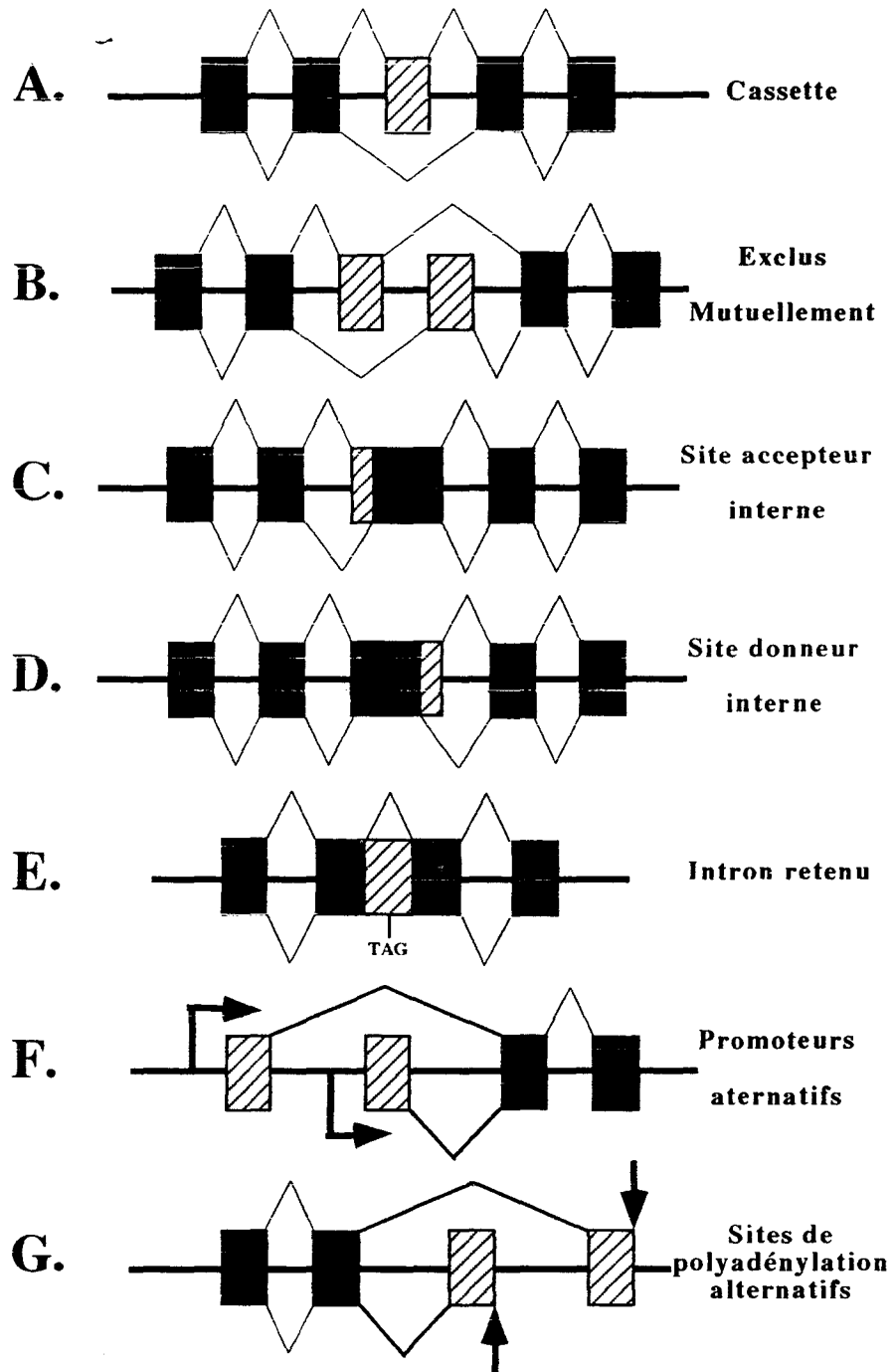


Figure 6: Modes d'épissage alternatifs.

Les exons constitutifs sont en noir et les exons alternatifs en hachurés. Les promoteurs (flèches horizontales) et les sites de polyadénylation alternatifs (flèches verticales) sont indiquées (d'après McKeown, 1992).

al., 1984). Chez les nématodes, l'épissage en trans a lieu pour certains ARN pré-messagers (Krause et Hirsh, 1987; Huang et Hirsh, 1989). Chez les nématodes *Caenorhabditis elegans* et *Ascaris lumbricoides*, les cellules musculaires peuvent effectuer les deux types d'épissage, en cis et en trans, de manière exclusive ou concomitante sur une même molécule d'ARN pré-messager. Il est possible que les séquences *leader* aient une influence sur la stabilité ou l'adressage intracellulaire des ARN qui les possèdent.

A ce jour, quelques cas d'épissage en trans ont été reportés chez les eucaryotes supérieurs (Bonen, 1993). Un des cas les plus intéressants semble être celui de l'oncogène *c-myc*. Chez le poulet et l'Homme, Vellard et coll. ont montré qu'il existait un phénomène de recombinaison en trans entre un exon appelé ET et l'ARN pré-messager du proto-oncogène *c-myc* (Vellard *et al.*, 1991). Par ailleurs, le brin complémentaire au niveau de l'exon ET code pour le facteur d'épissage SC35. Vellard et coll. ont proposé que la séquence ET et le facteur SC35 soient impliqués dans la régulation de l'expression du gène *c-myc* au cours de l'hématopoïèse (Vellard *et al.*, 1992).

L'épissage et l'organisation nucléaire.

Depuis ces dernières années, des études tendent à prouver qu'il est nécessaire de considérer le noyau de la cellule comme un espace hautement structuré. Cette organisation nucléaire participe sans aucun doute à la réalisation de la synthèse des ARN pré-messagers et à leur maturation (Hoffman, 1993). Dans ce sens, différents auteurs ont montré qu'il existait un mécanisme de recrutement de la machinerie de transcription et d'épissage au niveau des gènes actifs (Sass et Pederson, 1984; Fakan *et al.*, 1986; Beyer et Osheim, 1988). Différents facteurs protéiques et ribonucléoprotéiques ont pu être caractérisés au niveau de ce processus de recrutement, tels que le facteur d'épissage SC35 et des snRNP (Fu et Maniatis, 1990; Spector *et al.*, 1991). De récents travaux réalisés dans le laboratoire de J. Lawrence et d'autres travaux plus anciens montrent que les ARN pré-messagers subissent leur maturation le long d'une voie bien déterminée qui les conduira à une zone de l'enveloppe nucléaire, où ils seront exportés dans

le cytoplasme via les complexes de pores nucléaires (Carter *et al.*, 1993; Xing *et al.*, 1993). Une telle organisation nucléaire, qui a été nommée matrice nucléaire (nuclear matrix) ou échaffaudage nucléaire (nuclear scaffold) joue sans aucun doute un grand rôle dans le mécanisme d'épissage, participant de par ses interactions avec les transcrits et les complexes ribonucléoprotéiques à la parfaite réalisation de l'épissage.

B. De nouveaux rôles biologiques pour les ARN messagers.

Avec la découverte par Watson et Crick de la structure en double hélice de l'ADN, l'étape cruciale de l'existence d'une molécule pouvant être répliquée et portant l'information nécessaire à la perpétuation de tout système biologique était franchie. Cependant, entre cette molécule garante de la conservation des caractères héréditaires et les catalyseurs biochimiques que sont la plupart des protéines, manquaient un lien assurant le transfert d'information et une molécule originelle pouvant à la fois coder pour une information génétique et ayant la capacité de s'autorépliquer. L'hypothèse de l'adaptateur moléculaire de Francis Crick et la découverte de l'ARN messager par Jacques Monod, François Jacob et André Lwoff apporta la solution au problème du transfert d'information du gène à la protéine. Francis Crick fut l'un des premiers (Crick, 1967) à envisager que l'ARN pouvait être le matériel génétique autorépliatif originel, devançant d'une quinzaine d'années la découverte de l'activité catalytique de certaines molécules d'ARN, dans les laboratoires de Sydney Altman et Thomas Cech. En effet, Altman et Cech mirent respectivement en évidence l'activité catalytique de la fraction ARN de la RNase P et l'existence d'introns auto-épissables chez *Tetrahymena thermophyla* (Kruger *et al.*, 1982; Guerrier-Takada *et al.*, 1983). Cette découverte valut le prix Nobel de Chimie en 1989 à ces deux scientifiques et ouvrit la voie à l'étude de cette nouvelle classe de molécules d'ARN : les ribozymes.

Depuis la découverte des introns auto-épissables, le nombre de processus cellulaires impliquant de manière essentielle des molécules d'ARN n'a cessé d'augmenter. L'ARN n'est

Molécules d'ARN et fonctions biologiques.

(d'après Lamond et Gibson, 1990)

ARN	Fonctions
ARNm	Message spécifiant l'ordre d'assemblage des acides aminés au cours de la synthèse protéique. Fonctions ARN ? ex: Xist, HI9, 3'UTR
ARNr	1- Décodage de l'ARNm. 2- Peptidyl transférase ? 3- Translocation du peptidyl-ARNt.
ARNt	Adaptateur moléculaire amenant les acides aminés au cours de la traduction et d'autres voies d'anabolisme.
snRNA U1, U2, U4/6 et U5	Epissage des ARN pré-messagers.
snRNA U3	Maturation de l'ARNr ?
snRNA U7	Coupure de l'extrémité 3' des ARNm d'histones.
snRNA U11	Maturation de l'extrémité 3' des ARN pré-messagers.
ARN 2,5 S	Fraction catalytique de la <i>glucano transferase RNP</i> .
ARN M1	Fraction catalytique de la RNase P (maturation ARNt).
ARN de télomérase	Matrice pour la synthèse des télomères.
ARN 7S	Composant de la <i>signal recognition particle</i> (translocation des protéines).
Amorce ARN	Initiation de la réplication.
<i>micF</i> , etc	ARN de régulation procaryotes capables de moduler l'efficacité de traduction.

pas uniquement un acteur du dogme central au sein des processus de transcription et traduction, il joue également un rôle dans bien d'autres processus (Lamond et Gibson, 1990; tableau). L'existence de nouvelles fonctions pour des molécules d'ARN est ainsi à envisager.

I. L'inactivation du chromosome X et le gène Xist.

Le phénomène d'inactivation d'un chromosome X chez les mammifères femelles permet de compenser l'expression des gènes liés au chromosome X entre les males XY et les femelles XX. Les cellules somatiques femelles ne sont pas les seules où se déroule ce processus. En effet, durant la spermatogénèse, le seul X présent, encore actif au cours de la vie foetale et périnatale subit une inactivation au moment de la méiose pour n'être réactivé que beaucoup plus tardivement, après la fécondation.

L'expression du gène Xist, un des gènes exprimés à partir de l'X inactif, est étroitement corrélée à l'inactivation du chromosome X aussi bien dans les cellules somatiques femelles que dans les cellules germinales des deux sexes (McCarrey *et al.*, 1992; Richler *et al.*, 1992; Salido *et al.*, 1992). Kay et coll. ont de plus démontré que l'expression de Xist précédait l'inactivation du chromosome X, renforçant ainsi l'hypothèse que Xist serait, au moins en partie, à l'origine de l'inactivation chromosomique (Kay *et al.*, 1993).

Deux équipes ont dernièrement séquencé les homologues humain et murin de XIST. Les transcrits de 8 exons chez l'homme et 6 exons chez la souris, ont la caractéristique d'être de grande taille (15 kb et 17 kb) et la particularité de ne montrer aucun cadre de lecture ouvert (ORF pour *Open Reading Frame*) évident. En effet, le seul ORF conservé entre l'homme et la souris est de 25 résidus. Une phase de lecture ouverte de 400 pb a été identifiée chez l'homme, mais elle n'est pas conservée chez la souris. De plus, le transcrit Xist n'est pas associé aux polysomes et est localisé dans le noyau au niveau du corpuscule de Barr, spectre de l'X inactif. Ce résultat renforce l'hypothèse d'un rôle direct du transcrit Xist dans l'inactivation du chromosome (Brockdorff *et al.*, 1992; Brown C. J. *et al.*, 1992). Il a été proposé que le transcrit puisse être fonctionnel en tant que molécule d'ARN. L'ARN messager Xist pourrait agir localement avec

la chromatine du chromosome dont il est issu, peut être par l'intermédiaire d'autres facteurs nucléaires qui propageraient le processus d'hétérochromatisation. Une autre hypothèse envisagée par différents auteurs est que Xist serait un organisateur chromatinien. C'est la transcription en elle-même du locus qui pourrait amener des facteurs nucléaires à se fixer sur le chromosome et à propager l'inactivation (Brockdorff *et al.*, 1992; Brown C. J. *et al.*, 1992).

II. Le gène H19.

Le gène murin H19 a été initialement identifié dans une banque d'ADNc de foie foetal. Dans l'embryon, le gène H19 est un des gènes les plus fortement exprimé dans l'endoderme et le mésoderme. A la naissance, cette expression élevée est réprimée dans tous les tissus à l'exception du muscle squelettique (Pachnis *et al.*, 1988).

L'analyse de la séquence du gène murin H19 révèle la présence d'une multitude de petits ORFs dont aucun ne s'étend sur plus de deux des cinq exons du locus. Le plus grand de ces ORFs est entièrement localisé dans le premier exon et pourrait potentiellement coder une protéine de 132 acides aminés. Brannan et al ont cloné et séquencé le gène H19 humain, dans le but de comparer les homologues murin et humain et de trouver les régions conservées, donc potentiellement importantes pour la fonction du gène. A leur grande surprise ils ne trouvèrent pas d'ORF commun, bien que les séquences nucléotidiques des transcrits murin et humain soient bien conservées (Brannan *et al.*, 1990).

L'ARNm H19 de 2,5 kb est comme les autres ARNms nucléaires, transcrit par la polymérase II, épissé et polyadénylé. Cependant, il n'est pas associé aux ribosomes dans les tissus et les stades de développement examinés par les auteurs. L'analyse de sa distribution subcellulaire montre qu'un grand pourcentage d'ARN se situe dans une particule cytoplasmique ribonucléoprotéique de constante de sédimentation de 28S. Afin de savoir si ce gène est fonctionnel, Brunkow et coll. ont introduit des copies en excès du gène H19 dans des zygotes murins (Brunkow *et al.*, 1991). L'expression ectopique provoque la mort entre le 14ème jour embryonnaire et la naissance. S'il est peu probable que le gène H19 soit traduit, sa conservation

entre espèces ainsi que son motif d'expression au niveau embryonnaire et les résultats des expériences de transgénèse laissent présager un rôle important pour son transcrit. Sans avancer de mode d'action, les auteurs envisagent la possibilité que le produit du gène HI9 fonctionne en tant que molécule d'ARN.

Sous réserve de plus amples résultats, H19 et Xist représentent des exemples d'ARNm complètement maturés qui pourraient avoir un rôle au delà du simple transfert d'information de l'ADN à la protéine. Nous allons voir que de tous récents travaux montrent sous un nouveau jour, non plus cette fois l'ensemble du messenger mais une de ces régions: les séquences 3' non traduites.

III. Les séquences 3'UTR.

La région comprise entre le codon d'arrêt de la traduction et la queue de polyadénylation (3' UnTranslated Region; 3' UTR) des ARN messagers, après avoir été longtemps considérée comme une séquence sans rôle particulier, s'est révélée être le siège de différents contrôles concernant aussi bien la traduction, la stabilité et la localisation cellulaire du messenger considéré (Jackson, 1993). Ces trois points font donc intervenir des séquences en cis de l'ARNm et des facteurs en trans, impliqués dans la traduction, la protection ou la dégradation de la molécule d'ARN et l'adressage cellulaire de cette molécule. Rastinejad et Blau viennent récemment de démontrer que ces séquences 3'UTR, parties à part entière des ARN messagers, pouvaient non seulement contrôler les ARNms dans lesquels elles se trouvent mais également d'autres gènes en trans (Rastinejad et Blau, 1993).

La démarche initiale des auteurs était d'identifier les gènes impliqués dans la différenciation musculaire. Une variété de gènes spécifiques du tissu musculaire commencent à être exprimés lorsque la prolifération de myoblastes en culture est arrêtée. Ces myoblastes fusionnent et tendent même à former des myotubes pourvus de propriétés contractiles. Rastinejad et Blau ont utilisé une lignée cellulaire de myoblaste mutante incapable de se différencier et de fusionner.

Le promoteur du gène de l'actine, spécifique du tissu musculaire, ne peut être activé dans une telle lignée cellulaire. En cotransfectant dans ces cellules un gène marqueur pourvu du promoteur de l'Actine, et des vecteurs contenant l'ADNc d'une banque de cellules musculaires différenciées, ils ont pu identifier les gènes induisant une différenciation cellulaire. D'autres auteurs avaient déjà montré que certains gènes tels que MyoD, la myogénine, MFR4 et myf5 pouvaient convertir des fibroblastes en myoblastes, impliquant le fait que ces gènes pourraient réguler la différenciation musculaire (Thayer *et al.*, 1989). Les séquences isolées par Rastinejad et Blau ne correspondent en rien à de nouveaux gènes, mais représentent les séquences 3' UTRs d'ARNms spécifiques du tissu musculaire: l'actine, la tropomyosine et la troponine I. Chacune de ces séquences 3' UTR peut stimuler l'expression d'ARNm d'autres gènes spécifiques du tissu musculaire.

Si d'autres expériences doivent être effectuées afin de mieux comprendre le rôle des séquences 3'UTR dans la différenciation cellulaire, plusieurs modes d'actions ont été proposés comme la titration de facteurs se liant à des séquences d'ARN, une action directe de ces séquences sur d'autres ARNm par appariement de bases ou action catalytique (Rastinejad et Blau, 1993).

Les ARN messagers ne comportant pas d'ORF et les séquences 3' non traduites semblent donc avoir une fonction autre que le transfert d'information pour la synthèse protéique. Ceci montre la richesse de fonctions potentielles que pourraient avoir les molécules d'ARN dits messagers. De telles fonctions, sous réserve qu'elles soient démontrées, pourraient être des reliques du «monde de l'ARN» originel, ou de nouveaux essais dans le processus d'évolution cellulaire. Dans ce sens, certains auteurs ont démontré la propriété qu'avaient certaines séquences synthétiques d'ARN de lier de manière très spécifique de petites molécules organiques, telle l'ATP, sans pour autant retrouver ces motifs ARN dans le patrimoine cellulaire (Sassanfar et Szostak, 1993).

C. Le proto-oncogène c-ets-1.

Bon nombre de proto-oncogènes cellulaires (c-onc) ont été identifiés comme étant les homologues normaux de gènes transduits par des rétrovirus, générant ainsi des gènes transformants appelés oncogènes viraux (v-onc). Les produits de ces proto-oncogènes cellulaires appartiennent aux différentes classes de protéines participant au contrôle de la prolifération et de la différenciation cellulaires. Ces protéines exercent diverses fonctions aux différents niveaux des voies de transduction qui propagent un signal parvenant à la surface cellulaire (facteur de croissance) jusqu'au noyau (facteur de transcription) (pour revue voir Hunter, 1991). De manière générale, ces proto-oncogènes sont hautement conservés au cours de l'évolution, révélant ainsi leurs rôles essentiels dans le développement et la différenciation cellulaires. Nous verrons dans ce chapitre, que le proto-oncogène c-ets-1 en est un parfait exemple.

I. Du rétrovirus E26 à la famille des gènes ets.

I.1 Le rétrovirus E26.

E26 est un rétrovirus déficient pour sa réplication qui induit *in vivo* des leucémies mixtes érythroïdes et myéloïdes chez le Poulet (Moscovici *et al.*, 1981; Radke *et al.*, 1982) et transforme *in vitro* les précurseurs des lignées érythroïdes et myélomonocytaires (Radke *et al.*, 1982; Moscovici *et al.*, 1983).

La particularité de ce rétrovirus est qu'il a transduit deux proto-oncogènes cellulaires. La première séquence identifiée fut celle de l'oncogène v-myb, dérivée du proto-oncogène cellulaire c-myb tronquée dans sa partie carboxy-terminale (Klempnauer *et al.*, 1982; Gonda et Bishop, 1983; Nunn *et al.*, 1983; Rosson et Reddy, 1986). c-myb et v-myb codent tous les deux des protéines nucléaires transactivatrices. Des auteurs ont montré que l'expression du proto-oncogène c-myb est élevée dans les cellules hématopoïétiques immatures et inhibée au cours du processus de différenciation cellulaire, suggérant ainsi un rôle spécifique de c-myb

durant la différenciation hématopoïétique (pour revue Luscher et Eisenman, 1990). C'est en 1983 que deux équipes mirent en évidence l'existence d'une seconde séquence oncogénique, v-ets, fusionnée à la séquence v-myb dans le rétrovirus leucémogène aviaire E26 (lequel lui donna son nom: E-Twenty-Six) (Leprince *et al.*, 1983; Nunn *et al.*, 1983).

Le génome de E26 est transcrit en un ARNm de 5,7 kb codant une protéine de fusion unique à localisation nucléaire et se fixant à l'ADN: la p135^{gag-myb-ets} (Klempnauer *et al.*, 1984; Moelling *et al.*, 1985) (Figure 7). La p135^{gag-myb-ets} comprend ainsi la protéine d'enveloppe virale gag, une séquence tronquée du proto-oncogène cellulaire c-myb et une séquence dérivée du proto-oncogène c-ets. La p135^{gag-myb-ets} possède à chacune de ses extrémités un domaine de liaison à l'ADN et au centre un domaine de transactivation, agissant probablement comme un régulateur transcriptionnel (Aurigemma, 1992).

L'activité transformante complète de E26 requiert la présence des deux séquences v-myb et v-ets fusionnées au sein de cette protéine (Metz et Graf, 1991a, 1991b). De nombreux travaux ont été réalisés afin de connaître les rôles respectifs des séquences v-myb et v-ets dans la transformation par E26 de la lignée érythroïde d'une part et myéloïde d'autre part. Pour la transformation de la lignée érythroïde, l'intégralité de la séquence codante de v-ets est nécessaire, mais la présence de v-myb en trans ou en cis est également requise (Golay *et al.*, 1988; Beug *et al.*, 1984; Metz et Graf, 1991a; Domenget *et al.*, 1992). Cependant il n'existe pas de rôle absolu pour l'un ou l'autre des domaines, puisque la coopération des deux séquences semble exister pour l'efficacité et la stabilité de la transformation (Domenget *et al.* 1992).

1.2 Le proto-oncogène aviaire c-ets-1.

Le progéniteur cellulaire de v-ets est le gène Ck-c-ets-1, dont le locus est composé d'au moins 10 exons, répartis sur 60 kb (Figure 7). Il est transcrit en un doublet d'ARNm de 7,5 kb et de deux ARNm mineurs de 2,2 kb et 1,5 kb. La synthèse des messagers principaux est générée par un épissage alternatif en 5' qui accole soit les exons α et β (ou I et II), soit l'exon I⁵⁴ (ou exon A) aux exons 3' communs a à F (ou III à IX) (Leprince *et al.*, 1988). Deux protéines

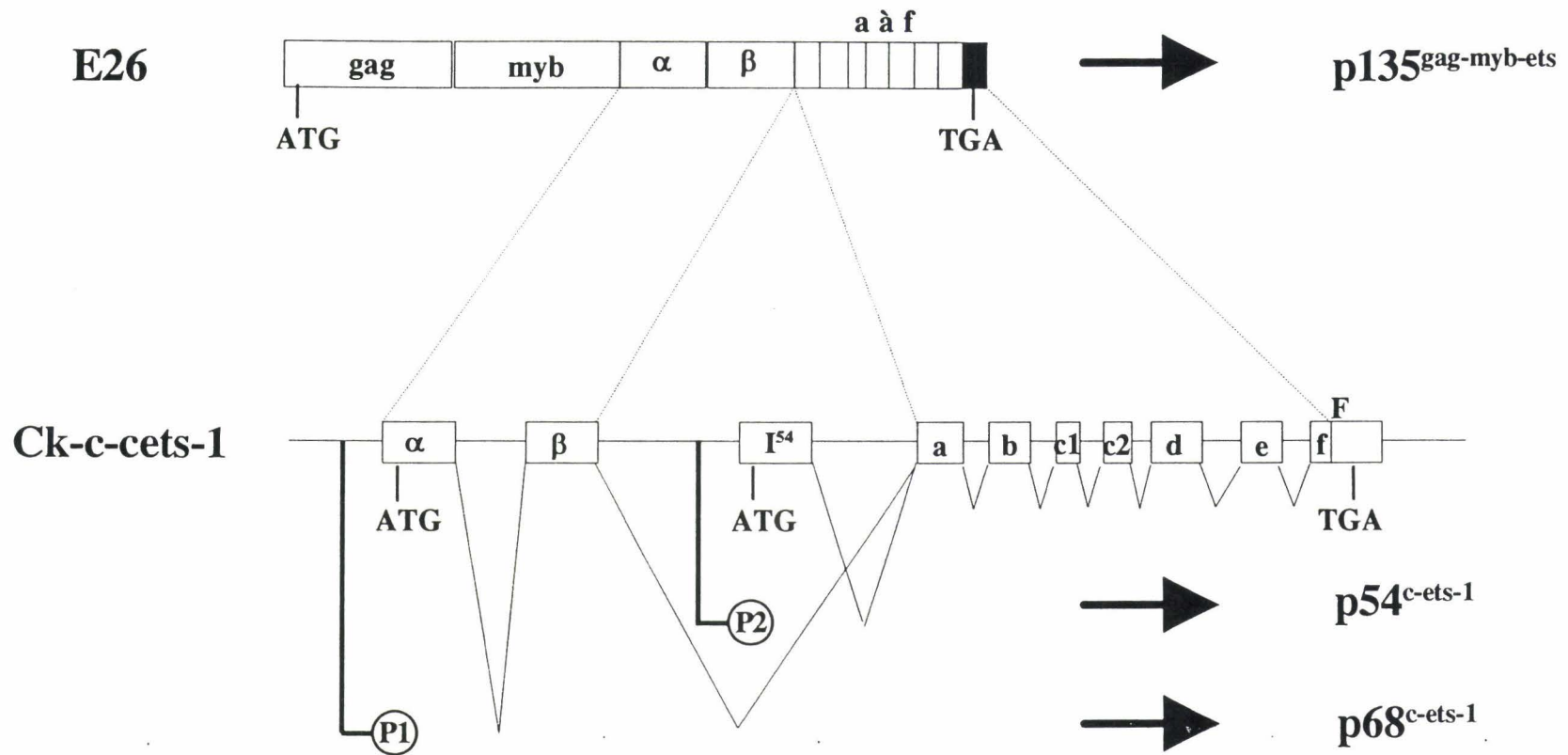


Figure 7. Organisations du locus *c-ets-1* aviaire et du génôme du virus E26.

Les séquences LTR ne sont pas représentées pour le génôme du virus E26. Les exons α et β et les exons de a à F ont été transduits (l'exon F n'est pas transduit dans sa totalité, la séquence transduite est notée f; la boîte noire du génôme de E26 représente une séquence d'origine inconnue). E26 code pour la $p135^{\text{gag-myb-ets}}$. Les promoteurs (P1 et P2) et l'épissage alternatifs pour le gène *Ck-c-ets-1* sont représentés. Le transcrit comprenant les exons α et β code pour la $p54^{\text{c-ets-1}}$, et celui commençant par l'exon I^{54} code pour la $p68^{\text{c-ets-1}}$. Les codons d'initiation de la traduction et le codon stop sont indiqués.

distinctes mais apparentées en sont traduites: il s'agit respectivement de la p68^{Ck-c-ets-1} (Gégonne *et al.*, 1987a) et de la p54^{Ck-c-ets-1} (Duterque-Coquillaud *et al.*, 1988) (Figure 7).

C'est l'ARNm de la p68^{Ck-c-ets-1} qui est le véritable progéniteur cellulaire de v-ets puisqu'il possède les deux exons α et β 5' spécifiques. Mais il en diffère, par sa fusion aux séquences gag et myb dans la p135^{gag-myb-ets} de E26 et par plusieurs mutations (Gégonne *et al.*, 1987a; Watson *et al.*, 1988a; Gégonne *et al.*, 1987b; Leprince *et al.*, 1988; Leprince *et al.*, 1992; Leprince *et al.*, 1993).

La p54^{Ck-c-ets-1} est exprimée à un taux modéré dans la plupart des tissus et à un taux élevé dans les cellules lymphoïdes T et B du thymus, de la rate et de la bourse de Fabricius (Bhat *et al.*, 1987; Bhat *et al.*, 1989; Bhat *et al.*, 1990; Chen, 1985; Ghysdael *et al.*, 1986). A l'inverse, la p68^{Ck-c-ets-1} a été initialement détectée à un faible taux dans la rate, dans une fraction de cellules correspondant aux vaisseaux sanguins et au stroma. Elle n'est détectée ni dans les cellules lymphoïdes de la rate, ni dans les autres cellules hématopoïétiques (macrophages, granulocytes, érythrocytes) (Leprince *et al.*, 1988; Leprince *et al.*, 1990; Queva *et al.*, 1993). Ces résultats ont été obtenus à l'aide des techniques de Northern-blot et d'immunoprécipitation. La technique d'hybridation *in situ*, appliquée à des coupes d'embryons de Poulet, a permis de détecter au début de l'organogénèse un signal intense et transitoire dans une grande variété de tissus d'origine mésodermique, et dans les cellules de la crête neurale pendant la neurulation. De plus, tout au long du développement embryonnaire, c-ets-1 s'exprime dans les cellules endothéliales sur la paroi des vaisseaux sanguins en formation, puis s'éteint quand la mise en place du vaisseau s'est accomplie. Cependant, de par la nature des sondes ribonucléotidiques employées, il est impossible d'attribuer les signaux observés à l'un ou l'autre des transcrits (Vandenbunder *et al.*, 1989; Desbiens *et al.*, 1991).

I.3 La famille des gènes ets.

Depuis la découverte du proto-oncogène c-ets-1 chez le Poulet, un grand nombre de gènes codant des facteurs de transcription homologues ont été identifiés. La famille des gènes ets est

définie d'après le domaine de liaison à l'ADN nommé ETS (Karin *et al.*, 1990). Plusieurs auteurs ont déterminé la très grande conservation phylogénétique de cette famille de facteurs de transcription (Lautenberger *et al.*, 1992; ; Laudet *et al.*, 1993). Degnan et coll. en utilisant la technique de PCR à l'aide d'amorces oligonucléotidiques dégénérées dans le domaine ETS, ont détecté les gènes ets chez plusieurs métazoaires inférieurs et supérieurs (Degnan *et al.*, 1993). Ils n'ont pu détecter les gènes ets chez les champignons, les végétaux et plusieurs génomes de protozoaires. Les auteurs concluent que la duplication et la divergence des gènes ets s'est déroulée de façon concomitante avec le développement des métazoaires animaux.

II. Le proto-oncogène humain c-ets-1: ETS-1.

II.1 Localisation chromosomique et expression.

Le gène ETS-1, homologue de la p54^{Ck-c-ets-1}, est localisé sur le chromosome 11 en 11q23-24, et comporte huit exons sur plus de 60 kb (de Taisne *et al.*, 1984; Jorcyk *et al.*, 1991) (Figure 8). ETS-1 est transcrit en un ARN majeur de 6,8 kb (Watson *et al.*, 1985). Un second transcrit de 2,7 kb dû à l'emploi d'un site alternatif de polyadénylation a été mis en évidence parallèlement par plusieurs auteurs (Wernert *et al.*, 1992; Collyn-d'Hooghe *et al.*, 1993). L'ARNm de 6,8 kb est exprimé préférentiellement dans le tissu lymphoïde adulte (Chen *et al.*, 1985; Sacchi *et al.*, 1988), à un haut niveau dans le thymus et dans les lymphocytes T périphériques (Bhat *et al.*, 1990) et à des niveaux moindres dans le poumon et la rate (Bath *et al.* 1987; Seth *et al.*, 1990). L'expression de ETS-1 est également trouvée dans des cellules T immatures leucémiques, dans différentes cellules leucémiques ou lignées cellulaires (Sacchi *et al.*, 1988). *In vitro*, lorsque des lymphocytes T périphériques (moins de 48 h de culture) sont stimulés par la phytohématagglutinine ou par un ionophore, l'expression de ETS-1 diminue rapidement (Sacchi *et al.*, 1988). Ainsi, il a été avancé que ETS-1 pourrait avoir un rôle de maintien des cellules dans un état quiescent. De plus, tout comme pour Ck-c-ets-1, l'expression de ETS-1 a été visualisée sur des coupes d'embryons humains au niveau des cellules endothéliales

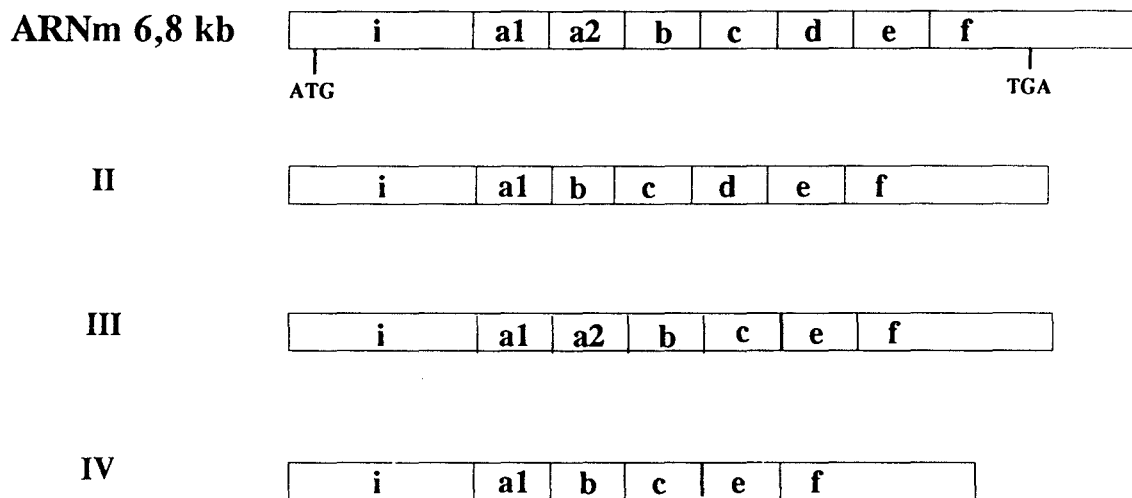
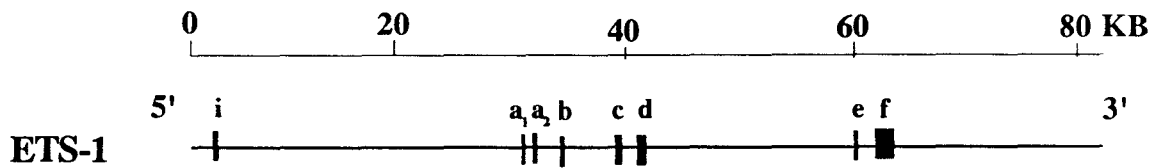


Figure 8. Organisation du locus c-ets-1 humain et transcrits alternatifs exprimés.

Le locus contient huit exons sur plus de 60 kb. Quatre transcrits sont exprimés par épissage alternatif. Le transcrit de 6,8 kb contient tous les exons du locus. Les exons a2 et d sont la cible d'épissage alternatif. L'excision de l'un ou l'autre exon, ou des deux, est à l'origine des autres transcrits (II, III, IV). Le codon d'initiation de la traduction et le codon stop sont indiqués.

des vaisseaux en formation (Wernert *et al.*, 1992).

L'ARN pré-messager de ETS-1 peut être différemment épissé. En effet, il a été montré que les exons a2 et d et les deux exons ensemble peuvent subir un épissage alternatif (Reddy *et al.*, 1988; Koizumi *et al.*, 1990; Jorcyk *et al.*, 1991; Collyn-d'Hooghe *et al.*, 1993; figure 8).

La région 11q23-24 est le lieu de nombreux remaniements (translocations, délétions, amplifications ou insertions inversées) qui sont impliqués dans différents types de leucémies, de lymphomes, de syndromes myéloprolifératifs, mais plus particulièrement dans les leucémies aiguës monocytaires (Berger *et al.*, 1980; Berger *et al.*, 1985; Yunis *et al.*, 1986). De par sa localisation chromosomique en 11q23-24, ETS-1 a été longtemps suspecté de jouer un rôle dans ces pathologies. Cependant l'absence de réarrangement du locus de ETS-1 dans ces pathologies a été démontrée par différents auteurs (Diaz *et al.*, 1986; Sacchi *et al.*, 1986; Ohyashiki *et al.*, 1987; Katz *et al.*, 1988; Kennedy *et al.*, 1989; Shipley et Sheer, 1989; Akao *et al.*, 1990). Récemment, les gènes situés aux points de cassure en 11q23 ont été identifiés (Cimino *et al.*, 1991; Djabali, 1992; Ziemer-Van der Piels *et al.*, 1991).

II.3 Les protéines ets-1.

Domaines fonctionnels des protéines ets.

La comparaison de la séquence en acides aminés de la famille des gènes ets avec d'autres facteurs de transcription a révélé plusieurs régions d'homologie significative:

- le domaine de dimérisation des protéines de la famille Helix-Loop-Helix (HLH).
- le site de reconnaissance de la protéine kinase II calmoduline-dépendante (CCKII kinase).
- les signaux de localisation nucléaire de certains virus transformants à ADN tels que SV40 et le virus du polyome.
- un domaine de liaison à l'ADN unique.

La région C-terminale comprenant le domaine de liaison à l'ADN et le signal de localisation nucléaire est très conservée phylogénétiquement. Elle permet de définir la famille

des protéines ets et argue pour une fonction cellulaire importante (Watson *et al.*, 1988).

Dans leur partie N-terminale les protéines ets (ets-1, ets-2 et erg) contiennent 60 acides aminés qui ont une faible homologie avec le domaine HLH de la famille des protéines HLH, telles que myc, max, myoD, Id, emc, et E12 (Seth *et al.*, 1990; Murre *et al.*, 1989). La région HLH est une caractéristique essentielle d'un certain nombre de facteurs de transcription semblant être requise pour des interactions protéine-protéine (Seth *et al.*, 1990; Murre *et al.*, 1989; Lassar *et al.*, 1989).

La grande majorité des protéines de la famille ets contiennent des sites de phosphorylation RXXS/T, cible de la protéine kinase II dépendante de la calmoduline (Watson *et al.*, 1988; Koizumi *et al.*, 1990). Des expériences de marquage *in vitro* ont démontré que ces sites pouvaient être en effet phosphorylés par la CCKII (Watson *et al.*, 1988; Koizumi *et al.*, 1990). Le traitement des cellules avec un ionophore au calcium, aussi bien que le traitement avec un anticorps anti récepteur des cellules T augmente le taux de phosphorylation (Koizumi *et al.*, 1990; Fujiwara *et al.*, 1988; Pognonec *et al.*, 1988). Dans les cellules humaines, différentes protéines, issues d'épissage différentiels et/ou de forme différemment phosphorylées ont été mises en évidence par immunoprécipitation: des protéines de 48 kDa et 51 kDa, leurs formes phosphorylées respectives de 49 kDa et 52 kDa, ainsi que deux protéines de 39 kDa et 42 kDa traduites à partir d'ARNm sans exon d (Koizumi *et al.*, 1990; Pognonec *et al.*, 1990). Fischer et coll ont démontré que tous les isoformes de ETS-1 commençaient au même ATG et avaient la même extrémité c-terminale (Fischer *et al.*, 1992). Ils ont également montré que les p48 et p39 pouvaient être dérivées des p51 et p42 par la modification covalente de groupes SH. Fleischman et coll. ont montré que ets-1 était hyperphosphorylée en début de mitose sans que cela n'altère la liaison à l'ADN (Fleischman *et al.*, 1993). La phosphorylation nécessite la présence de la région correspondant à l'exon d, où sont localisés les sites consensus de phosphorylation Ca²⁺/Calmoduline Kinase II (CCKII). Les protéines traduites à partir des ARNm sans exon d ne sont pas phosphorylables mais se fixent plus efficacement à l'ADN et gardent la capacité de transiter vers le noyau (Wasylyk *et al.*, 1992). Par contre, la phosphorylation rapide et dépendante du calcium de la p48 abolit sa liaison spécifique à l'ADN

in vitro, diminue son affinité pour la chromatine et affecte sa localisation subcellulaire : elle reste cytoplasmique (Pognonec *et al.*, 1988).

Les protéines ets contiennent des signaux de localisation nucléaire similaires à ceux trouvés dans l'antigène grand T de SV40 (Kalderon *et al.*, 1984; Seth *et al.*, 1990). Boulukos et coll ont montré que ces signaux étaient suffisants pour que la protéine ets-1 soit transloquée au noyau (Boulukos *et al.*, 1989).

Le domaine de liaison à l'ADN, riche en résidus basiques Lysine et Arginine, ne correspond à aucun type de domaine de liaison à l'ADN déjà caractérisé tels que: le domaine HLH (Laughon et Scott, 1984), le domaine homéo (Laughon et Scott, 1984), le domaine en doigt de zinc (pour revue Evans et Hollenberg, 1988), le motif de «fermeture à glissière à Leucines» (Landschulz, 1988). Le domaine de liaison à l'ADN des protéines ets est de type «tryptophane-repeat» (trois résidus tryptophanes séparés par 17 ou 18 acides aminés) et pourrait avoir une structure de type HLH. Ce type de motif a été montré pour être fonctionnellement important dans le domaine de liaison à l'ADN du facteur de transcription myb (Saikumar *et al.*, 1990; Anton, 1988). De manière significative, cette région C contient également deux motifs basiques riches en Arginine et Lysine. De tels résidus chargés positivement pourraient être directement impliqués dans l'activité de liaison à l'ADN (Seth, 1992).

Les protéines ets, facteurs de transcription.

Le produit du gène ets-1 se lie de manière spécifique à une séquence consensus GCA/CGGAAG présente dans la séquence activatrice du promoteur précoce du polyomavirus (Wasylyk *et al.*, 1990), dans le LTR du virus du sarcome murin de Moloney (MSV-LTR) (Gunther *et al.*, 1990), à la séquence activatrice du gène codant le récepteur alpha des cellules T (Ho *et al.*, 1990) et au LTR du virus leucémogène humain des cellules T de type 1 (HTLV-1; Bosselut *et al.*, 1990). Il a été démontré de même que le produit de ets-1 activait la transcription des gènes placée sous le contrôle de ces séquences. Plus récemment, les sites de liaison consensus ETS ont été définis comme 5'-A/GCCGGAA/TGT/C-3' (Woods *et al.*, 1992; Nye *et al.*, 1992).

La transactivation par le site PEA3 est l'un des événements clé pour la transformation cellulaire. C'est une cible nucléaire d'oncogènes localisés plus en amont dans la transduction du signal (v-raf, v-mos, v-src, moyen T du polyome, c-Ha-ras), de promoteurs tumoraux et de composants du sérum (Wasylyk *et al.*, 1989). Des sites PEA3 sont présents en amont de certains promoteurs parmi lesquels ceux des gènes codant la collagénase I, la stromélysine I ou l'activateur du plasminogène de type urokinase. Les protéines ets-1 transactivent effectivement l'expression de ces gènes *in vitro* (Gutman *et al.*, 1991; Wasylyk *et al.*, 1991) mais également l'expression de son propre gène (Majérus *et al.*, 1992).

OBJECTIFS

Lors de notre DEA en 1989/1990, alors que la fonction des protéines ets-1 était encore inconnue, nous nous proposons de rechercher d'éventuels exons alternatifs apparentés aux exons α et β de Ck-c-ets-1, en amont de l'exon I. Les exons I et de a à f étant parfaitement conservés en taille et en séquence entre l'Homme et le poulet (Watson *et al.*, 1988), il semblait curieux que des études en Northern et Southern-blot n'aient pu révéler la présence des exons α et β dans le locus humain. De plus, dans l'optique où ETS-1 pouvait être impliqué dans des hémopathies malignes, il était important de déterminer exactement la structure complète du locus et les différents produits exprimés.

La technique d'amplification génique par la polymérase ou Polymerase Chain Reaction (PCR) appliquée aux ARN (RT-PCR) fut employée dans ce but. La grande sensibilité et la spécificité de la RT-PCR, expérimentalement modulables par le choix des amorces oligonucléotidiques et les conditions de stringence, devaient ainsi nous permettre de mettre en évidence d'éventuels transcrits alternatifs, aussi faiblement exprimés soient-ils et comportant des exons de plus ou moins grande homologie de séquence avec les exons α et β de poulet.

Les résultats de nos travaux furent tout à fait surprenants. En effet, nous obtinrent un produit d'amplification unique dans lequel aucune séquence ne peut s'apparenter aux exons α ou β . En contrepartie, la séquence de ce produit d'amplification révéla l'existence d'une jonction exonique parfaite entre les exons a1 et d du locus de ETS-1, mais de polarité opposée à celle du locus. Le site donneur d'épissage de l'exon d et le site accepteur d'épissage de l'exon a1 semblaient avoir participé à un processus d'épissage normal. Une telle jonction exonique est cependant en parfait désaccord avec le principe d'ordre connu pour le mécanisme d'épissage.

Peu de temps après, des travaux réalisés dans le laboratoire de Bert Vogelstein aux Etats-Unis concernant le gène DCC (Deleted in Colorectal Carcinoma), un gène suppresseur de tumeur dans les carcinomes colorectaux (Fearon *et al.*, 1990), mirent en évidence plusieurs transcrits comprenant une jonction exonique inversant l'ordre des exons, comme dans le cas de ETS-1 (Nigro *et al.*, 1991).

Sous réserve que ce phénomène puisse être étendu à d'autres gènes, nous sommes donc apparemment en présence d'un mécanisme particulier prenant part dans l'épissage des ARN pré-

messagers. La compréhension du processus et du rôle éventuel d'un tel mécanisme dans l'expression d'oncogènes et de gènes suppresseurs de tumeur est ainsi du plus grand intérêt.

Notre travail aura ainsi pour objectifs:

- de mettre en évidence les différents transcrits possibles de ETS-1 comprenant un tel type de jonction exonique inversant l'ordre des exons.

- de mieux caractériser ces différents transcrits possibles, en obtenant leur séquence complète et en étudiant leur localisation cellulaire et leur stabilité.

- d'envisager quels pourraient être les «facteurs» (séquences, facteurs protéiques, contexte cellulaire) permettant un tel mécanisme.

A la suite de ces travaux nous discuterons de l'éventuel rôle biologique de ce nouveau type de transcrits.

RESULTATS

A. La découverte de l'épissage inversé.

Les résultats et l'observation décrits dans la publication, EMBOJ. vol. 11 n°3. 1095-1098, en 1992 ont été obtenus suite à l'étude de l'organisation du locus de ETS-1 et la comparaison avec l'organisation du locus chez le poulet. Par la technique de PCR, technique sensible mais aussi expérimentalement modulable par le choix des amorces oligonucléotides et les conditions de stringence, nous avons mis en évidence deux transcrits comportant des jonctions tout à fait inattendues entre des exons, ceci à partir de la fraction poly A⁻ des ARN totaux. En effet, la séquence partielle de ces transcrits montre que des sites donneur d'épissage d'exons en 3' du gène ETS-1 ont fusionnés avec le site accepteur d'épissage d'un exon situé en amont. Nous avons confirmé l'existence de l'un de ces transcrits, à l'aide de la technique de protection aux ribonucléases et nous avons pu quantifier sa présence à approximativement 1/100^{ème} du transcrit normalement épissé. De plus l'étude des séquences génomiques et la position des exons impliqués dans cette réaction démontrent que les sites d'épissage de la maturation normale ont été utilisés dans un ordre différent de celui du génome, pour former ce transcrit. Ces produits sont observés dans plusieurs lignées cellulaires humaines dans lesquelles nous avons vérifié que le locus de ETS-1 était non réarrangé. Les conclusions de cet article dans EMBO J., ont été reprises par l'éditeur de Trends in Genetics en mai 1992 dans la rubrique *Monitor*.

Une observation similaire a été décrite presque simultanément dans la littérature par le groupe de Bert Vogelstein pour le gène DCC (Nigro *et al.*, 1991).

Splicing with inverted order of exons occurs proximal to large introns

Claude Cocquerelle, Pierre Daubersies,
Marie-Ange Majerus, Jean-Pierre Kerckaert and
Bernard Bailleul¹

Unité 124 INSERM, Institut de Recherches sur le Cancer, Place de Verdun, 59045 Lille Cedex, France

¹Corresponding author

Communicated by E. Brody

Following our studies which showed that the α and β exons of the chicken *c-ets-1* gene are not conserved in the human homologue, we succeeded in identifying a novel human *c-ets-1* transcript in which the normal order of exons is scrambled. By PCR and RNase protection assays, we demonstrated that while the order of exons is different from that in genomic DNA, splicing of these exons in aberrant order occurs in pairs and at the same conserved consensus splice sites used in the normally spliced transcript. The scrambled transcript is non-polyadenylated and is expressed at much lower levels than the normal transcript. It is not the consequence of genomic rearrangement at the *ets-1* locus nor is it due to the transcription of any *ets-1* pseudogene. These results confirm previous observations of scrambled splicing.

Key words: alternative splicing/intron/oncogene/*c-ets-1*

Introduction

The *c-ets-1* proto-oncogene encodes a transcription factor that cooperates with AP-1 in transcriptional activation (Wasylyk *et al.*, 1990). The chicken *c-ets-1* locus encodes two alternatively spliced transcripts which encode p54^{*ets-1*} and p68^{*ets-1*} respectively (Duterque-Coquillaud *et al.*, 1988; Gegonne *et al.*, 1987). p54^{*ets-1*} is homologous to the human p51^{*ets-1*} (Watson *et al.*, 1988). p68^{*ets-1*} contains two alternatively spliced exons, α and β (Leprince *et al.*, 1988), which to date have not been detected in human. It is the p68^{*ets-1*} transcript which has been transduced by the E26 retrovirus (Leprince *et al.*, 1983; Nunn *et al.*, 1983). PCR experiments were performed in order to determine whether these exons exist in human. We failed to detect exons α and β in the human genome but in the process of our studies, we detected abnormally spliced products. In these transcripts, exon d or c splice donor sites were joined at the exon a1 splice acceptor site.

Splicing is mediated by a RNA-protein complex, the spliceosome, which catalyses splicing between exons at consensus splice sites (Steitz *et al.*, 1987; Brody and Abelson, 1985). As many pre-mRNAs contain more than one intron, a mechanism has been postulated which ensures that cognate 5' and 3' splice sites are appropriately used such that aberrant splicing does not occur (Aebi and Weissmann, 1987).

In this paper we present evidence from PCR and RNase

protection experiments that pairs of exons of the human *ETS-1* gene are joined at consensus splice sites, but in a different order from that present in genomic DNA. Similar observations have recently been described for the *DCC* gene (deleted in colorectal carcinomas) in which the normal order of exons is scrambled (Nigro *et al.*, 1991). It is noteworthy that abnormal splicing occurs only with *ETS-1* exons c, d and a1. These exons are close to the largest introns (Kerckaert *et al.*, 1990; Jorcyk *et al.*, 1991) and the exon d skips during pre-mRNA maturation, at relatively high frequency (Jorcyk *et al.*, 1991; Reddy and Rao, 1988).

Results and discussion

We performed double PCR amplifications with anti-sense primers to exon a1 and with a sense primer to exon I of human *ETS-1* (Reddy and Rao, 1988) or with sense primers homologous to the additional 5' exons, α and β , that were identified in the chicken gene (Leprince *et al.*, 1988) (Figure 1a and b). PCR products of the expected size (93 bp) amplified from the *ETS-1* 6.8 kb mRNA (Reddy and Rao, 1988; Watson *et al.*, 1985) were found in different cell lines, including CEM, HeLa and A431 (Figure 1c). Using the oligomer O27 (27mer homologous to chicken exon β) as sense primer, double PCR amplification of cDNA of different cell lines yielded an amplification product of 158 bp (Figure 1d). We verified that the products obtained from different cell lines were identical by Southern blotting hybridization (Figure 1d). To do this, we used a probe derived from a PCR-amplified product and showed that it hybridized on pulsed field electrophoresis blots, to a specific 400 kb *NruI* restriction fragment containing the *ETS-1* locus (data not shown). We concluded from these data that we have amplified an identical product in the different cell lines which is very probably an *ETS-1* transcript.

The sequence of the 158 bp amplification product revealed that the 5' exon amplified from the exon a1 was not the 5' alternative exon as expected but to our surprise, the sequence was identical to exon d of *ETS-1* (Reddy and Rao, 1988) which is 10 kb downstream of exon a1 (Kerckaert *et al.*, 1990; Jorcyk *et al.*, 1991). To explain this result, sequence comparison between primer O27 and exon d was performed. The result demonstrated high homology between O27 and exon d and this homology probably allowed for PCR amplification in the low stringency conditions used. We confirmed that no rearrangement of the *ETS-1* locus had occurred in any of the cell lines used to prepare cDNA; indeed, probes from exon a1 and d show a single band in Southern blot hybridization experiments (data not shown). Similarly we failed to identify any potential pseudogene which may explain this transcript.

We designed PCR experiments using high stringency conditions with a new set of primers derived from sequences within exon d and a1 to determine whether the same products were obtained (Figure 2a). The PCR product of the predicted

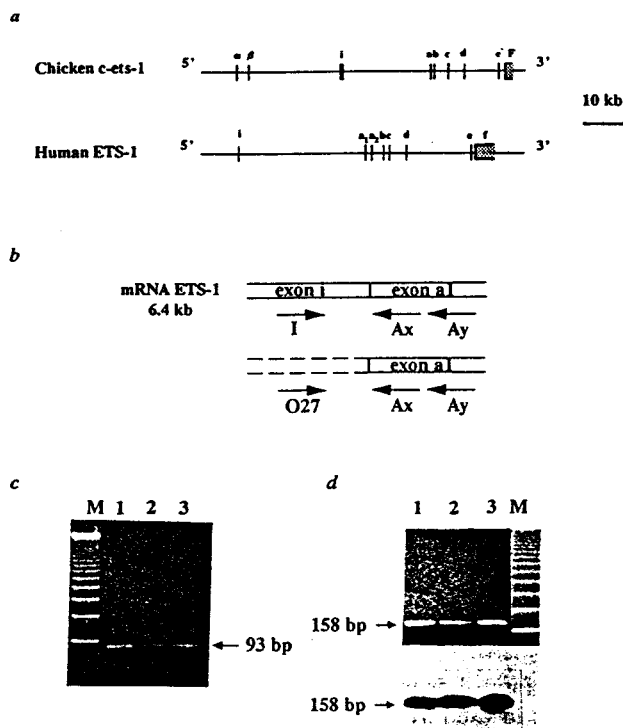


Fig. 1. Two stage PCR experiments to detect 5' alternative exons to exon a1. (a) Comparison of genomic organization of chicken and human *ets-1* loci. (b) The diagram shows the set of primers used in the two-step amplification strategy from total RNA, for control amplifications and for studying 5' alternative exons. The primers I and O27 are homologous to exon i of the human *ETS-1* and to exon β of the chicken *ets-1*, respectively. Primers to exon a1 of *ETS-1* (Ax and Az) are common to both reactions. (c) Amplification results with the primers I; an amplified product of 93 bp is observed. (d) Amplification results with the primers O27. RNAs show an amplified product of 158 bp. (Bottom) Southern blot of the agarose gel above, using the 158 bp amplification product of A431 cell line as a probe (1, CEM; 2, A431. 3, HeLa. M; multimer of 123 bp).

size (242 bp) was found after 30 cycles from the poly(A)⁻ RNA and was not observed with poly(A)⁺ RNA from the three cell lines (Figure 2b). Sequence comparison of this 242 bp amplified product, with genomic clones of exons a1 and d was carried out and showed 100% homology between the PCR product and the genomic exons a1 or d which both end correctly at the exon-intron junction (Figure 2c). This is in agreement with a splicing reaction occurring between splice donor site of exon d and splice acceptor site of exon a1 (i.e. no mutation or base addition was observed, and both cleavage positions of the donor and acceptor splicing sites are correct). We concluded from these data that this product is the result of an abnormal splicing reaction, in which the predicted order of exons is scrambled relative to the genomic order of exons. Additionally, this product is not polyadenylated and consensus splice sites are used.

To rule out the possibility that amplified and cloned sequences were artefacts from PCR, we performed RNase protection assays which allowed quantification of the scrambled transcript with regard to the normal transcript. An antisense RNA probe, that included the scrambled transcript containing exon a1 and d (Figure 3a), was hybridized to total RNA. The RNase-resistant products were analysed by gel electrophoresis. Protected fragments corresponding to the abnormally spliced product (242 bp) and normally spliced products (130 and 122 bp for exon d

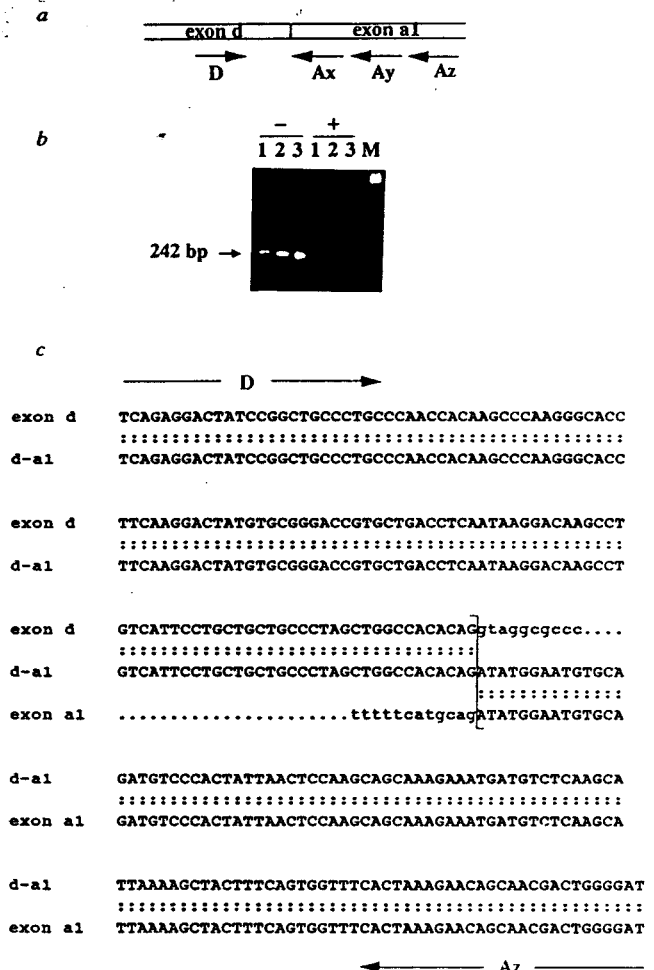


Fig. 2. PCR amplification and sequence comparison of scrambled splicing product with genomic DNA. (a) Diagram shows the new primers D and Az used to amplify the scrambled splicing product. Both primers were totally equivalent to their respective exons. (b) PCR analysis using one-step amplification with primers D and Az, onto poly(A)⁺ (+) and poly(A)⁻ (-) RNA. 1, CEM; 2, A431; 3, HeLa; M, marker. (c) Sequence comparison of the amplification product and genomic sequences. Exon-exon and exon-intron junctions are indicated by a vertical line. Homology is indicated by dots. Intron sequences are in small letters. Arrows correspond to primers.

and a1 respectively), were observed (Figure 3b). This result is the first observation of scrambled exons by an RNase protection experiment. The level of scrambled transcript was estimated to be close to 0.01 the level of the normally spliced transcript.

Recently, similar abnormally spliced transcripts were identified for a candidate tumour suppressor gene, the *DCC* gene, by PCR experiments (Nigro *et al.*, 1991) and the authors also demonstrated that these products are not polyadenylated. Both results clearly establish that scrambling of exons may occur during the splicing process *in vivo*. In both cases, the scrambled exons were joined at the same splice sites, with the same precision as normal splicing (Mount, 1982), except that the exon order was changed. The first step in assembly of the spliceosome involves splice site recognition (Robberson *et al.*, 1990) and is carried out normally, but during scrambled splicing, the subsequent step, which pairs sequential exons appropriately, is executed aberrantly. Aberrant annealing of homologous sequences or

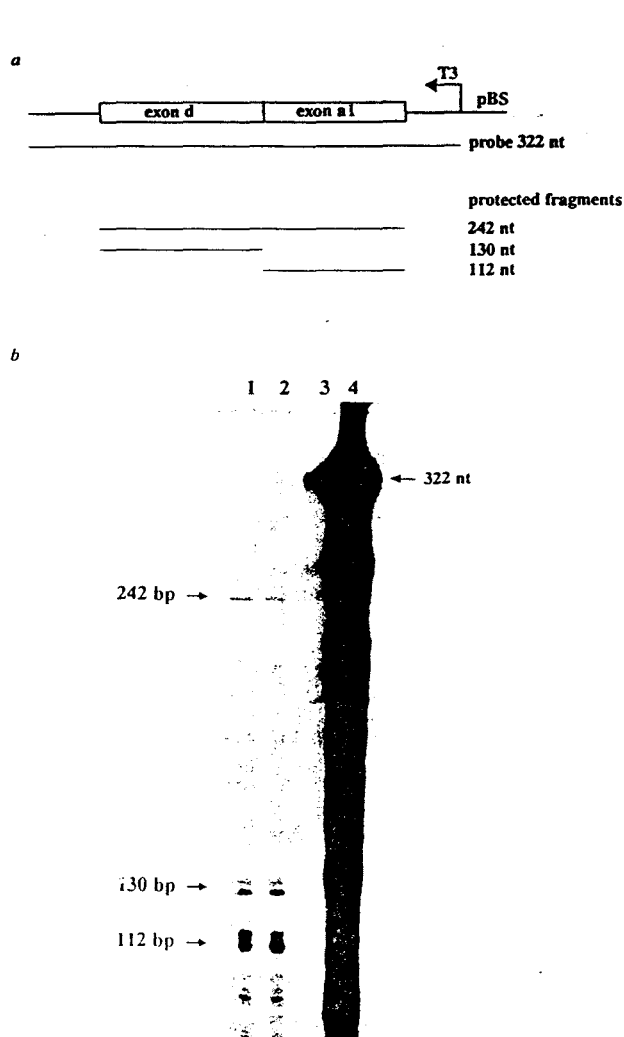


Fig. 3. RNase protection analysis of scrambled splicing product. (a) The top line shows the construct used to generate specific probe. Open boxes indicate exon structure in the scrambled splicing product and the lines correspond to the polylinker of the pBluescript II SK⁻vector. The second line shows the RNase protection probe of 322 nucleotides. The expected protected fragments over the scrambled splicing product, exon a1 and exon d are drawn in the third, fourth and fifth lines, respectively, and the sizes are indicated. (b) Autoradiogram of RNase protection analysis products of 20 μ g total RNA sample from CEM cell line (1–2), 20 μ g tRNA (3) and undigested probe (4). Arrows indicate protected fragments which correspond, from top to bottom, to the scrambled product and the normally spliced exons d and a1.

fortuitous protein interactions within introns leading to a particular secondary structure of the pre-mRNA, might facilitate illegitimate interactions of snRNPs to incorrectly paired exons. A possible mechanism for intramolecular splicing generating abnormally spliced product has been proposed (Nigro *et al.*, 1991) which would generate circular RNAs, but an intermolecular reaction is also possible. *Trans*-splicing mechanisms described in lower eukaryotes (Agabian, 1990) bear no comparison to the scrambled splicing observed here.

Using a combination of sense and antisense primers to the different exons of *ETS-1*, which could allow visualization of scrambled products between donor splice sites of exons c, d, e and the acceptor site of the exons a1, a2, b and c, we performed PCR to determine whether scrambling may be observed with other exons of this locus. We observed

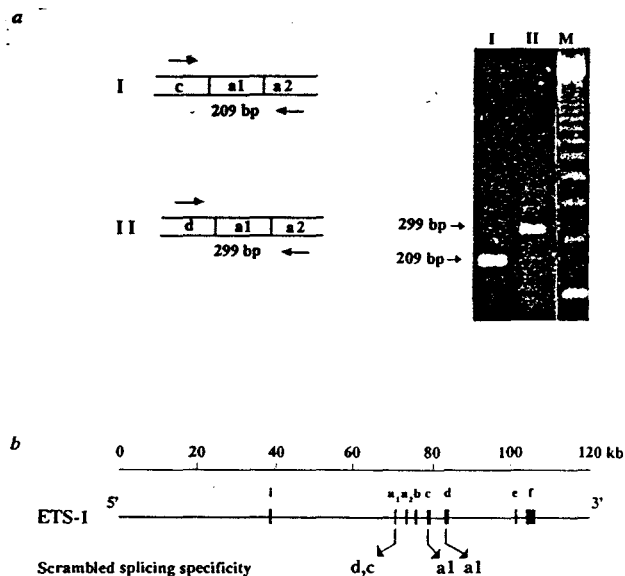


Fig. 4. Specificity of scrambling splicing. (a) Only two scrambled products (I and II) were observed by PCR from CEM total RNA, using a combination of primers which could allow visualization of scrambled products between donor sites of the exons c, d and e and the acceptor splice site of the exons a1, a2, b and c. These products correspond to splicing of exon d or exon c with exon a1 which normally splices 3' with exon a2. M, 123 bp mutimer. (b) Genomic organization of the human *ETS-1* locus and location of the splice sites used by scrambled splicing (see arrows).

only an additional unexpected product of splicing which involved the donor splice site of exon c and, as before, the acceptor splice site of exon a1 (Figure 4a, product I). Both scrambled transcripts I and II were detected using primers to exons a2 (Figure 4a) and in these experiments scrambled products between exon c or d and exon a2 were never observed. Other PCR experiments to detect scrambled splicing between the other exons were unsuccessful. Thus, only the acceptor splice site of exon a1 and the donor splice sites of exons c and d appear to participate in the scrambled splicing mechanism. The specificity observed in the scrambled splicing is represented in Figure 4b. It is noteworthy that the intervening sequences downstream of exon d and upstream of exon a1 are very large: 15 kb and 30 kb respectively (Kerckaert *et al.*, 1990; Jorczyk *et al.*, 1991). In addition, alternative splicing of exon d has been demonstrated in different cell lines (Jorczyk *et al.*, 1991; Reddy and Rao, 1988; Koizumi *et al.*, 1990) and this frequently occurs at a significant level with regard to the normal transcript (J.P. Kerckaert, unpublished data). It is interesting to note that the *DCC* gene also contains large intervening sequences (Fearon *et al.*, 1990). This gene shares sequence homology with the *NCAM* gene (Fearon *et al.*, 1990) which expresses a combination of alternatively spliced forms (Reyes *et al.*, 1991) and it is possible that some exons of *DCC* gene are alternatively spliced. A relationship between scrambled splicing and alternative splicing may exist.

By PCR amplification of aliquots from sucrose gradient separation of total RNA, we observed that the scrambled product (242 bp amplified product as in Figure 2b) is only detected in one or two fractions corresponding to ~2 kb size (data not shown). Although the *ETS-1* pre-mRNA spans over 70 kb (Kerckaert *et al.*, 1990; Jorczyk *et al.*, 1991),

it is unlikely that the signal observed results from degradation of a larger product which should be detected in more than two fractions of the sucrose gradient separation. In addition, the amplification of scrambled products I and II (Figure 4a) demonstrated that normal splicing occurs downstream of the scrambled exons. It may be concluded that the 2 kb molecule containing scrambled exons (d-a1) which came from large pre-mRNAs, has been processed by splicing to eliminate large intervening sequences. To date, we do not know the 5' and 3' limits of the aberrant transcripts. The open reading frame is preserved by joining exon c or d with exon a1 but it is unlikely that this has any biological relevance since in the case of scrambled exons of the *DCC* gene, two out of three products lose the open reading frame (Nigro *et al.*, 1991).

The expression of alternatively spliced forms is regulated according to the cell environment or the stage of development (Breitbart *et al.*, 1987) but relatively little is known about the mechanisms involved. The discovery and the study of scrambled exons may be a means to understanding the mechanisms responsible for alternate process and order in splicing.

Materials and methods

PCR amplification

Total RNA was prepared from different cell lines using the guanidium thiocyanate-caesium chloride method as previously described (Glisin *et al.*, 1974). Poly(A)⁺ and poly(A)⁻ RNA was isolated from total RNA using double passage over oligo(dT)-cellulose chromatography. Poly(A)⁺ RNA constituted 7.5% of the total RNA. 1.5 µg RNA was reverse transcribed in 10 µl containing 200 ng Ay primer, 1 µg BSA, 1.5 mM each dNTP, 50 mM Tris-HCl pH 8.3, 75 mM KCl, 10 mM DTT, 3 mM MgCl₂ and 100 U MuMLV reverse transcriptase (Gibco-BRL) for 1 h at 37°C. Reverse transcription mixture was included in a final volume (100 µl) of PCR reaction containing 100 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 1 U *Taq* polymerase (Stratagene) and 200 ng of 5' primer (I or O27). For second stage of PCR, 1 µl of first step reaction was added to a second PCR mixture containing 200 ng of each primer (Ax and I or Ax and O27). Amplification was performed by 30 cycles: 92°C (1 min), 45 or 65°C (2 min) for the first and second stage of PCR respectively and 72°C (2 min) in a Techne PHC-2 thermocycler. PCR reactions shown in Figures 2 and 4 were performed as previously, but with a 65°C hybridization temperature. Sequences of the primers are: I, 5'-GATCTCAAGCCGACTCTCAC-CATCATC-3'; O27, 5'-TGCTGTGAGGACCCCTGGATGCCATGC-3'; Ax, 5'-TAGTGGGACATCTACACATTCATATC-3'; Ay, 5'-TCCTT-TGCTGCTTGGAGTTAATAGTGGG-3'; Az, 5'-ATCCCCAGTCG-TTGCTGTGTTCTTTAGTG-3'; D, 5'-TCAGAGGACTATCCGG-CTGCCCTGC-3'; exon a2 primer, 5'-GCCCACATCACCCAGTC-CCGAA-3'; and exon c primer, 5'-GTCACCCAGACAACATGTG-3'.

Agarose gel analysis

A 10 µl aliquot of the PCR reaction was electrophoresed on a 2% agarose-TBE gel. The RNA was transferred onto a nylon membrane (Biotrace N) according to the protocol of the manufacturer and was then hybridized with 10⁶ c.p.m. of 158 bp probe. The membrane was autoradiographed overnight using HyperfilmTM MP (Amersham). The 158 bp probe was prepared as follows: 1 µl amplification reaction obtained from A431 RNA (diluted 20 times) was added to a 20 µl PCR mixture as described, with 200 ng of primer Ax and O27, 200 µM dATP, dGTP, dTTP and 6 µl [³²P]dCTP (10 µCi/µl).

Cloning and sequencing

The 242 bp amplification product was cloned into the *Sma*I site of pBluescriptII SK-vector (Stratagene) and sequenced by the dideoxynucleotide method. For genomic sequences, subclones were obtained and sequenced by conventional methods from 10 overlapping recombinant clones which cover 120 kb of the *ETS-1* locus (Kerckaert *et al.*, 1990 and our data).

RNase protection assay

The RNase protection was performed as previously described (Bailleul *et al.*, 1990). The amplification product of 242 bp (Figure 2c) was cloned into

the *Sma*I site of pBluescript II SK-. The plasmid was linearized with *Hind*III and used as a template for T3 RNA polymerase (Stratagene). 20 µg of total RNA was hybridized to 1 × 10⁵ c.p.m. of [³²P]uridine-labelled probe, overnight at 45°C. After RNase A and T1 digestion at 37°C for 30 min, samples were analysed by electrophoresis on a 6% acrylamide-8 M urea gel, and exposed to Cronex X-ray film in the presence of intensifying screens for three days.

Acknowledgements

We thank M.H.Loucheux-Lefebvre for encouragement and support, M.Collyn-d'Hooghe for discussion, S.Galiègue-Zouitina for pulsed field experiments and K.MacLeod for help in manuscript preparation. C.C. was supported by Institut de Recherches sur le Cancer de Lille and Région Nord-Pas de Calais (Conseil Régional) fellowship. This work was supported by the Association pour la Recherche sur le Cancer (ARC).


References

- Aebi, M. and Weissmann, C. (1987) *Trends Genet.*, **3**, 102-107.
 Agabian, N. (1990) *Cell*, **61**, 1157-1160.
 Bailleul, B., Surani, M.A., White, S., Barton, S.C., Brown, K., Blessing, M., Jorcano, J. and Balmain, A. (1990) *Cell*, **62**, 697-708.
 Breitbart, R.E., Andraedis, A. and Nadal-Ginard, B. (1987) *Annu. Rev. Biochem.*, **56**, 467-495.
 Brody, E. and Abelson, J. (1985) *Science*, **228**, 963-967.
 Duterque-Coquillaud, M., Leprince, D., Flourens, A., Henry, C., Ghysdael, J., Debuire, B. and Stehelin, D. (1988) *Oncogene Res.*, **2**, 335-344.
 Fearon, E.R. *et al.* (1990) *Science*, **247**, 49-56.
 Gegonne, A., Leprince, D., Duterque-Coquillaud, M., Vandenbunder, B., Flourens, A., Ghysdael, J., Debuire, B. and Stehelin, D. (1987) *Mol. Cell. Biol.*, **7**, 806-812.
 Glisin, V., Crkvenjakov, R. and Byus, C. (1974) *Biochemistry*, **13**, 2633-2637.
 Jorczyk, C.L., Watson, D.K., Mavrothalassitis, G.J. and Pappas, T. (1991) *Oncogene*, **6**, 523-532.
 Kerckaert, J.P., Duterque-Coquillaud, M., Collyn-d'Hooghe, M., Morel, P., Majerus, M.A., Lai, J.L., Fenaux, P., Batters, F., Debuire, B. and Loucheux-Lefebvre, M.H. (1990) *Leukemia*, **4**, 16-19.
 Koizumi, S., Fisher, R.J., Fujiwara, S., Jorczyk, C., Bhat, N.K., Seth, A. and Pappas, T.S. (1990) *Oncogene*, **5**, 675-681.
 Leprince, D., Gegonne, A., de Taisne, C., Schnurberger, A., Lagrou, C., and Stehelin, D. (1983) *Nature*, **306**, 395-397.
 Leprince, D., Duterque-Coquillaud, M., Li, R.P., Henry, C., Flourens, A., Debuire, B. and Stehelin, D. (1988) *J. Virol.*, **52**, 3233-3241.
 Lewin, B. (1980) *Cell*, **22**, 324-326.
 Mount, S.M. (1982) *Nucleic Acids Res.*, **10**, 459-472.
 Nigro, J.M., Cho, K.R., Fearon, E.R., Kern, S.E., Ruppert, J.M., Oliner, J.D., Kinzler, K.W. and Vogelstein, B. (1991) *Cell*, **64**, 607-613.
 Nunn, M.F., Seeburg, P.P.M., Moscovici, C. and Duesberg, P.H. (1983) *Nature*, **306**, 391-395.
 Reddy, E.S.P. and Rao, V.N. (1988) *Oncogene Res.*, **3**, 239-246.
 Reyes, A.A., Small, S.J. and Akeson, R. (1991) *Mol. Cell. Biol.*, **11**, 1654-1661.
 Roberson, B.L., Cote, G.J. and Berget, S.M. (1990) *Mol. Cell. Biol.*, **10**, 84-94.
 Sharp, P.A. (1987) *Science*, **235**, 766-771.
 Steitz, J.A. (1987) In Birnstiel, M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer-Verlag, pp 71-99.
 Wasylyk, B., Wasylyk, C., Flourens, P., Begue, A., Leprince, D. and Stehelin, D. (1990) *Nature*, **346**, 191-193.
 Watson, D.K., McWilliams-Smith, M.J., Nunn, M.F., Duesberg, P.H., O'Brien, S.J. and Pappas, T.S. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7294-7298.
 Watson, D.K., McWilliams, M.J., Lapis, P., Lautenberger, J.A., Schweinfest, C.W. and Pappas, T.S. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7862-7866.

Received on October 10, 1991; revised on November 25, 1991

MONITOR**Splicing with inverted order of exons occurs proximal to large introns**C. COCQUERELLE *ET AL.**EMBO J.* 11, 1095-1098

Splicing links exons together in a strict 5' to 3' direction. However, there have been two recent examples of exon shuffling during splicing: in the *DCC* (deleted in colorectal carcinomas) gene and, as discussed here, in the human *ETS* gene. The odd couple was discovered by accident. Chicken *c-ets-1* has two alternatively spliced exons, α and β . Using primers to the human $\alpha 1$ exon and to the chicken β exon, Cocquerelle *et al.* tried to amplify a human fragment corresponding to an alternative exon. Sequencing revealed that their fragment was

in fact the d exon from further down the gene linked to the $\alpha 1$ exon; d and β are similar and PCR was carried out at low stringency. The RNA was present in the poly(A)⁻ pool only, and genuine splice sites had been used. It was not a PCR artefact either: RNase protection assays also identified the abnormally spliced species and suggested that it is present at ~1% of the levels of normal transcripts. Attempts to identify other abnormal splices by PCR provided only one other example: a c to $\alpha 1$ join. Of potential significance, the intron before $\alpha 1$ and that after d are huge (30 and 15 kb, respectively). An understanding of this unusual exon scrambling could perhaps help to resolve the mechanism of alternative splicing. 

TIG MAY 1992 VOL. 8 NO. 5

157

B. L'épissage inversé forme des transcrits circulaires.

La démonstration que ces réactions d'épissage forment des molécules d'ARN circulaires a été acceptée pour publication dans le numéro spécial de Janvier 1993 intitulé «The new age of RNA» de la revue FASEB Journal, vol 7 n°1 155-160.

En effet, nous avons poursuivi l'amplification par PCR de ces transcrits, dans la fraction poly A⁻ des ARN totaux et montrés que les exons en amont et en aval des jonctions d'épissage inversé sont normalement épissés dans la molécule d'ARN. Ces transcrits ont une structure circulaire parce que, d'une part il est possible de les observer par PCR à l'aide d'amorces dans un même exon, et d'autre part ces molécules sont présentes dans une fraction de poids moléculaire en accord avec la taille pour une structure circulaire comportant uniquement les exons: c'est à dire 648 et 909 bases.

Nous avons également montré que ce type d'épissage se produit uniquement entre certains exons du locus de ETS-1. La spécificité observée est très instructive: elle suggère que certains critères doivent être nécessaires pour engendrer une réaction d'épissage modifiant l'ordre des exons. Sur cette base, nous pouvons remarquer que les introns à proximité des exons du gène ETS-1 impliqués dans ce type d'épissage sont de très grande taille (supérieure à 15 kb).

Dans ce papier, nous décrivons également la localisation cytoplasmique et la relative stabilité des transcrits circulaires.

Mis-splicing yields circular RNA molecules

CLAUDE COCQUERELLE, BÉNÉDICTE MASCREZ, DOMINIQUE HÉTUIN, AND BERNARD BAILLEUL¹

Unité 124 INSERM, Institut de Recherches sur le Cancer, 59045 Lille Cedex, France

ABSTRACT We previously identified novel human *ets-1* transcripts in which the normal order of exons is inverted, and demonstrated that although the order of exons is different than in the genomic DNA, splicing of these exons out of order occurs in pairs using genuine splice sites (1). Here we determine the structure of these novel transcripts, showing that they correspond to circular RNA molecules containing only exons in genomic order. These transcripts are stable molecules, localized in the cytoplasmic component of the cells. To our knowledge, this is the first case of circular transcripts being processed from nuclear pre-mRNA in eukaryotes. This new type of transcript might represent a novel aspect of gene expression and hold some interesting clues about the splicing mechanism.—Cocquerelle, C., Mascrez, B., Hétuin, D., Bailleul, B. Mis-splicing yields circular RNA molecules. *FASEB J.* 7: 155-160; 1993.

Key Words: circular RNA • *c-ets-1* • mis-splicing • proto-oncogene

SPlicing is a fundamental process of RNA biosynthesis in eukaryotic cells in which excision of intervening sequences is mediated by the spliceosome (2). The two main features of splicing are precision and orderliness to ensure that cleavage and rejoining at the exon-intron boundaries conserve the reading frame and that exon skipping is avoided (3). In the case of multi-intronic genes, or genes with large introns such as the Abelson proto-oncogene (*c-abl*), whose first intervening sequence is at least 200 kb long (4), elements other than known components of the spliceosome must exist to ensure precision and orderliness in removal of such large introns.

In contrast to splicing orderliness, we previously identified new transcripts for the human proto-oncogene *ets-1*, in which pairs of exons are joined at consensus splice sites but in a different order from that of the genomic DNA (1). These transcripts are nonpolyadenylated and are expressed at much lower levels than the normal transcript. Similar observations were described for the DCC gene (deleted in colorectal carcinoma)² in Vogelstein's lab (5).

In this paper we demonstrate that this splicing with inverted order of exons, called mis-splicing, occurs specifically with three splice sites of the *ets-1* locus. In addition, we present evidence that mis-splicing is an intramolecular process leading to circular RNA molecules.

MATERIALS AND METHODS

RNA preparation and sucrose gradient separation

Total RNA from a CEM cell line was prepared using the guanidinium thiocyanate-cesium chloride method as previously described (6). Poly (A)⁻ and poly (A)⁺ RNA was isolated from total RNA by double-passage chromatography over oligo (dT)-cellulose (Pharmacia, Piscataway, N.J.). Poly (A)⁻ RNA (100 µg) was separated by sedimentation through

a 10–20% sucrose linear gradient in 5 mM Tris HCl pH 7.4, 2.5 mM EDTA, 0.5% SDS, and RNA in the 10 collected fractions and precipitated with ethanol. Linear single-stranded riboprobes labeled with [³²P]uridine were generated by T7 transcription of pBluescript SK-vector (Stratagene, San Diego, Calif.).

Preparation of cytoplasmic and nuclear RNA

For cytoplasmic RNA, cytoplasm was prepared using ribonucleoside-vanadyl complexes, and RNA was isolated by a variation of the classic phenol extraction procedure as previously described (7). Nuclei were prepared by hypotonic lysis. Cells were collected by centrifugation at 4°C and washed several times with PBS (2.16 g/l Na₂HPO₄, 7 H₂O, 0.2 g/l KH₂PO₄, 8 g/l NaCl, and 0.2 g/l KCl). Cells were resuspended in modified RSB (10 mM Tris-HCl at pH 7.9, 10 mM NaCl, and 5 mM MgCl₂) in a Dounce homogenizer at a concentration of 3–5 × 10⁷ cells/ml and swollen in ice for 5–10 min. Cells were broken with 10–15 strokes of the pestle and KCl was added to 100 mM. The nuclei were recovered by centrifugation at 1000 × g for 3 min at 4°C. Lysis of nuclei was in high-salt solutions, followed by DNase treatment to reduce viscosity, after which RNA was extracted with hot phenol.

PCR experiments and sequencing

PCR experiments were performed as previously described (1) with oligonucleotides spanning *ets-1* cDNA positions (8) 353-380, 446-472 (a1 exon antisense), 508-529, 510-534 (a2 exon antisense and sense), 709-730 (b exon antisense), 781-801 (c exon antisense), 828-847, 969-998 (c exon sense), 1093-1108 (d exon antisense), and 1140-1164, 1232-1256 (d exon sense). Primers at position 446-472 (a1 exon), 1232-1256 (d exon), and 828-847 (c exon) were used to observe mis-splicing products in nuclear fractions and during actinomycin D treatments. The amplified products through c-a1 and d-a1 mis-splicing junctions were cloned into the SmaI site of the pBluescript SK-vector (Stratagene) and sequenced by the dideoxynucleotide method. For genomic sequences, subclones were obtained and sequenced by conventional methods from recombinant clones containing the *ets-1* locus (9).

RESULTS

Mis-splicing yields stable molecules in cytoplasm

The observation described previously (1) that mis-splicing occurs in the *ets-1* locus using the acceptor splice site of a1

¹To whom correspondence should be addressed, at: Unité 124 INSERM, Institut de Recherches sur le Cancer, Place de Verdun, 59045 Lille Cedex, France.

²Abbreviation: DCC, deleted in colorectal carcinoma gene.

exon and both c and d donor splice sites is represented in Fig. 1a. PCR amplifications were performed in an attempt to determine the cellular localization of these transcripts. Nuclear and cytoplasmic RNA was prepared from the CEM cell line and analyzed for mis-spliced products by PCR amplification. We found that c-a1 and d-a1 mis-spliced transcripts are both contained in the cytoplasm (Fig. 1b), even though these transcripts are not polyadenylated (1). To deter-

mine the relative stability of these transcripts, we treated CEM cells with actinomycin D for up to 48 h. Although the results of PCR experiments are not completely quantitative, the fact that both mis-spliced transcripts are still observed in the RNA of cells after 48 h of actinomycin D treatment argues that the half-lives of these molecules are probably high (Fig. 1c). Indeed, these transcripts are expressed at low levels (1), and a short half-life would rapidly lead to a undetectable

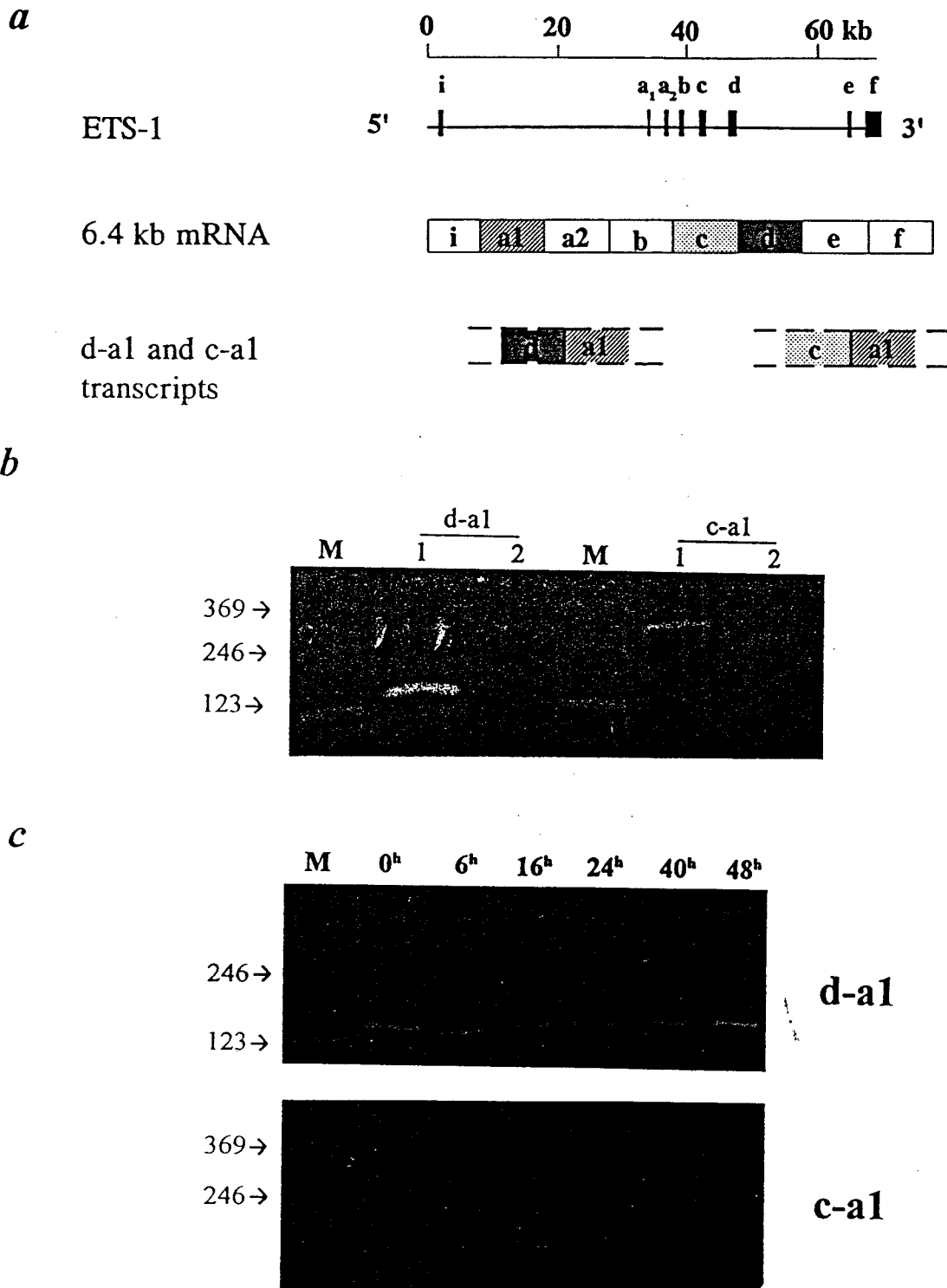


Figure 1. a) Genomic organization of the human *ets-1* locus showing the structures of the normal *ets-1* transcript and the two mis-splicing products (d-a1 and c-a1). b) PCR experiments demonstrating localization of the both d-a1 and c-a1 mis-spliced transcripts; 1. Cytoplasmic 2. nuclear RNA c) PCR experiments demonstrating the high stability of both d-a1 and c-a1 mis-spliced transcripts.

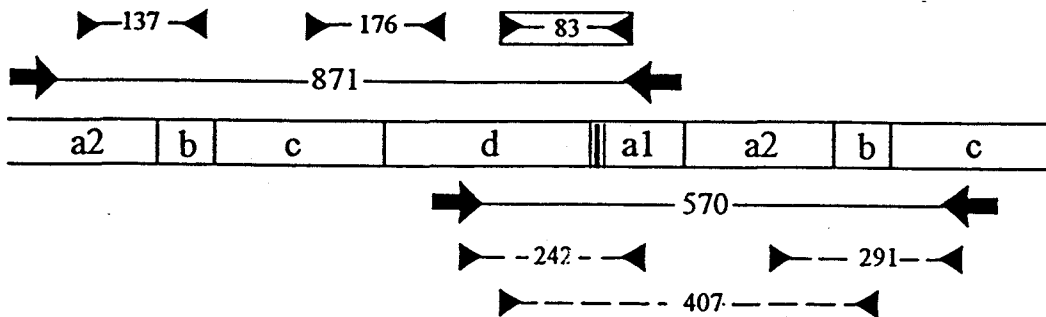
signal after 24 h of actinomycin D treatment. Therefore, we may conclude that the mis-spliced transcripts are relatively stable molecules.

Mis-splicing occurs at specific splice sites

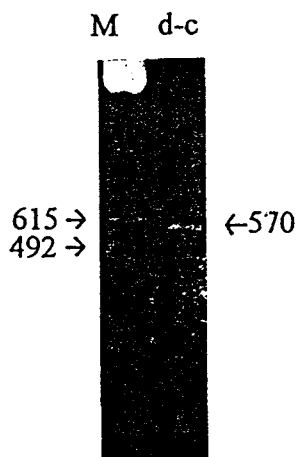
A central question is whether all splice sites of any gene may participate in mis-splicing. Therefore, we determined

whether mis-splicing might be observed with all exons of the *ets-1* locus. Mis-splicing has been described for transcripts of DCC (5) as well as *ets-1* (1), and these transcripts (found in poly (A)⁺ RNA) are expressed at much lower levels than the normal transcripts (1, 5). To overcome this technical barrier, we partially concentrated the mis-spliced transcripts by sucrose gradient separation from poly (A)⁺ RNA, aiming to

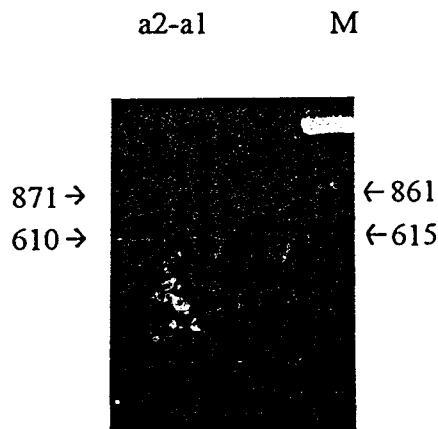
a



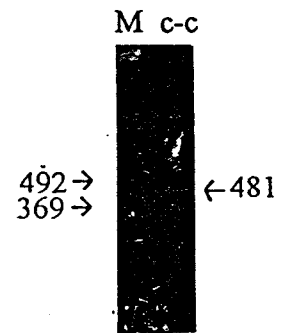
b



c



d



e

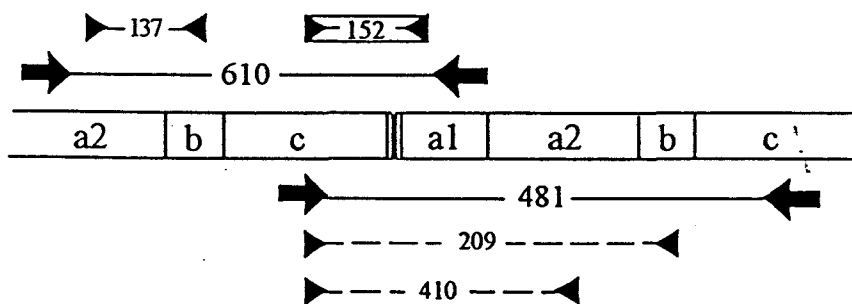


Figure 2. Mis-splicing in transcripts of the *ets-1* locus. Detection of mis-spliced products by PCR amplification. *a*) PCR amplification of RNA molecules containing splice junctions with inverted order of d and a1 exons is represented by large arrows. The content of the sequence was confirmed by nested PCR (arrowheads and dashed lines) using sets of primers inside the PCR products. *b*) PCR amplification downstream of the d-a1 mis-spliced product using d and c primers yields a 570-bp ampligene. *c*) PCR amplification upstream of both mis-spliced products using primer a2 and a1 generates two bands, 871 and 610 bp, for mis-spliced products d-a1 and c-a1, respectively. *d*) PCR amplification downstream of the c-a1 mis-spliced product using two opposite strand c primers yields a 481-bp ampligene. *e*) PCR amplification of RNA molecules (large arrows) containing splice junction with inverted order of c and a1 exons and nested PCR (arrowhead). The sizes of the amplification products were determined by comparison with a multimer of 123 bp. The mis-splicings are indicated by a particular mark [|||]. The products of both nested PCRs through the mis-spliced junctions indicated by boxes have been cloned and sequenced.

reduce nonspecific PCR amplifications. Sucrose gradient-sedimented poly (A)⁻ RNA was collected in 10 fractions, which were tested by PCR to detect mis-splicing products.

Using a combination of sense and antisense primers to the different exons of *ets-1* locus to allow visualization of mis-splicing, we performed PCR experiments with partially purified fractions. We observed that the amplifications using primers described in Fig. 2a and Fig. 2e reveal amplification products characteristic in length of transcripts containing exons in genomic order upstream and downstream of either c-a1 or d-a1 mis-spliced junctions. The sequence content of the PCR products was determined by nested PCR. Indeed, other sets of primers located inside the PCR products was used to reamplify it (see Figs. 2a and 2e). For example, the 871 bp PCR product was reamplified by three sets of primers: in a2 and b exons, c and d exons, and d and a1 exons to yield the expected length PCR products of length 137, 176, and 83 bp respectively. In addition, sequencing of the d-a1 and c-a1 junctions was performed to ensure that they correspond to use of the genuine splice sites. Figure 3 shows the sequence of mis-spliced products d-a1 and c-a1 and the comparison of mis-spliced junctions with genomic sequences, confirming that the exon junctions of mis-spliced RNAs correspond to a splicing reaction that involved the genuine donor and acceptor splice sites. Indeed, the d-a1 and c-a1 mis-spliced junctions correspond to the use of the a1 acceptor splice site (uuuucaugcag/AU), the d exon donor splice site (CAG/guaggc), and the c exon donor splice site (CUG/guaggu), respectively. These sequences match precisely the consensus splice sequences (2) and correspond to splice sites used in the maturation of the normal transcripts (10).

The observation of a unique PCR product for each amplification described in Figs. 2b and 2d, and only two PCR products for Fig. 2c, clearly excludes the existence of transcripts with mis-spliced junctions other than c-a1 and d-a1 products. For example, the d exon to c exon amplification does not reveal bands shorter than 570 bp (Fig. 1a and 1b), excluding the existence of splicing between the donor splice site of d exon and acceptor splice sites of a2, b, or c exons. Indeed, mis-splicing using these splice sites would yield smaller PCR products; i.e., d-a2-b-c, d-b-c, and d-c. In this case, it is appropriate to interpret a negative result. Indeed, we argue that the PCR conditions are obviously correct, because we observe in the reaction mixture a PCR product of the expected size but not a smaller one, whereas both PCR products should be generated from the same set of primers. All PCR experiments to detect mis-splicing using the donor splice site of e exon were unsuccessful. Both i and f exons were excluded from the mis-splicing analysis because they correspond to the first and last exons of the locus, respectively (10, 11). Consequently, mis-splicing occurs with only

exon d	TGCTGCTGCCCTAGCTGGCCACACAG	taggcgccc
d-a1	TGCTGCTGCCCTAGCTGGCCACACAG	TATGGAATGTGCAGATGTCCCACTA
exon a1	tttttcatgcag	TATGGAATGTGCAGATGTCCCACTA
c-a1	CATGTGCATGGGGAGGACCACTGCTG	TATGGAATGTGCAGATGTCCCACTA
exon c	CATGTGCATGGGGAGGACCACTGCTG	taggtcata

Figure 3. Sequence comparison of c-a1 and c-a1 mis-spliced products with genomic DNA. Exon-exon and exon-intron are indicated by the vertical line. Identity is indicated by dots and intron sequences by small letters.

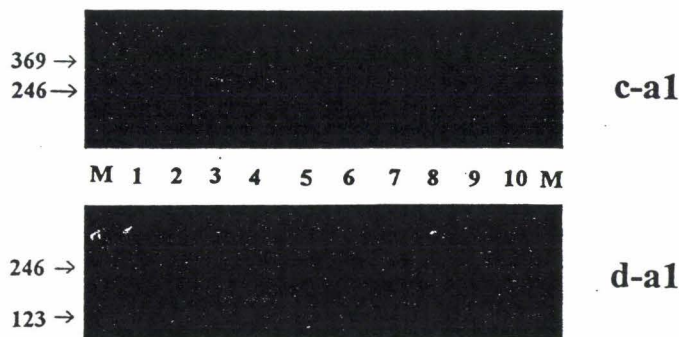


Figure 4. Amplification of c-a1 and d-a1 products from a sucrose gradient of CEM poly (A)⁻ RNA. The 10 fractions collected from the top (n°1) of each gradient are indicated. M, multimer of 123 bp.

particular splice sites in the *ets-1* transcript, i.e., the acceptor site of a1 exon and the donor sites of exons c and d.

Ets-1 mis-splicing products are circular molecules

We have demonstrated that the mis-spliced products contain normal splice junctions upstream and downstream the site of mis-splicing, i.e., a2, b, and c exons are observed upstream and downstream of the mis-splicing junctions (Figs. 2b and 2e), and these exons are localized between the a1 exon and d exon in the genomic organization. In addition, PCR in both directions from the c exon was able to amplify a 481-bp fragment (Fig. 2e) corresponding to c-a1-a2-b-c product, suggesting that the mis-spliced transcript could be circular. To ensure that this molecule does not contain sequences other than a1 to c exons, we analyzed the size of this molecule by sucrose gradient centrifugation.

Mis-spliced species were observed in the low molecular weight fraction of poly (A)⁻ RNA. Indeed, the set of primers into c and a1 exons allowed the visualization of the c-a1 mis-spliced transcript in fractions 2 to 4 from the top of the sucrose gradient of poly (A)⁻ RNA. Similarly, the d-a1 mis-spliced product was observed in fractions 3 to 5 (Fig. 4). In addition, we determined the sedimentation of 300- and 900-base [³²P]labeled linear riboprobes in the sucrose gradient. The 900-base RNA probe comigrated with the d-a1 mis-spliced product in fractions 3 to 5, whereas the 300-base RNA was localized in fractions 1 and 2. So it is likely that the transcripts contained only one of each exon (a1 exon to c or d exon) in a circular form. Indeed, the sums of the length of exons from a1 to c or from a1 to d are 648 and 909 bases, respectively (8, 10). In addition, this PCR amplification is observed solely in sucrose gradient fractions 2 to 4, arguing that a multimeric structure does not exist. These data are in agreement with a circular form formed by an intramolecular splicing mechanism (Fig. 5).

DISCUSSION

In this paper we propose that the *ets-1* mis-spliced products are circular RNA molecules containing only normally spliced exons, with the exception of the circulating event. In addition, similar preliminary experiments with DCC mis-spliced products show that they are also small-sized molecules in which normal spliced junctions are observed (data not shown). Because the introns of DCC are very large (12; and B. Vogelstein, personal communication), it is likely that the mis-spliced products from DCC are also circular molecules containing only exonic sequences. The cytoplasm-

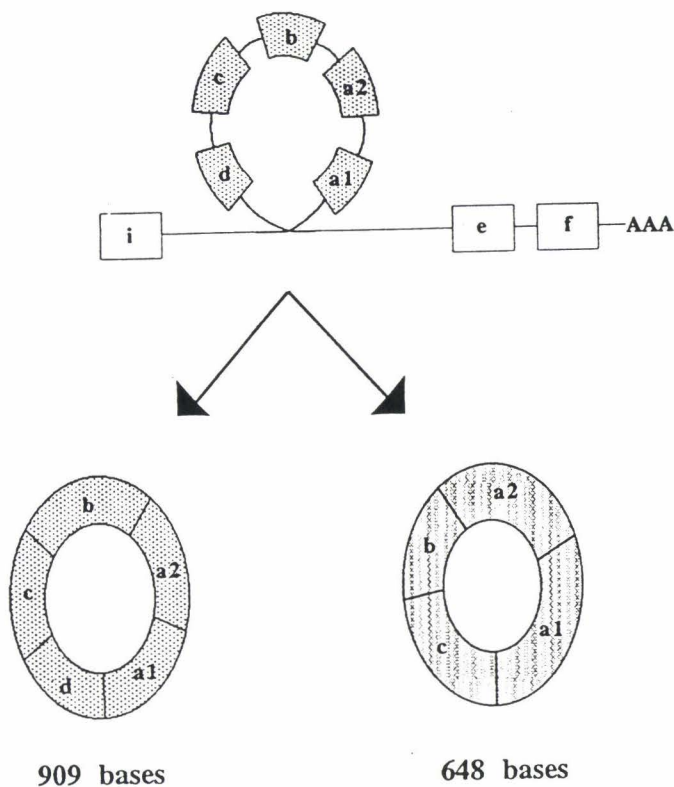


Figure 5. Schematic representation of the structure of both circular mis-splicing products (d-a1 and c-a1). The loop structure of the fully immature pre-mRNA is hypothetical. The sizes of the circular molecules are indicated.

mic localisation of *ets-1* and also of DCC (5) mis-spliced molecules is in agreement with the lack of intervening sequences in these molecules. Indeed, splicing factors are known to prevent pre-mRNA transport into the cytoplasm (13). In the present case, the poly(A) tail is not required for transport to the cytoplasm. In addition, the relatively high stability of these molecule lacking 3' and 5' ends could be explained by their protection against any exonuclease activity. Direct evidence of these molecules in Northern blots is difficult to obtain because they are expressed at a very low level (1).

We may consider different hypotheses to explain the mis-splicing specificity (i.e., acceptor splice site of a1 exon and donor splice sites of c and d exons). We may suppose that a particular sequence conformation around these splice sites could allow the splicing with inverted order of exons, as already described for alternative splicing (14). If there is a sequence downstream of the d exon that favors or allows mis-splicing between d exon and a1 exon, the same sequence could participate in mis-splicing between c and a1 exons by d exon skipping, which has been observed during normal pre-mRNA maturation at a relatively high frequency (8, 10). If such sequences do exist, they must have little effect on the 5' to 3' direction splicing. Indeed, the 5' to 3' splicing reaction occurs efficiently in *ets-1* transcript of the locus. A more attractive hypothesis to account for the observed mis-splicing specificity is that the mis-splicing reaction occurs at random, but that in the nucleus the steady-state levels of the intermediate in which both intervening sequences (including d exon) are not spliced is higher compared with other intermediates. Consequently, the acceptor splice sites of a1 exon and donor splice sites of c and d exons could represent the main targets of the mis-splicing reaction. Indeed, the inter-

vening sequences downstream of the d exon and upstream of the a1 exon are very large; 15 and 30 kb, respectively (9, 10), whereas the other intervening sequences are less than 3 kb. In support of this argument, long intervening sequences (15) and intervening sequences containing alternative exons (16) have been suggested as determining factors for the order of intron splicing. Indeed, these exons seem to be spliced last because a more difficult recognition of the intron, and consequently a less rapid excision, probably occur. This convincing explanation of the mis-splicing specificity does not require the existence of unlikely RNA signals in the introns.

Consequently, the mis-splicing mechanism is mainly an intramolecular process. It is unlikely that intermolecular splicing between two different *ets-1* transcripts occurs because amplifications of mis-splicing products are not observed with high-molecular-weight RNA, either poly(A)⁻ or poly(A)⁺. The basis for this intramolecular specificity is currently unclear.

Several questions arise: What kind of molecular components determine the physical proximity of splice sites to allow the circularizing splicing? Are the elements involved in the physical proximity of the splice sites components of the spliceosome (17)? Do these circular RNAs have a biological significance or effect? It is not possible to answer these questions at present, but it is noteworthy that mis-splicing is not restricted to the cell line used here. Indeed DCC mis-spliced products have been identified in rodent and human cells, in both normal and neoplastic cell types, and in different tissues (5). It is tempting to think that the splicing mechanism yielding circular RNAs simply represents errors in the normal splicing process. But occasionally errors generate biological effects in cells, as in mutations generated by DNA replication (18), chromosomal breakages by recombinase in B or T cells (19), and oncogene transduction by retrovirus (20).

Evidence presented herein indicates unusual features at the junctions of mis-spliced *ets-1* exons leading to circular RNA molecules. We suggest that these observations are only the visible part of an iceberg. The most basic questions about the molecular mechanism of nuclear pre-mRNA splicing remain unanswered. Indeed, we have only a limited understanding of how splice sites are chosen and none as to how they are juxtaposed with orderliness. In this context, mis-splicing is an interesting model from which to gain insight into the splicing mechanism. [FJ]

We thank members of the laboratory for helpful discussions, M. H. Loucheux-Lefebvre for encouragement, and support and staff services for invaluable help. C. C. was supported by Institut de Recherches sur le Cancer de Lille and Région Nord-Pas de Calais (Conseil Régional) fellowship. This work was supported by the Association pour la Recherche sur le Cancer and the comité du Nord and du Pas-de-Calais de la ligue contre le Cancer.

REFERENCES

- Cocquerelle, C., Daubersies, P., Majérus, M.-A., Kerckaert, J.-P., and Bailleul, B. (1992) Splicing with inverted order of exons occurs proximal to large introns. *EMBO J.* 11, 1095-1098
- Green, M. R. (1991) Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annu. Rev. Cell Biol.* 7, 559-599
- Aebi, M., and Weissmann, C. (1987) Precision and orderliness in splicing. *Trends Genet.* 3, 102-107
- Bernards, A., Rubin, C. M., Westbrook, C. A., Paskind, M., and Baltimore, D. (1987) The first intron in the human *c-abl* gene is at least 200 kilobases long and is a target for translocations in chronic myelogenous leukemia. *Mol. Cell. Biol.* 7, 3231-3236

5. Nigro, J. M., Cho, K. R., Fearon, E. R., Kern, S. E., Ruppert, J. M., Oliner, J. D., Kinzler, K. W., and Vogelstein, B. (1991) Scrambled exons. *Cell* **64**, 607-613
6. Glisin, V., Crkvenjakov, R., and Byus, C. (1974) Ribonucleic acid isolated by cesium chloride centrifugation. *Biochemistry* **13**, 2633-2637
7. Berger, S. L., and Chirgwin, J. M. (1989) Isolation of RNA. *Methods Enzymol.* **180**, 3-10
8. Reddy, E. S. P., and Rao, V. N. (1988) Structure expression and alternative splicing of the human c-ets-1 proto-oncogene. *Oncogene Res.* **3**, 239-246
9. Kerckaert, J. P., Duterque-Coquillaud, M., Collyn-d'Hooghe, M., Morel, P., Majérus, M. A., Lai, J. L., Fenaux, P., Bauters, F., Debuire, B., and Loucheux-Lefebvre, M. H. (1990) Polymorphism of the proto-oncogene ETS-1 in hematological malignancies. *Leukemia* **4**, 16-19
10. Jorcyk, C. L., Watson, D. K., Mavrothalassitis, G. J., and Papas, T. (1991) The human ETS-1 gene: genomic structure, promoter characterization and alternative splicing. *Oncogene* **6**, 523-532
11. Majérus, M.-A., Bibollet-Ruche, F., Telliez, J.-B., Wasylyk, B., and Bailleul, B. (1992) Serum, AP-1 and ets-1 stimulate the human ets-1 promoter. *Nucleic Acids Res.* **20**, 2699-2703
12. Fearon, E. R., Cho, K. R., Nigro, J. M., Kern, S. E., Simons, J. W., Ruppert, J. M., Hamilton, S. R., Preisinger, A. C., Thomas, J., Kinzler, K. W., and Vogelstein, B. (1990) Identification of a chromosome 18q gene that is altered in colorectal cancers. *Science* **247**, 49-56
13. Legrain, P., and Rosbach, M. (1989) Some *cis* and *trans* acting mutants for splicing target pre-mRNA to the cytoplasm. *Cell* **58**, 573-583
14. Clouet-d'Orval, B., d'Aubenton-Carafan, Y., Marie, J., and Brody, E. (1991) Determination of an RNA structure involved in splicing inhibition of a muscle-specific exon. *J. Mol. Biol.* **221**, 837-856
15. Gattoni, R., Keohavong, P., and Stévenin, J. (1986) Splicing of the E2A premessenger RNA of adenovirus serotype 2. *J. Mol. Biol.* **187**, 379-397
16. Bovenberg, R. A. L., Van de Meerendonk, W. P. M., Baas, P. D., Steenbergh, P. H., Leeps, C. J. M., and Jansz, H. S. (1986) Model for alternative RNA processing in human calcitonin gene expression. *Nucleic Acid Res.* **14**, 8785-8803
17. Guthrie, C. (1992) Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science* **253**, 157-163
18. Kunkel, T. A., and Alexander, P. S. (1986) The base substitution fidelity of eucaryotic DNA polymerases. *J. Mol. Biol.* **261**, 160-166
19. Haluska, F. G., Tsujimoto, Y., and Croce, C. M. (1987) Mechanisms of chromosome translocation in B and T-cell neoplasia. *Trends Genet.* **3**, 11-15
20. Bishop, J. M. (1990) Nobel lecture. Retroviruses and oncogenes II. *Biosci. Rep.* **10**, 473-491

Received for publication October 1, 1992.

Accepted for publication November 11, 1992.

C. La spécificité de formation des transcrits circulaires.

Les résultats que nous présentons ici sont soumis à publication dans *Molecular and Cellular Biology*. Ils confirment par une preuve directe la structure circulaire de ces transcrits. En effet, nous avons visualisé à la taille attendue, l'un des transcrits circulaires sur un Northern-blot en gel d'acrylamide. De plus, le retard de migration à haut pourcentage d'acrylamide de cette molécule est caractéristique d'une structure circulaire. Nous avons poursuivi l'étude de la spécificité d'épissage inversé, en abordant une hypothèse expliquant la formation spécifique de certains transcrits qui tient compte de la chronologie de l'épissage des introns. L'étude des formes immatures d'épissage dans le noyau a été effectuée par protection aux ribonucléases et montre que les réactions d'épissage inversé ont lieu avec un transcrit immature dans le noyau qui n'est pas une espèce prédominante.

L'expression de l'un des transcrits circulaires a été analysée dans plusieurs types de lignées humaines et comparé à celle du transcrit linéaire. Cette étude montre que l'expression de ce transcrit circulaire est faible par rapport au transcrit linéaire. Nous observons des variations de l'expression du transcrit circulaire, mais elles reflètent en fait le taux de transcription du locus plutôt que des variations de fréquence dans la réaction d'épissage inversé.

CIRCULAR RNA TRANSCRIPTS FORMATION BY SPLICING

MARIELLE CAVROIS, CLAUDE COCQUERELLE, DOMINIQUE HETUIN and
BERNARD BAILLEUL.

Unité 124 INSERM, Institut de Recherches sur le Cancer de
Lille, Place de Verdun, 59045 Lille Cedex, France.

Running title; Circular RNA yielded by mis-splicing.

Keywords; circular RNA molecule/ mis-splicing/ oncogene/ ETS1.

* To whom correspondence should be addressed (Telephone (33) 20 52
97 00, FAX (33) 20 52 70 83).

ABSTRACT

We have previously identified novel transcripts of the human *c-ets-1* gene in which the normal order of exons is inverted and demonstrated that while the order of exons is different from the genomic DNA, splicing of these exons in aberrant order (called mis-splicing) occurs in pairs using genuine splice sites albeit at low level. Here, we unambiguously confirm the circular structure of these transcripts by Northern blots experiments onto acrylamide gels and PCR experiments. This data show that mis-splicing yields to circular RNA molecules containing only exons sequences. To investigate the circular transcripts formation, we analysed the expression and the stability of circular transcripts versus linear transcripts in different cell lines. We found that during the splicing process of the *ETS1* transcript about 1% of the pre-mRNA is spliced in reverse order and that variations on circular transcript level observed in different cells reflect the level of transcription of the *c-ets-1* locus. In addition, to elucidate the mis-splicing specificity observed in *ets-1* locus, we determined within the nucleus the level of the different precursor mRNAs. The results disagree with the initially proposed model involving a higher steady-state level of the intermediate pre-mRNA in which the mis-splice sites are present.

INTRODUCTION

Splicing is the process by which introns are removed from precursor messenger RNA (pre-mRNA) molecules and exons are joined to form mature mRNA (15). Since many pre-mRNA molecules contain more than one intron, mechanisms have been postulated which ensure that cognate 5' and 3' splice sites are appropriately used such that aberrant splicing does not occur (1). Nevertheless, junctions of exons in nongenic order have been described in cells for DCC (Deleted in Colorectal Carcinoma) (21) and ETS1 (8). The DCC and ETS1 variant transcripts have been detected by RT-PCR in nonpolyadenylated RNA, from the cytoplasmic component of the cells. RNase protection assays of ETS1 variant transcript confirmed that they were not PCR artefacts (8). In both variant transcripts, the exon junctions in nongenic order result from splicing reactions at low level occurring at genuine splice sites which are utilised during normal maturation. PCR data suggest that in each variant transcript, one exon junction in nongenic order is present and that splicing of the other ETS1 exons occurs in genic order (9). Consequently we prefer call this reaction, mis-splicing rather than scrambled exons as originally termed (21). From this PCR data, we proposed that the ETS1 mis-spliced products are circular RNA molecules resulting from an intramolecular reaction (9). To date, mis-splicing of ETS1 transcripts has been observed at the acceptor splice site of the a1 exon and the donor splice sites of c and d exons. Of potential significance, introns

near these mis-spliced sites are huge (30 and 15 kb) whereas the other intervening sequences of c-ets-1 locus are less than 3kb (18,29).

Here, we report the analysis of the major ETS1 mis-spliced transcript by extensive PCR and Northern blot experiments which confirms the circular structure of the transcript. We attempted to determine within the nucleus, the level of the precursors of the major mis-spliced products to elucidate the mis-splicing specificity of the c-ets-1 locus. Finally, we used different cell lines to compare the formation level of circular RNAs to linear transcripts.

MATERIALS AND METHODS

RNA preparation. Total RNA was prepared using the guanidinium thiocyanate cesium chloride method (13). Nuclear RNA was prepared as already described (3). Poly (A)⁻ and poly (A)⁺ RNA was isolated from total RNA by double passage chromatography over oligo (dT)-cellulose (Pharmacia Biotech S.A. France). Poly (A)⁻ RNA (100 µg) was separated by sedimentation through a 10-20% sucrose linear gradient in 5mM Tris HCl pH 7.4, 2.5 mM EDTA, 0.5 % SDS. The 6 out of 30 fractions positive for d-a1 mis-splicing PCR experiments were pooled, ethanol precipitated and solved in water to obtain partially purified fractions.

Northern blots onto acrylamide gels. Concentrated d-a1 mis-splicing products from 400 µg RNA Poly (A)⁻ were electrophoresed in the same conditions of buffer (8M urea, 88mM Tris borate pH 8.3, 2.5 mM EDTA) and electrophoresis (34 mA, 55W, 6 hours) for both 4 and 5 % acrylamide gels. Following electrophoresis, the urea was removed by soaking the gel in 500 ml of 7% formaldehyde, 25mM Sodium phosphate (pH 6.5) at room temperature, 15 min. The RNA was electroblotted onto a nitrocellulose filter (Genscreen from New England Nuclear) in 25 mM sodium phosphate (pH 6.5) using an

electroblotting apparatus from Biorad, at 200mA, 4°C, overnight. Hybridization with antisense riboprobe corresponding to the d-a1 mis-splicing junction (8) was carried out as indicated by manufacturer. Amersham-HP film were exposed overnight to membranes in the presence of intensifying screen.

PCR Experiments. The amplification products were obtained by RT-PCR from sucrose gradient separation fractions of poly (A)⁻ RNA. PCR experiments were performed as previously described (9) with oligonucleotides spanning ETS1 cDNA position (17,25) ; 353-380, 446-472 (a1 exon antisense), 508-529, 510-534 (a2 exon antisense and sense), 709-730 (b exon antisense), 781-801, 969-998 (c exon antisense and sense), 1093-1108 (d exon antisense) and 1140-1164, 1232-1256 (d exon sense).

Cloning and sequencing. For genomic sequences, subclones were obtained and sequenced by conventional methods from 10 overlapping recombinant clones which cover 120KB of the c-ets-1 locus (18 and unpublished data). Exon e amplification product was performed by PCR using to intron primer; AGGAGTTTTCTCGGATCCTTCCTC and GTAATGATTCACAATGGC. The amplified products containing c-a1 and d-a1 mis-splicing junctions or e exon were cloned into the SmaI site of the pBluescript SK-vector (Stratagene) and the sequences were performed by dideoxynucleotide method.

RNase protection assay. The ³²P labelled riboprobe and the RNase protection (figure 3 to 5) were performed as previously described (8). The plasmid containing e exon fragment was linearized with Cla I and used as a template for T3 RNA polymerase (Stratagene). From the riboprobe synthesis, the ³²P Uridine specific radioactivities were 60Ci/mole and 1000Ci/mole for total and nuclear RNA respectively. 100 µg of total RNA or 33µg of nuclear RNA was hybridized to 10⁵ c.p.m of Uridine ³²P labelled probe, overnight at 45°C. After RNase A and T1 digestion at 30°C for 30 minutes, samples were analyzed by electrophoresis on a 6 % acrylamide 8 M urea gel, and exposed to Amersham-HP film in the presence of intensifying screens for three days.

RESULTS

Circular RNA revealed by Northern blot experiments. To confirm the circular structure of the ETS1 mis-spliced products, we analyzed poly (A)⁻ RNA from CEM cells, by Northern blot analysis using polyacrylamide gels and a specific probe. To eliminate the normal transcripts and to concentrate the mis-spliced transcripts, we performed sucrose gradient separation from CEM poly (A)⁻ RNA (See materials and methods). A ³²P labelled probe complementary to the mis-spliced d-a1 product (see figure 1) was used to reveal the RNA species onto 4 and 5% polyacrylamide gel (figure 1). As expected, a product of roughly 900 bases was observed (figure 1). Indeed, the sum of the ETS1 exons from a1 to d is 909 bases (17,25). Moreover, although both experiments were performed in the same conditions of electric field and time, we observed a greater shift between RNA species and ³²P labelled 902 bases linear marker in the 5% polyacrylamide gel than in the 4% gel. The aberrant migration behavior of the molecule in the high percentage polyacrylamide gel is characteristic of circular molecules like lariat RNA structures (14,26).

RT-PCR around the circular ETS1 transcripts. RT-PCR using sets of primers to visualize circular products (antisense and sense primers from a1 and a2 exons respectively (fig 2a) or antisense c exon and sense c and d exons (fig 2b)), revealed amplification products containing d-a1 and c-a1 mis-spliced junctions. The order of the exons in the amplified products was checked by nested PCR amplification experiments (Open boxes; fig 2). In addition, the sequences of d-a1 and c-a1 mis-spliced junctions (CACACAG/ATATGGAA and AGTCGTG/ATATGGAA respectively) confirm the use of splice sites

as in the normal transcript maturation (12,23). We thought that the complete sequence of circular transcripts was not necessary after results of nested PCRs and the identical a1 to a2 exon junction sequence from mis-spliced molecule (CCAAAG/ATCCCC) to normal transcripts. The 871 bp and 570 bp RT-PCR products correspond to the circular d-a1 molecule, whereas the 610 bp and 481 bp products correspond to circular c-a1 molecule. The 481 bp amplification in which both primers are in opposite direction from c exon, clearly confirms the circular structure of the mis-spliced products.

Mis-splicing specificity. The RT-PCR from a2 to a1 exons through the c-a1 or d-a1 mis-spliced junction does not reveal shorter bands than the 610 and 871 bp fragments (fig 2a). Consequently the existence of splicing between the acceptor splice site of a1 exon and the donor splice sites of a2 and b exons can be excluded. Similarly, RT-PCR shown in figure 2b, excludes the existence of splicing between the donor splice site of c or d exons and acceptor sites of a2, b and c exons. All PCR experiments used to detect mis-splicing of the donor splice site of e exon were unsuccessful. Mis-splicing of i and f exons can also be excluded since they correspond to the first and last exon of the locus, respectively (17,20). In conclusion, mis-splicing occurs preferentially with the acceptor splice site of a1 exon and the donor splice sites of c and d exons.

Steady-state levels of nuclear ETS1 RNA in CEM cells. Mis-splicing events may occur at random between splice sites, but in the nucleus the steady-state level of an intermediate mRNA in which intervening sequences adjacent to mis-splice sites are not rapidly

excised, would be higher as compared to other intermediates. To understand the mis-splicing specificity, we analyzed the level of intermediate ETS1 mRNA in nucleus. The temporal splicing order of the large intron between d and e exons and the 1 kb intron between exon e and f was determined. An RNase protection was performed with a 291 bases uniformly labelled probe carrying the complete e exon and adjacent parts of upstream and downstream introns. A protection of 123 bases was generated which corresponds to completely spliced ETS1 RNA in CEM cells (figure 3). Additionally, three types of c-ETS1 precursor RNA molecules leading to bands of 155, 179 and 209 bases are predicted to be present in the nucleus. The larger fragment corresponding to ETS1 RNA containing both introns, and the smaller one corresponding to c-ETS1 RNA carrying an unspliced 5' intron and spliced intron between e and f exons could not be detected. However a medium fragment (179 bases) was visible (figure 3) which corresponds to ETS1 RNA carrying a downstream unspliced intron and spliced intron of 15kb. This result disagrees with a «random temporal splicing model» to account for mis-spliced specificity, I.e., c-ETS1 RNA is rapidly spliced at small introns, leading to a high steady-state level of intermediate RNAs in which both large intervening sequences are not spliced. In these RNA molecules, the mis-spliced sites are still present; i.e., acceptor site of exon a1 and donor sites of exon d. Consequently, other factors may be responsible of mis-splicing specificity observed in ETS1 maturation.

Circular RNA levels in cell lines. Expression of d-a1 circular transcript in cells from different origin was performed using RNase protection experiments that allowed quantification of mis-spliced

RNA with regard to the normal transcript. A uniformly 322 bases labelled probe (figure 4a and 4b lane 2) carrying the d-a1 mis-spliced junction was hybridized to total RNA. The RNase-resistant fragments were analyzed by polyacrylamide gel electrophoresis. Protected fragments corresponding to the d-a1 mis-spliced product (242 bases) and normally spliced transcripts (130 and 122 bases for d and a1 exons respectively) were clearly observed using 20 μ g of total RNA from CEM, HSB2, Jurkat and MDA cells (figure 4, lanes 4 to 7). The protected fragments were undetectable in K562 cells (Data not shown) and visible only as faint bands, in HeLa cells (lane 8) although 100 μ g were used during hybridization. Additional protection was performed using 100 μ g of total RNA from HeLa cells following 48 hour's serum starvation and a 3 hour's serum stimulation. In these conditions, both signals from normal and circular transcripts are stronger (lane 9). The high expression of ETS1 in CEM, HSB2, Jurkat and MDA cells is as previously described (10,27) and the serum stimulation of ETS1 transcription in HeLa cells are in agreement with a previous study on c-ets-1 promoter stimulation by serum (20). In the described experiments, we analyzed the ratio between the d-a1 circular transcript and the normal linear transcript. A comparison of the 242 bases protected signal and protected signals from 130 to 112 bases indicates that approximately 1% of the ETS1 RNA is circular in each of the cell types tested, despite different levels of transcription of the c-ets-1 locus, in each case.

In addition, similar RNase protection experiments were performed to analyze the stability of both transcripts in CEM cells following actinomycin D treatment (Figure 5). We observed during

actinomycin D treatment that the signals corresponding to normal and circular transcripts are both coordinately reduced and the relative ratio of these transcripts is maintained. After control of the RNA integrity was performed by agarose gels (Data not shown) we concluded that both transcripts have a similar stability in CEM cytoplasm despite the obvious difference of structure. Indeed the linear transcript is 6400 bases long and presents modifications associated with the ends of messenger whereas the circular transcript is 906 bases long with no end.

Similar stability of circular and linear transcripts in CEM cells indicates that the steady-state level of these transcripts corresponds to splicing formation in nucleus.

DISCUSSION

Splicing errors yields circular RNA molecules. The Northern blot experiments and RT-PCR data (figure 1 and 2) unambiguously demonstrate that pre-mRNA can mis-splice to produce circular RNA molecules containing only exon sequences. It is noteworthy that DCC and ETS1 mis-spliced junctions were identified in rodent and human cells in both normal and neoplastic cell types and in different types of cells (8,21). It is tempting to think that the splicing which yields ETS1 circular RNA simply represents an error in the normal splicing process. Indeed, in cells from different origins, a low percentage of ETS1 transcript mis-splices and when variations are observed (figure 4), they are more likely to reflect the level of transcription of the c-ets-1 locus than an increase in mis-splicing reactions. Thus, systematically, during the

splicing process of the ETS1 transcript, about 1% of the pre-mRNA is spliced in reverse order. Because screening of circular transcripts expression has not been performed in adult and embryonic tissues, we cannot exclude that the ETS1 circular RNA may be required by cells during a differentiation process.

What kinds of molecular components ensure the physical proximity of splice sites in a strict 5' to 3' direction? Spliceosome assembly is directed in part by highly conserved pre-mRNA sequences at the 5' and 3' splice sites and the branch site (28). However, the components and the mechanism that ensure assembly in a strict 5' to 3' direction are unclear, although different models have been proposed (1).

Mis-splicing specificity. Here, we showed by RT-PCR that mis-splicing occurs specifically at three exons of the *c-ets-1* locus. Indeed, the acceptor splice site of a1 exon and the donor splice sites of d and c exons are solely implicated in this mechanism. The proposal that mis-splicing occurs at random between different splice sites, but in the nucleus the steady-state level of the mis-spliced intermediate mRNA (I.e., a pre-mRNA in which adjacent introns to mis-splice sites are not excised) would be higher as compared to other intermediates, was analyzed to explain ETS1 mis-splicing specificity. Evidence in favor of this hypothesis are that large intervening sequences (12) and intervening sequences containing alternative exons (4) are proposed as less efficient excised introns. Indeed, intervening sequences in the DCC gene are large (11), as are introns next to acceptor splice site of a1 exon and donor splice site of d exon (30 and 15 kb respectively) (18,28). In addition, skipping of d exon has been observed during normal

maturation at relatively high frequency (17,25). Surprisingly, RNase protection experiments to study the temporal splicing order of introns adjacent to e exon reveal that the 15kb intron between d and e exons is more efficiently excised than the 1 kb intron between e and f exons. Thus, the existence of other factors explaining the mis-splicing specificity may be postulated such as RNA signals favoring the physical proximity of splice site. Recently, inverted splicing has been postulated as a mechanism of alternative splicing to delete the d exon from certain ETS1 mRNA and errors in this process result in mis-splicing (16). At this point, it is noteworthy to describe results recently obtained in sry expression in adult testis. Sry is the mammalian sex determining master regulator gene (19). The critical site of sry action is the genital ridge but Sry is expressed at high level in the adult testis where to date, no function has been determined. The structure of the sry transcripts has been demonstrated to be circular in testis as a result of a splicing process (7). The authors propose that the peculiar inverted repeat structure of the Sry locus could be involved in the generation of circular transcripts. This inverted repeat in the transcript could allow formation of a stem loop structure facilitating the splicing process. Interestingly, the genital ridge transcript is linear probably by use of alternative promoter (7). These results strongly support a biological significance of the circulating splicing mechanism and circular RNAs.

Biological significance of circular RNA. The circular structure of transcripts could prevent its translation, without affecting its putative function as a RNA molecule. Indeed Sry

circular transcripts are not associated with polysomes (7). Some transcripts do not contain any significant conserved ORFs and thus do not appear to encode a protein, suggesting that they may function as a structural RNA. For example, the XIST gene locus is transcribed only from the inactive X chromosome, suggesting that it may play a role in the inactivation process (6). The same is true for H19 gene which is highly expressed during fetal development, but does not appear to encode a protein and may play a role in chromosome condensation (2,5). In addition, data on differentiated skeletal muscle cells suggest that 3'UTR of certain differentiation-specific RNAs are trans-acting regulators in a feedback loop that inhibits cell division and promotes differentiation (24). We emphasize that reactions catalyzed by the spliceosome complex or by ribosomes demonstrate that RNA molecules are not just passive carriers of genetic information but may play active catalytic roles (22,23). We postulate that an RNA function of the ETS1 circular transcripts may be required by cells despite its low abundance and the simultaneous presence of the protein-encoding linear transcript.

Characterization of additional circular molecules in eukaryotic cells should gain insight into the biological significance of this splicing mechanism and of circular RNAs containing exons.

ACKNOWLEDGEMENTS

We thank M.H. Loucheux-Lefebvre for encouragement and support, J. Marie for useful discussions and Northern blot experiment contribution, Ken McElreavey for help in manuscript

preparation and M-C Duvieuxbourg for photographic work. CC was supported by Institut de Recherches sur le Cancer de Lille and Région Nord-Pas de Calais (Conseil Régional) fellowship.

This work was supported by the Association pour la Recherche sur le Cancer (ARC) and the comité du Nord and du Pas-de-Calais de la ligue contre le Cancer.

LITERATURE CITED

1. **Aebi, M., and C. Weissmann.** 1987. Precision and orderliness in splicing. *Trends in Genetics* **3**:102-107.
2. **Bartolomei. MS., S. Zemel and SM. Tilghman.** 1991. Parental imprinting of the mouse H19 gene. *Nature (London)* **351**:153-155.
3. **Berger, S.L. and J.M. Chirgwin.** 1989. Isolation of RNA. *Methods Enzymol.* **180**:3-10.
4. **Bovenberg, R. A. L., W. P. M. Van de Meerendonk, P. D. Baas, P. H. Steenbergh, C. J. M. Leeps, and H. S. Jansz.** 1986. Model for alternative RNA processing in human calcitonin gene expression. *Nucleic Acid Res.* **14**:8785-8803.
5. **Brannan, C., E. Dees, R. Ingram and S. Tilghman.** 1990. The product of H19 gene may function as an RNA. *Mol. Cell. Biol.* **10**:28-36.

6. Brockdorff, N., A. Ashworth, G. Kay, V. McCabe, D. Norris, P. Copper, S. Swift and S. Rastan. 1992. The product of the mouse Xist gene is a 15kb inactive X-specific transcript containing no ORF and located in the nucleus. *Cell* 71:515-526.

7. Capel, B., A. Swain, S. Nicolis, A. Hacker, M. Walter, P. Koopman, P. Goodfellow and R. Loweell-Badge. 1993. Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell* 73:1019-1030.

8. Cocquerelle, C., P. Daubersies, M.A. Majérus, J.P. Kerckaert and B. Bailleul. 1992. Splicing with inverted order of exons occurs proximal large introns. *EMBO J.* 11:1095-1098.

9. Cocquerelle, C., B. Mascrez, D. Héтуin and B. Bailleul. 1993. Mis-splicing yields circular RNA molecules. *FASEB J.* 7:155-160.

10. Collyn d'Hooge, M., S. Galiegue-Zoutina, D. Szymiczek, D. Lantoine, S. Quief, MH. Loucheux-Lefebvre and JP. Kerkaert. 1993. Quantitative and qualitative variation of ETS-1 transcripts in hematologic malignancies. *Leukemia* 7: in press.

11. Fearon, E. R., K.R. Cho, J.M. Nigro, S.E. Kern, J.W. Simons, J.M. Ruppert, S.R. Hamilton, A.C. Preisinger, J. Thomas, K.W. Kinzler, and B. Vogelstein. 1990. Identification of a chromosome 18q gene that is altered in colorectal cancers. *Science* 247:49-56.

12. **Gattoni, R., P. Keohavong, and J. Stévenin.** 1986. Splicing of the E2A premessenger RNA of Adenovirus serotype 2, Multiple pathways in spite of excision of the entire large intron. *J. Mol. Biol.* **187**:379-397.

13. **Glisin, V., R Crkvenjakov and C. Byus.** 1974. Ribonucleic acid isolated by cesium chloride centrifugation. *Biochemistry* **13**:2633-2637.

14. **Grabowski, P. J., R. A. Padgett, and P. A. Sharp.** 1984. Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate. *Cell* **37**:415-427.

15. **Green, M. R.** 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA Splicing. *Annu. Rev. Cell Biol.* **7**:559-599.

16. **Jarrell, K.A.** 1993. Inverse splicing of group II intron. *Proc. Natl. Acad. Sci. USA* **90**:8624-8627.

17. **Jorcyk, C. J., D.K. Watson, G.J. Mavrothalassitis, and T. Papas.** 1991. The human ETS-1 gene: genomic structure, promoter characterization and alternative splicing. *Oncogene* **6**:523-532.

18. **Kerckaert, J. P., M. Duterque-Coquillaud, M. Collyn-d'Hooghe, P. Morel, M.A. Majérus, J.L. Lai, P. Fenaux, F. Bauters, B. Debuire, and M.H. Loucheux-Lefebvre.** 1990. Polymorphism of the

proto-Oncogene ETS-1 in hematological malignancies. *Leukemia* **4**:16-19.

19. **McElreavey. K., E. Vilain, C. Cotinot ,E. Payen and M. fellous.** Control of Sex determination in Mammals. *Eur J. Biochem.* in press.

20. **Majérus, M-A., F. Bibollet-Ruche, J. B. Telliez, B. Wasylyk, and B. Bailleul.** 1992. Serum, AP-1 and Ets-1 stimulate the human ets-1 promoter. *Nucleic Acids Res.* **20**:2699-2703.

21. **Nigro, J. M., K. R. Cho, E. R. Fearon, S. E. Kern, J. M. Ruppert, J. D. Oliner, K. W. Kinzler, and B. Vogelstein.** 1991. Scrambled exons. *Cell* **64**:607-613.

22. **Noller, H. F., V. Hoffarth, L. Zimniak.** 1992. Unusual Resistance of Peptidyl Transferase to Protein Extraction Procedures. *Science* **256**:1416-1420.

23. **Piccirilli, J. A., T. S. McConnell, A. J. Zaug, H. F. Noller, and T. R. Cech.** 1992. Aminoacyl Esterase Activity of the *Tetrahymena* Ribozyme. *Science* **256**:1420-1424.

24. **Rastinejad. F. and H. Blau.** 1993. Genetic complementation reveals a novel regulatory role for 3' untranslated regions in growth and differentiation. *Cell* **72**:903-917

25. Reddy, E. S. P., and V.N. Rao. 1988. Structure, Expression and Alternative Splicing of the Human *c-ets-1* Proto-oncogene. *Oncogene Res.* 3:239-246.

26. Ruskin, B., A. R. Krainer, T. Maniatis, and M. R. Green. 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* 38:317-331.

27. Sacchi, N., A. DeKlein, S.D. Showalter, G. Bigi and T.S. Papas. 1988. High expression of *ets-1* gene in Human thymocytes and immature T leukemic cells. *Leukemia* 2:12-18.

28. Wassarman, D.A., and J. A. Steitz. 1992. Interactions of small nuclear RNA's with Precursor Messenger RNA during in vitro Splicing. *Science* 257:1918-1925.

29. Watson, D.K., M. J. McWilliams, and T. S. Papas. 1988. Molecular organization of the Chicken *ets* locus. *Virology* 164:99-105.

LEGENDS

Figure 1. A) Genomic organization of the human *ets-1* locus, the 6.4kb mRNA, the d-a1 mis-spliced transcript and the ^{32}P labelled riboprobe as previously used for RNase protection experiments (8) are indicated. B) Northern blots from partially purified fractions of RNA poly (A)- isolated from CEM cells using a ^{32}P labelled d-a1 probe. M= linear single strand riboprobe and RNA = RNA sample. The acrylamide concentration of the electrophoresis blots are indicated in percentage.

Figure 2. Detection of mis-spliced products by RT-PCR amplifications. *Left*, PCR amplifications results of RNA molecules containing mis-spliced junction using sense a2 and antisense a1 primers (A) or sense d or c primers and c antisense primer (B). The 610bp and 481bp amplicon correspond to c-a1 mis-spliced products; Δ d indicates the absence of d exon in the amplicons. The set of primers used in the RT-PCR are indicated by arrows. The mis-splicing are indicated by a particular mark ([]). Nested PCR indicated by open boxes were performed to ensure the order of the exons. *Right*, The amplification products (a2-a1, d-c and c-c) and the 123 bp ladder marker (M) are indicated by arrows on agarose gels.

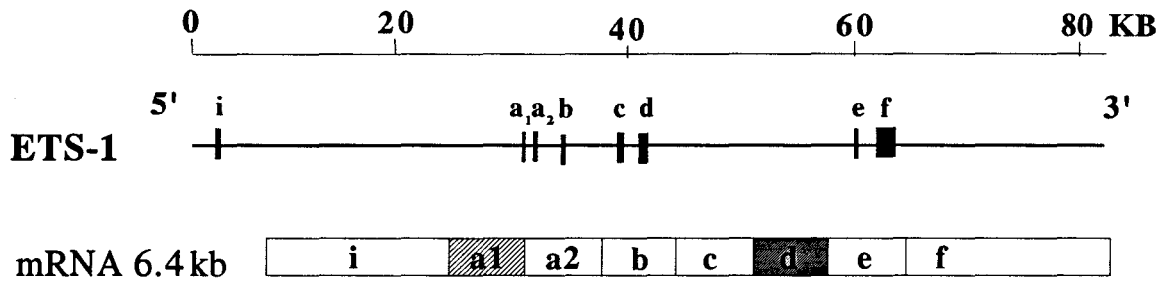
Figure 3. Temporal order of intron adjacent to e exon splicing by RNase protection analysis. a) the pBS construct generating specific probe spanning e exon, the riboprobe and the possible protected fragments are indicated. b) Autoradiogram of RNase protection products of 33 μg CEM nuclear RNA (lanes 3 and 5) or 33 μg of *Saccharomyces cerevisiae* tRNA (lanes 2 and 6). In addition to

the expected 123 nt protected fragment (see arrows on lanes 3 and 5), we observed a 179 nt protected fragment which corresponds to ETS1 RNA carrying the downstream unspliced intron and upstream spliced intron. Lane 1 and 4 are the 291 nt riboprobe and the molecular weight standart (Hinf I digested Ψ X174) respectively.

Figure 4. Linear and d-a1 circular transcript expression in cell lines by RNase protection analysis. a) the pBS construct generating d-a1 riboprobe, the RNase protection probe and the expected protected fragments are indicated. b) Autoradiogram of RNase protection analysis products of 20 μ g (lanes 4 to 7) or 100 μ g (lanes 8 and 9) total RNA : lane 4 to 6 from CEM, HSB2 and Jurkat cells respectively are T cell lines, lane 7. MDA cells and lanes 8 and 9. Hela cells are epithelial cell lines. The RNA preparations were performed with cells in exponential growth (lanes 4 to 8) and with Hela cells following 48 hours serum starvation and 3 hours serum stimulation (lane 9). The lane 1 to 3 are the molecular weight standart, the 322 nt ARN uniformly labelled probe and Rnase protection with 100 μ g of *Saccharomyces cerevisiae* tRNA, respectively. The arrows indicate the mis-spliced transcript (242 nt) and the d exon and a1 exon protections (130 and 112 nt). Bands between 130 and 112 nt are likely to correspond to uncorrect digestions of the hybrids.

Figure 5. Linear and d-a1 circular transcript stability in CEM cell lines by RNase protection analysis. Lanes 1 to 3 are as indicated in figure 4. Total RNA were prepared from CEM cells following 0, 3, 6, 9 and 12 hours (lanes 4 to 8) of actinomycin D treatment (4 μ g/ml).

A)



Mis spliced transcript d-a1.
Riboprobe.



B)

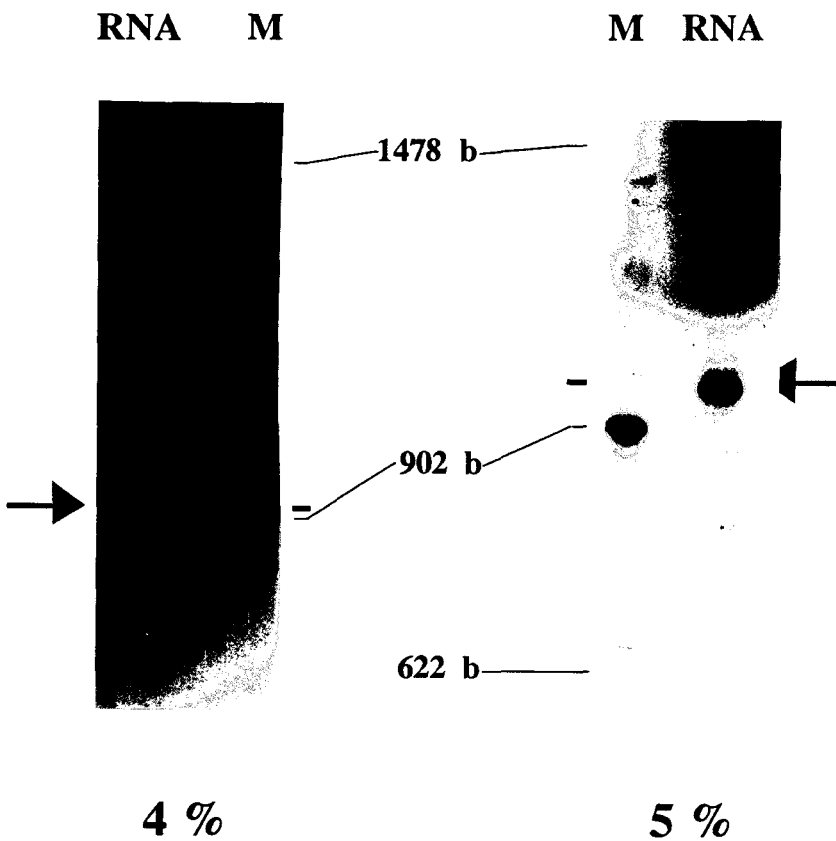


FIGURE 1, Cavrois et al.

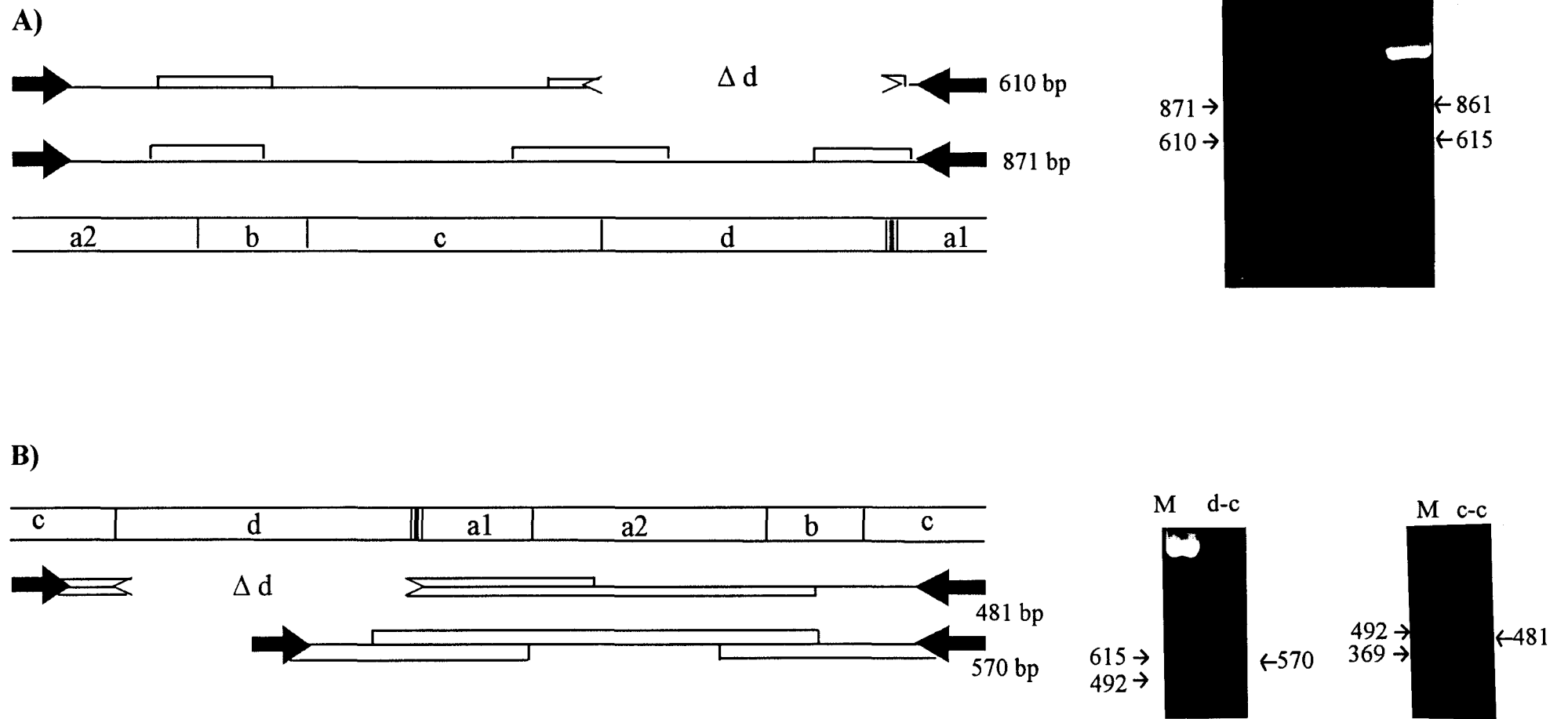
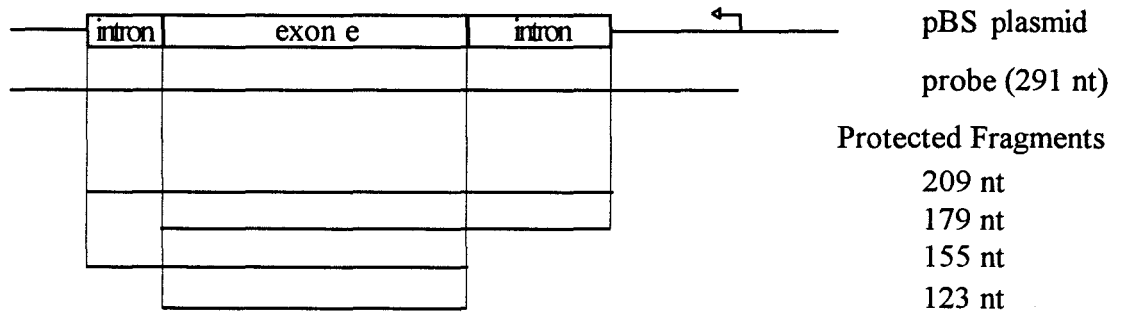


FIGURE 2, Cavrois et al.

a



b

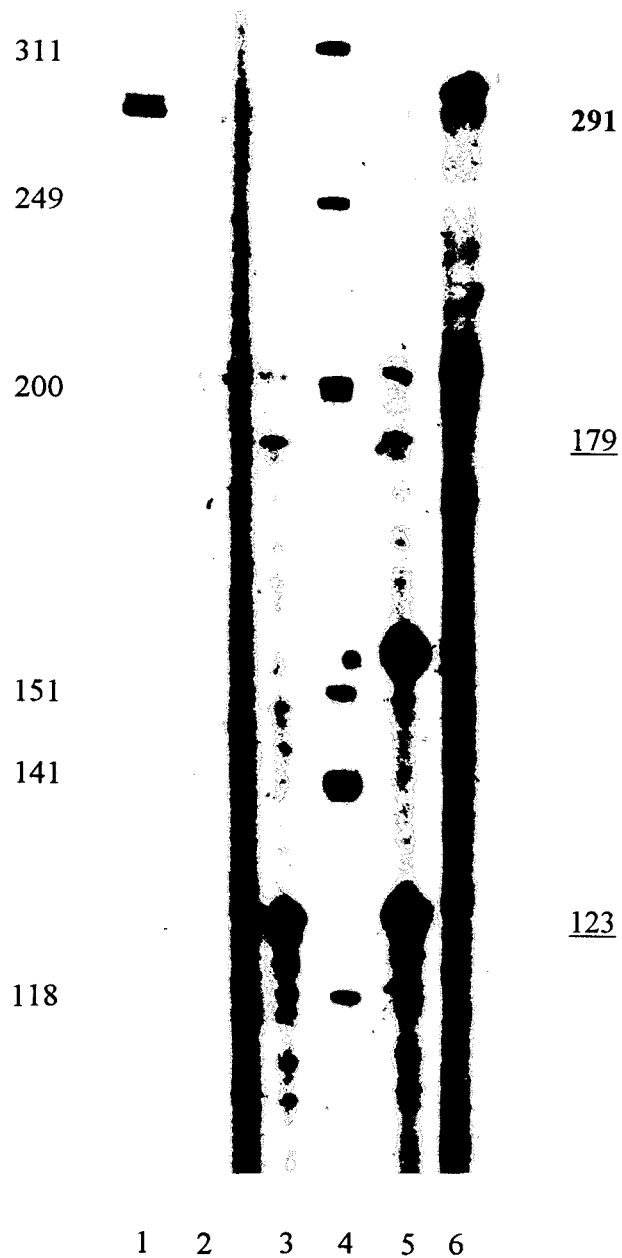


FIGURE 3, Cavrois et al.

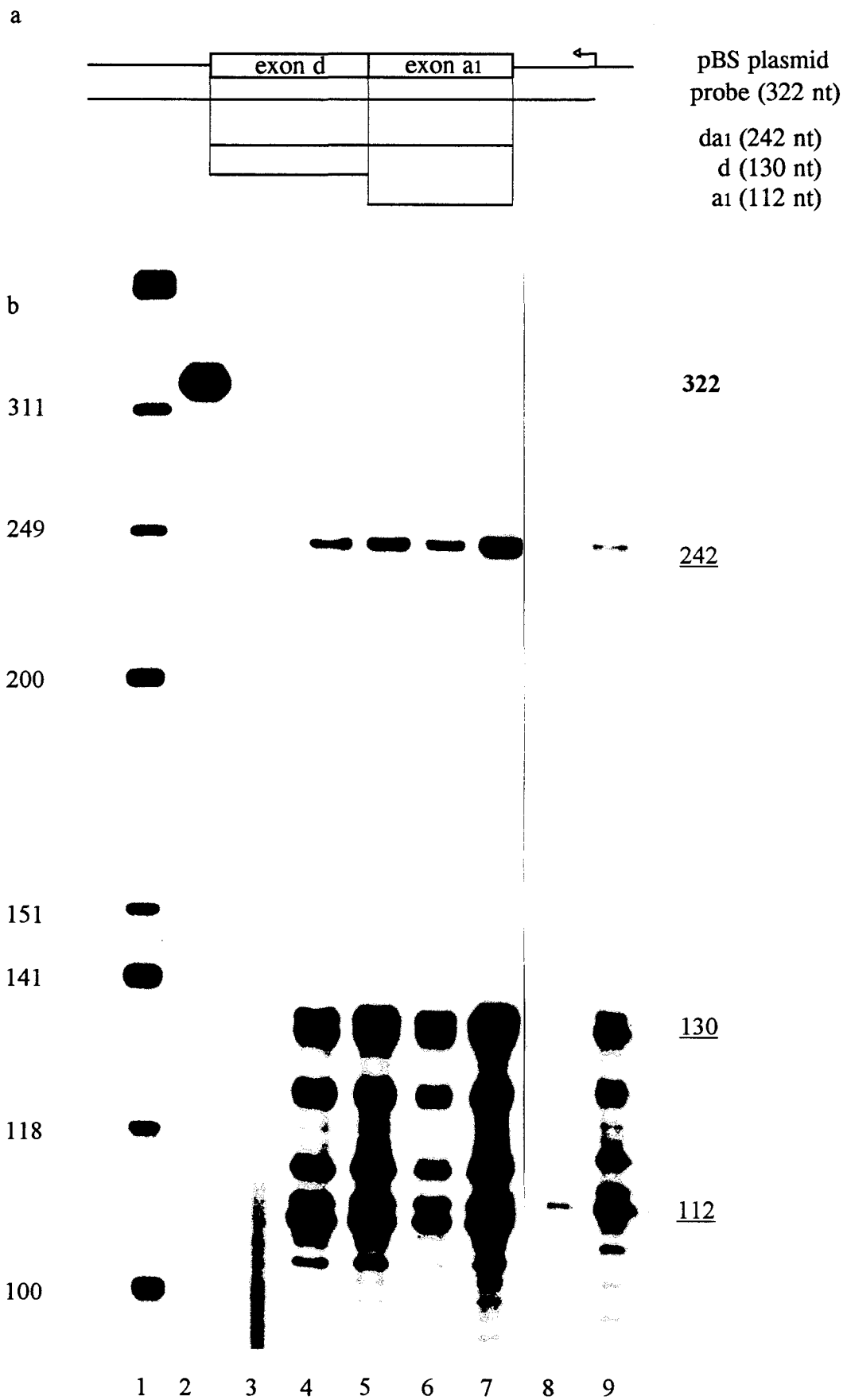


FIGURE 4, Cavrois et al.

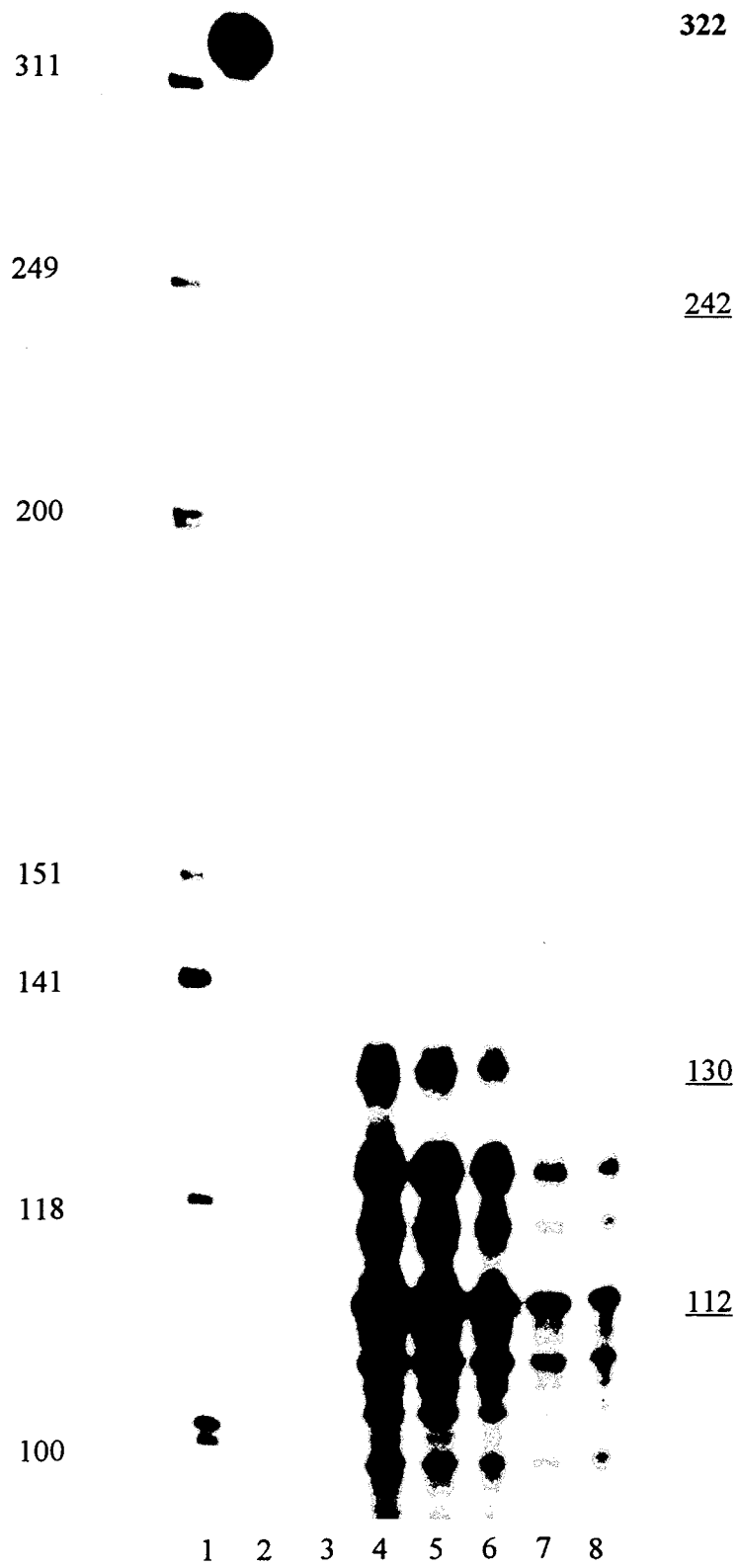


FIGURE 5, Cavrois et al.

DISCUSSION

Au cours de notre travail de Thèse, nous avons mis en évidence la formation, par un mécanisme d'épissage, de transcrits circulaires pour le proto-oncogène humain ETS-1. Une réaction d'épissage inversé intramoléculaire entre le site donneur d'épissage d'un exon en 3' et le site accepteur d'épissage d'un exon en 5' permet la formation de transcrits circulaires. Les jonctions exoniques inversées, à l'origine de la formation de ces transcrits, ont été mises en évidence par séquençage des ADNc double brin obtenus par une transcription inverse suivie d'une réaction de PCR (RT-PCR). Ces transcrits ont été observés dans l'ARN de lignées cellulaires humaines d'origine hématopoïétique ou épithéliale, mais également dans l'ARN provenant de tissus hépatiques humains normaux (figure 2b, page 57; Cocquerelle, 1990). La visualisation de ces jonctions par la technique de protection aux ribonucléases nous a permis de rejeter rapidement l'éventualité d'un artéfact de PCR (figure 3, page 58).

Avant de discuter nos résultats, il nous paraît tout à fait intéressant de décrire les autres gènes pour lesquels un mécanisme d'épissage inversé a pu être mis en évidence. A l'heure actuelle, les travaux démontrant l'existence d'une structure circulaire issue de ce mécanisme d'épissage particulier sont ceux concernant les gènes ETS-1 (Cocquerelle *et al.*, 1993) et SRY (Capel *et al.*, 1993). Les travaux sur les gènes DCC (Nigro *et al.*, 1991) et *limb deformity* (Philip Leder, communication personnelle) démontrent l'existence de jonctions exoniques inversées dans des transcrits, sans déterminer la structure complète de ces derniers. Toutefois, vu les similitudes réactionnelles entre ces cas, il nous paraît vraisemblable que les transcrits de DCC et *limb deformity* soient également circulaires. Afin d'alléger le style de notre discussion, nous parlerons des jonctions exoniques inversées pour les transcrits de DCC et *limb deformity* en terme de transcrits circulaires.

Durant notre Thèse, trois cas de réaction d'épissage inversé ont été observés par différentes équipes. A l'aide de la technique de PCR, Nigro et coll. ont été les premiers à décrire dans la littérature l'existence de transcrits comprenant une jonction exonique inversée pour le gène DCC, gène suppresseur de tumeur délété dans les carcinomes colorectaux, dont le produit est apparenté aux molécules d'adhésion cellulaire comme les N-CAM (Neural Cell Adhesive

Molécule) (Fearon *et al.*, 1990; Nigro *et al.*, 1991). La présence de ces transcrits a pu être mise en évidence dans des tissus normaux et tumoraux chez la souris et l'homme (Nigro *et al.*, 1991). Le groupe de Philip Leder à Boston a mis en évidence l'existence de plusieurs transcrits comprenant une jonction exonique inversée pour le gène *limb deformity* (P. Leder, communication personnelle). Ces transcrits ont été identifiés en criblant une banque d'ADNc de bouton embryonnaire de membre de souris (Jackson-Grusby *et al.*, 1992). Le gène *limb deformity* code une nouvelle classe de protéines, les formines (Woychik *et al.*, 1990), et il est impliqué en partie dans le développement antéro-postérieur des membres chez les embryons de poulet et de souris (Woychik *et al.*, 1990; Jackson-Grusby *et al.*, 1992; Trumpp *et al.*, 1992). Des observations très intéressantes sont celles décrites récemment par Capel et coll. pour le gène SRY, responsable du déterminisme sexuel mâle chez les mammifères (Capel *et al.*, 1993). Capel et coll ont non seulement mis en évidence l'existence de transcrits comprenant une jonction exonique inversée chez la Souris, mais également caractérisé la structure circulaire de ces molécules d'ARN, confirmant l'existence de transcrits circulaires chez les eucaryotes supérieurs comme l'ont montré nos travaux originaux. Le gène SRY est le seul gène du chromosome Y requis pour le développement des caractères mâles chez les mammifères (Berta *et al.*, 1990; Gubbay *et al.*, 1990; Sinclair *et al.*, 1990; Koopman *et al.*, 1991; Foster *et al.*, 1992; Gubbay *et al.*, 1992; Harley *et al.*, 1992; Hawkins *et al.*, 1992a, 1992b; McElreavey *et al.*, 1993). La particularité du gène SRY est que le cadre de lecture est contenu dans une région génomique continue de 2,7 kb, flanquée chez la souris de deux séquences inversées répétées (Capel *et al.*, 1993; figure 9). Le cadre de lecture montre que ce gène code une protéine possédant un domaine central HMG (High Mobility Group), domaine qui est capable d'interagir avec l'ADN *in vitro* (Ferrari *et al.*, 1992; Harley *et al.*, 1992). De par son interaction avec l'ADN, le produit de ce gène initie probablement la différenciation sexuelle mâle en régulant l'expression de certains gènes (Ferrari *et al.*, 1992; Harley *et al.*, 1992). Chez la souris, SRY est exprimé sous forme linéaire dans les cellules des crêtes génitales en développement (Koopman *et al.*, 1990) et sous forme circulaire dans le testicule adulte (Capel *et al.*, 1993). Capel et coll. avancent l'hypothèse suivante: la formation des transcrits circulaires serait la conséquence de l'utilisation d'un

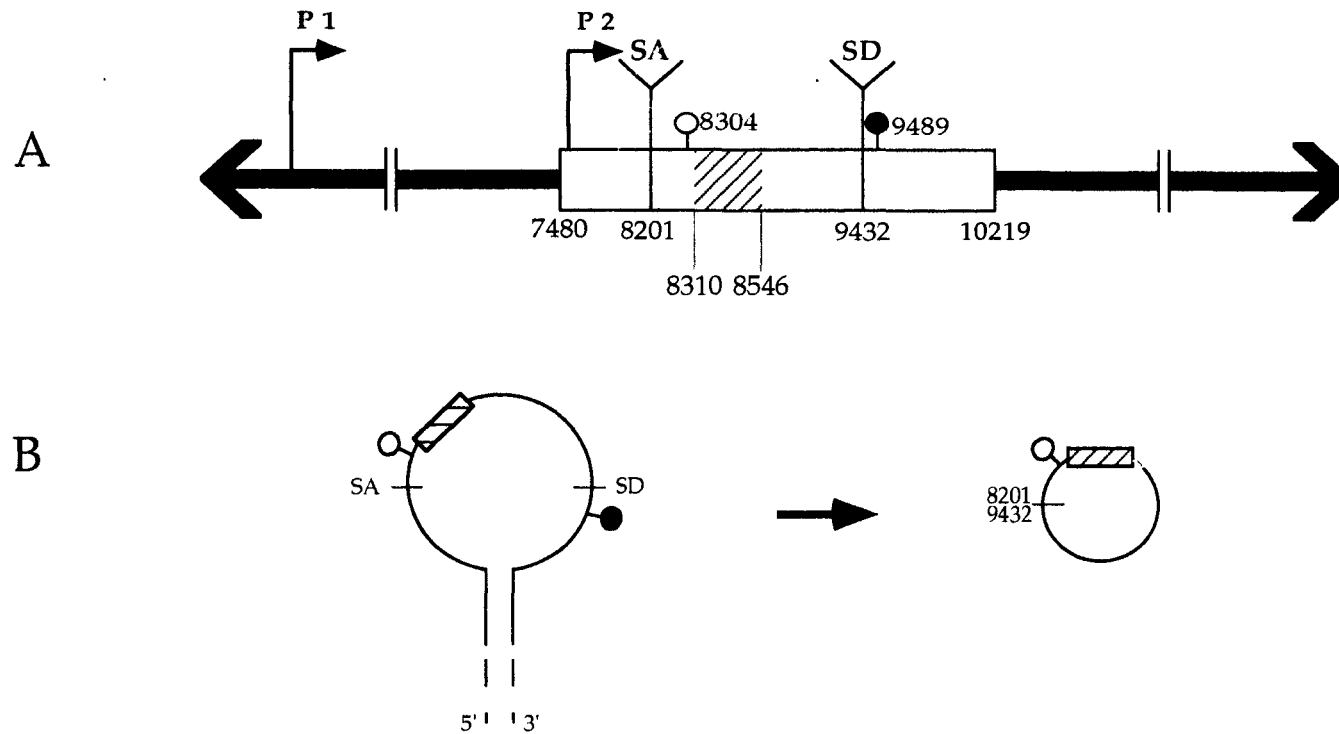


Figure 9: Locus du gène SRY et modèle expliquant la circularisation de son ARN.

(A) La séquence codant pour la boîte HMG (figure en hachurés) se trouve à l'intérieur d'une région de 2,7 kb représentée par la boîte blanche. Cette région est flanquée de deux séquences inversées répétées de plus de 15 kb (flèches noires). Les sites accepteur (SA) et donneur (SD) d'épissage sont indiqués. Le codon d'initiation de la traduction est figuré par un cercle blanc, et le codon stop par un cercle noir. Les positions nucléotidiques de ces différents éléments sont indiquées d'après Capel et coll. (1993). Les promoteurs hypothétiques dans les crêtes génitales (P2) et dans les testicules adultes (P1) sont positionnés. (B) Dans les testicules adultes, le promoteur P1 serait situé en amont de la séquence inverse répétée. L'ARN prémessager comprendrait les deux séquences inverses répétées, qui en s'appariant faciliterait le rapprochement des sites d'épissage et permettrait la réaction d'épissage circularisant l'ARN. (Modifié de Capel et al., 1993).

promoteur localisé en amont de la séquence répétée dans les testicules adultes. Dans les cellules des crêtes génitales, le promoteur fonctionnel serait situé en aval de la région répétée (Capel *et al.*, 1993; figure 9).

Peu d'hypothèses ont été proposées pour expliquer la formation des transcrits circulaires et la fonction potentielle de tels transcrits dans les cellules. En tenant compte des hypothèses déjà avancées, nous allons discuter les résultats concernant le mécanisme d'épissage inversé dans le cadre général de l'épissage, la spécificité de ce mécanisme pour certains exons du locus de ETS-1 et les conséquences de la structure circulaire de l'ARN, en termes de stabilité et de rôle biologique potentiel.

Les deux points fondamentaux que doit respecter le mécanisme d'épissage sont, d'une part, une reconnaissance correcte des sites d'épissage avec une précision à la base près pour l'excision des introns, et d'autre part, un appariement correct des sites dans une polarité 5'-3'. Si les nombreuses études réalisées sur la réaction d'épissage et les facteurs protéiques et/ou ribonucléoprotéiques constituant le spliceosome permettent de mieux appréhender la précision (Green, 1991; Moore *et al.*, 1993), peu de choses sont connues concernant l'appariement adéquat des exons, surtout dans le cas de gènes pourvus de nombreux introns et/ou de grandes régions introniques.

Les séquences des jonctions d-a1 et c-a1 des transcrits circulaires de ETS-1 révèlent la présence de jonctions exoniques (figure 2c, page 57; figure 3, page 65), impliquant les sites donneurs d'épissage des exons d et c et le site accepteur d'épissage de l'exon a1, normalement utilisés pour la formation du transcrit linéaire (Jorcyk *et al.*, 1991). Pour les gènes DCC et *limb deformity*, les sites d'épissage inversé sont également ceux utilisés lors de la maturation normale des ARN pré-messagers (Nigro *et al.*, 1991; P. Leder, communication personnelle). En ce qui concerne SRY, la structure particulière du locus permet que des sites d'épissage cryptiques soient utilisés pour une réaction d'épissage circularisant l'ARN (Capel *et al.*, 1993; figure 9).

Les épissages inversés de ETS-1, SRY et DCC ont été observés *in vivo* et concernent des ARN pré-messagers nucléaires dont la maturation normale nécessite l'intervention d'un

complexe ribonucléoprotéique en trans, le spliceosome. D'autres cas d'épissage circularisant des séquences exoniques ont pu être observés *in vitro* pour des introns autocatalytiques. En effet, Puttaraju et coll. d'une part et Jarrell d'autre part ont montré qu'il était possible d'obtenir *in vitro* des séquences exoniques circulaires en utilisant les propriétés d'auto-épissage des introns de groupe I et II (Puttaraju *et al.*, 1992; Jarrell, 1993). Les introns de groupe I et II sont très structurés et plusieurs domaines importants pour l'activité autocatalytique ont été délimités. Ces domaines sont nommés P à S pour les introns de groupe I et d1 à d6 pour les introns de groupe II (Davies *et al.*, 1982; Michel *et al.*, 1982; figure 1, page 13). Les expériences de Puttaraju et coll. et de Jarrell consistent, à l'aide de constructions plasmidiques, à permuter l'ordre de ces séquences introniques, essentielles à l'épissage, l'ordre des sites d'épissage et celui des séquences exoniques (Puttaraju *et al.*, 1992; Jarrell, 1993; figure 10). Les auteurs observent que la réaction d'épissage qui a lieu suite à la permutation des sites d'épissage et des domaines structuraux circularise l'ARN (Puttaraju *et al.*, 1992; Jarrell, 1993; figure 9). Des travaux plus anciens de Price et coll. sur l'intron de groupe I de l'ARNr de *Tetrahymena thermophyla* ont révélé que l'absence d'un site d'épissage fonctionnel en 5' de l'intron peut mener à l'utilisation d'un site cryptique dans la séquence exonique en aval de l'intron. En conséquence, la réaction d'auto-épissage entre ce site 5' cryptique et le site 3' circularise la séquence exonique (Price *et al.*, 1987). Ces expériences *in vitro* sur les introns auto-épissables montrent que la fonction des sites d'épissage de ces introns n'est pas déterminée par leurs positions relatives (5' ou 3'), mais par la séquence primaire de ces sites ainsi que par les structures secondaires et tertiaires de la molécule d'ARN qui se reconnaissent, se lient et activent ces sites (Price *et al.*, 1987). Ces résultats, associés à ceux concernant l'épissage inversé pour les transcrits primaires d'eucaryotes supérieurs, prouvent une fois de plus l'existence de similitudes entre le mécanisme d'épissage des ARN pré-messagers nucléaires et l'auto-épissage des introns de groupe I et II (Weiner, 1993). En effet, de manière générale ces mécanismes d'épissage sont basés sur deux réactions similaires de trans-estérification et l'on pense que l'épissage des ARN pré-messagers nucléaires est un processus catalysé par des molécules d'ARN, comme dans l'auto-épissage des introns de groupe I et II (Cech, 1986; Steitz, 1992; Moore *et al.*, 1993). De plus, l'auto-épissage des introns de



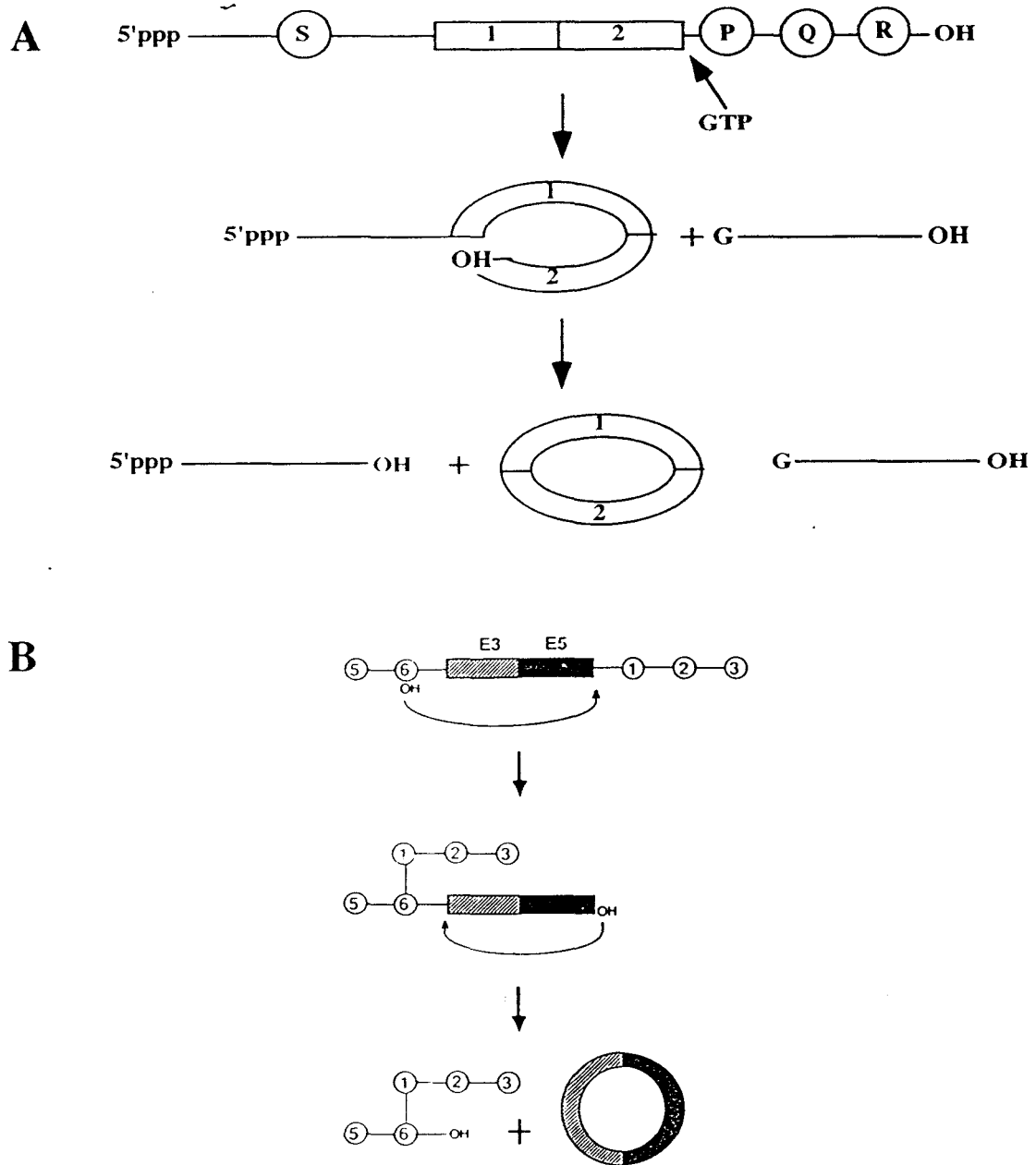


Figure 10. Epissage inversé pour les introns de groupes I et II.

(A) Intron de groupe I. Les éléments conservés de l'intron de groupe I sont représentés dans des cercles. Les exons inversés dans la construction sont numérotés selon leur ordre normal. Le GTP initiant la réaction est indiqué. (B) Intron de groupe II. Les domaines structuraux de l'intron sont indiqués par des chiffres encadrés. Le groupement OH initiant la réaction est indiqué. Dans le cas de l'intron de groupe II, une structure en Y se forme comme il est figuré. (modifié de Puttaraju et coll., 1992 et Jarrell, 1993)

groupe II mène à la formation d'une boucle de lariat comme l'épissage des ARN pré-messagers nucléaires (Cech, 1986). Cependant, la différence capitale entre ces mécanismes d'épissage réside dans le fait que l'épissage des introns de groupe I et II, à la différence des ARN pré-messagers nucléaires, ne nécessite pas l'intervention de complexe ribonucléoprotéique en trans. La molécule d'ARN de ces introns contient à elle seule tous les éléments nécessaires pour une structuration en vue de la réaction d'épissage. Une réaction d'épissage inversé pour ces introns ne peut donc avoir lieu qu'en présence d'une permutation des éléments introniques. L'existence de réactions d'épissage inversé pour les ARN pré-messagers nucléaires, dans lesquelles aucune permutation n'a été effectuée, démontre que la polarité 5'-3' observée lors de l'épissage doit être assurée par un ou plusieurs mécanismes faisant intervenir des éléments en trans, cela après la reconnaissance correcte des sites d'épissage dans le transcrit immature .

Le mécanisme d'épissage inversé peut s'inscrire comme une variante de l'épissage dans le modèle de «définition de l'exon» de Robberson, modèle le plus retenu à l'heure actuelle pour expliquer le mécanisme de sélection des sites d'épissage (Robberson *et al.*, 1990). Ce modèle veut que la séquence exonique soit l'unité de base reconnue initialement dans la réaction d'épissage. L'appariement de deux exons n'est donc plus dépendant d'une reconnaissance au préalable de l'intron les séparant, comme dans le modèle de «linear scanning» (Aebi et Weissmann, 1987). Le modèle de «définition de l'exon» régit la sélection des sites d'épissage à travers l'exon mais n'explique en rien l'appariement de ceux-ci. Nous observons une définition correcte des exons subissant un épissage inversé. En effet les sites à l'extrémité de ces exons participent à une sélection et à un épissage normal. Cependant, le mécanisme d'épissage inversé n'effectue pas l'association des sites d'épissage dans la polarité 5'-3'. Une étude plus approfondie de ce mécanisme conjointement au modèle de «définition de l'exon» pourrait être du plus grand intérêt pour élucider les mécanismes assurant la polarité 5'-3' de l'épissage des ARN pré-messagers nucléaires chez les eucaryotes supérieurs.

L'épissage inversé peut être considéré d'une certaine manière comme une réaction d'épissage en trans puisque sa réaction associe des sites d'épissage en dehors de la polarité 5'-3'. Bruzik et coll. ont récemment montré qu'un ARN SL (*Spliced leader RNA*) de nématode peut

être épissé en trans *in vitro* et *in vivo* à l'aide de la machinerie d'épissage des cellules HeLa (Bruzik *et al.*, 1992). Cet ARN SL a pu être épissé en trans avec des sites accepteur d'épissage de nématode ou de mammifère. Ces auteurs ont ainsi démontré la potentialité d'épissage en trans chez les cellules de mammifères, qui doit être assurée par des facteurs encore inconnus à l'heure actuelle. Par ailleurs, plusieurs cas d'épissage en trans éventuel chez les mammifères ont été publiés dans la littérature (Shimizu *et al.*, 1989; Vellard *et al.*, 1991). Vellard et coll. ont ainsi mis en évidence chez le poulet et l'homme un transcrite chimérique issu de deux molécules d'ARN distinctes, l'une provenant du proto-oncogène *c-myc* et l'autre d'un gène nommé ET (Vellard *et al.*, 1991). L'équipe de Bernard Perbal a par ailleurs démontré que le facteur d'épissage SC35 était codé par le brin complémentaire de celui du gène ET (Vellar *et al.*, 1992). Bien qu'il n'a pas été parfaitement démontré que cette réaction implique la machinerie d'épissage, elle révèle la potentialité de jonctions d'ARN en trans, ainsi que la diversité et la complexité de la maturation des ARN pré-messagers nucléaires chez les eucaryotes supérieurs. Les éventuels facteurs autorisant de telles réactions en trans chez les mammifères pourraient également être impliqués dans le mécanisme d'épissage inversé.

Comme cela a été montré pour SRY, DCC et ETS-1, le mécanisme d'épissage chez les eucaryotes supérieurs possède à l'évidence la potentialité de pouvoir inverser la polarité de la réaction d'épissage. Quelles sont les conditions requises pour une telle inversion de polarité, ou encore quels peuvent être les mécanismes garants de la polarité 5'-3' au cours de l'épissage? L'étude de la spécificité de la réaction d'épissage inversé pour le locus de ETS-1 et l'analyse de la structure de ce locus nous apportent des éléments de réponse à ces questions.

Après avoir découvert la jonction exonique inversée entre les exons d et a1 du locus de ETS-1 (figure 2, page 57), nous nous sommes demandés si tous les sites d'épissage des six exons internes du locus de ETS-1 pouvaient participer à des réactions d'épissage inversé. A l'aide de couples d'amorces oligonucléotidiques permettant de tester toutes les possibilités de jonctions inversées pour les exons du gène ETS-1, nous avons détecté uniquement deux jonctions exoniques impliquant le site accepteur d'épissage de l'exon a1 et les sites donneurs d'épissage

des exons c et d (figure 4, page 58; figure 2, page 64). Ces résultats montrent qu'il existe une spécificité de la réaction d'épissage inversé pour ces trois sites d'épissage. Il se peut que d'autres jonctions exoniques existent, mais le fait que nous ne les détectons pas à l'aide de la technique très sensible de PCR prouve que la potentialité d'épissage inversé pour ces sites est très faible, par rapport à celle des sites des exons a1, c et d.

La comparaison des gènes ETS-1, DCC et *limb deformity* montre qu'il pourrait exister une corrélation entre le mécanisme d'épissage inversé et la présence de grandes régions introniques. En effet, en ce qui concerne le gène ETS-1, le site accepteur d'épissage de l'exon a1 et les sites donneur d'épissage des exons c et d sont impliqués dans l'excision de grands introns. Le site accepteur de a1 est adjacent à un intron de 30 kb et le site donneur de d adjacent à un intron de 15 kb (Jorcyk *et al.*, 1991). En ce qui concerne le site donneur d'épissage de l'exon c, il est parfois impliqué, du fait de l'épissage alternatif de l'exon d, dans l'excision d'une séquence intronique de 20 kb. L'épissage alternatif de l'exon d est en effet observé avec une fréquence relativement élevée au niveau du transcrit linéaire de ETS-1 (Jorcyk *et al.*, 1991; Collyn-d'Hooghe *et al.*, 1993). La corrélation entre épissage inversé et grands introns peut être également envisagée pour les gènes DCC et *limb deformity*. En effet, le gène DCC comprend plus de vingt exons répartis dans une région de plus de 400 kb (Fearon *et al.*, 1990; Bert Vogelstein, communication personnelle) et le gène *limb deformity* comprend 12 exons sur plus de 200 kilobases (P. Leder, communication personnelle). Par ailleurs, Nigro et coll. ont montré que pour un gène de petite taille (le gène p53) aucun produit d'épissage inversé n'était détectable par PCR (Nigro *et al.*, 1991). Ce résultat est en faveur d'un rôle direct ou indirect des grands introns dans l'épissage inversé. Cependant il est également possible que le ou les mécanismes impliqués dans l'élimination d'exons alternatifs ait une relation avec la formation de transcrits circulaires.

En effet, pour expliquer l'épissage alternatif dans le cadre du modèle de «définition de l'exon», nous devons supposer que la définition de l'exon alternatif est, soit inefficace (Robberson *et al.*, 1990), soit «dissimulée» par un mécanisme particulier intervenant avant l'excision des introns adjacents à l'exon. Nous pensons que le mécanisme de «dissimulation» d'un exon alternatif peut avoir un rapport avec le mécanisme d'épissage inversé, car la spécificité

d'épissage inversé de ETS-1 est associée au site donneur d'épissage de l'exon d qui est fréquemment la cible d'épissage alternatif. De plus, le gène DCC code une protéine fortement similaire aux molécules d'adhésion cellulaire neuronale (N-CAM) et autres glycoprotéines de surface cellulaire apparentées (Fearon *et al.*, 1990). Il est ainsi très probable que les transcrits immatures de DCC soient la cible de nombreux épissages alternatifs, comme cela est le cas pour l'ARN prémessager du gène N-CAM, épissé en plus de 27 transcrits alternatifs différents lors du développement embryonnaire du coeur chez le Rat (Reyes *et al.*, 1991). En ce qui concerne le gène *limb deformity*, il est intéressant de remarquer que les transcrits possédant une jonction exonique inversée sont issus de la conservation d'un exon subissant un épissage alternatif dans les transcrits linéaires (P. Leder, communication personnelle).

Dans cet ordre d'idée, il est intéressant d'exposer l'hypothèse de Jarrell concernant la relation entre l'épissage inversé et l'épissage alternatif (Jarrell, 1993). Cet auteur propose que lors de l'épissage alternatif, il existerait une étape dans laquelle l'exon subirait un épissage inversé utilisant ses sites accepteur et donneur d'épissage. Cet épissage inversé circulariserait l'exon alternatif, en vue de son excision du transcrit primaire. La jonction entre les deux exons situés respectivement en 5' et en 3' de l'exon alternatif serait concomitante de l'épissage circularisant celui-ci. Il est intéressant de discuter certaines de nos expériences dans le cadre de cette hypothèse. Lors des expériences de PCR réalisées afin de démontrer le caractère circulaire des transcrits c-a1 et d-a1, la molécule c-a1 a pu être amplifiée à l'aide de deux amorces situées dans l'exon c (figure 2e, page 64). A l'opposé, nous n'avons jamais pu déterminer pourquoi une amplification du transcrit d-a1 n'était pas possible en utilisant deux amorces dans l'exon d. L'hypothèse de Jarrell pourrait être la solution de ce problème. En effet, selon cette hypothèse l'épissage alternatif de l'exon d devrait le circulariser. Puisque la fréquence d'épissage alternatif de l'exon d est élevée dans la lignée cellulaire CEM (Collin-d'Hooghe *et al.*, 1993), il est possible que l'utilisation d'amorces dans l'exon d engendre une compétition pour l'amplification de la molécule circulaire d-a1 et l'hypothétique molécule circulaire de l'exon d. Les conditions de nos expériences ne nous permettaient pas de visualiser un produit d'amplification de petite taille provenant de l'exon alternatif circulaire. Ces remarques sont d'autant plus intéressantes

que des expériences de protection aux ribonucléases montrent que la molécule circulaire d-a1 est exprimée en quantité beaucoup plus importante que la molécule c-a1 (Marielle Cavrois, communication personnelle).

Toutes ces observations vont dans le sens d'une relation étroite entre l'épissage inversé et l'épissage alternatif. Dans le cas des transcrits circulaires de ETS-1, une alternative dans l'épissage inversé de l'exon d selon le modèle de Jarrell pourrait entraîner la formation de jonctions exoniques entre les sites donneur d'épissage des exons c et d et le site accepteur d'épissage de l'exon a1. L'hypothèse de Jarrell pour les ARN pré-messagers nucléaires devrait être facilement confirmable en démontrant l'existence d'exons alternatifs circulaires.

Une des premières possibilités que nous avons envisagé pour expliquer la spécificité de la réaction d'épissage inversé, était que les sites d'épissage impliqués dans cette réaction pouvaient présenter une séquence nucléotidique particulière. En effet, des séquences différentes de celles habituellement observées pour les sites d'épissage 5' et 3', la séquence de branchement ou la séquence polypyrimidique pourraient être responsables de cette spécificité. Cependant, l'analyse et la comparaison des sites d'épissage des exons de ETS-1 n'ont révélé aucune séquence propre aux sites impliqués dans l'épissage inversé.

Les expériences de Price et coll. (1987) sur l'intron de groupe I de *Tetrahymena thermophyla* ont amené la réflexion suivante: l'épissage inversé menant aux transcrits circulaires d-a1 et c-a1 pourrait être la seule réaction d'épissage possible pour un ARN pré-messager tronqué de ETS-1. En effet, il se peut que les introns de grande taille puissent être, au niveau de l'ARN immature, des cibles plus fréquentes de cassures spontanées ou induites, ou d'autres phénomènes comme des initiations non spécifiques et/ou des arrêts prématurés de la transcription. L'ARN pré-messager de ETS-1 résultant de tels phénomènes ne comprendrait plus le ou les sites d'épissage des exons i et e, respectivement partenaires du site accepteur d'épissage de l'exon a1 et des sites donneur d'épissage des exons c ou d. L'épissage inversé entre ces derniers sites serait la seule possibilité pour la poursuite de la maturation de l'ARN. Des cassures induites volontairement dans ces introns et créant des ARN tronqués devraient augmenter le niveau d'épissage inversé. Des expériences de transfection d'oligonucléotides antisens s'hybridant aux

grandes régions introniques dans des cellules HeLa ont été réalisées dans notre équipe, cela afin de cliver l'ARN prémessager à l'aide de l'activité RNase H endogène des cellules. A l'heure actuelle, ces expériences n'ont donné aucun résultat positif.

Nous pouvons également penser que certaines espèces d'ARN prémessager - présentes en quantité importante dans le noyau du fait de l'existence d'une étape limitante dans le processus complet de maturation - peuvent être les cibles préférentielles pour une réaction d'épissage inversé. La réaction d'épissage inversé serait une erreur de la machinerie d'épissage, ayant lieu à une fréquence faible et de manière aléatoire avec les exons non épissés dans le noyau. L'existence d'une molécule d'ARN immature comprenant encore les grands introns en amont de a1 et en aval de d, et plus abondante que les autres formes partiellement épissées, pourrait s'expliquer par le fait que les introns de grande taille ou contenant des exons alternatifs semblent être excisés difficilement (Bovenberg *et al.* 1986, Gattoni *et al.* 1986). La grande taille du locus de ETS-1 ne nous a pas permis d'utiliser la technique de PCR pour analyser la chronologie de l'épissage, comme cela a été réalisé récemment pour le gène de l'Adenine Phosphoribosyltransferase (Kessler *et al.* 1993). C'est donc par des expériences de protection aux ribonucléases que nous avons analysé l'excision des introns adjacents à l'exon e. Nous avons pu montrer que l'intron de 15 kb entre les exons d et e est plus rapidement excisé dans le noyau que celui de 1 kb entre les introns e et f (Figure 3, page 91). Ces résultats infirment, à notre grand étonnement, l'excision difficile des grands introns et donc l'hypothèse de la spécificité de l'épissage inversé due à l'existence d'une molécule intermédiaire majoritaire dans le noyau. Il est évident que la réaction d'épissage inversé a lieu pour un ARN ne contenant plus que les deux grands introns ou pour un ARN encore plus immature. A l'inverse de notre hypothèse, certaines de ces molécules, qui ne semblent pas être en concentration élevée, sont peut-être de façon non aléatoire et plus fréquente la cible de l'épissage inversé pour une raison non déterminée.

En conclusion, nous devons supposer que des éléments dans ou à proximité des grands introns favorisent les réactions d'épissage inversé au niveau du locus de ETS-1. De façon tout à fait logique, nous pouvons imaginer l'existence, dans les grandes régions introniques, de séquences complémentaires favorisant le rapprochement de certains sites d'épissage, comme

cela a déjà été décrit concernant la formation du transcrite circulaire pour SRY (Capel *et al.*, 1993). Par ailleurs, pour un épissage 5'-3', il a été montré que l'appariement de séquences introniques dans l'intron du gène *RP51A* de Levure facilitait la sélection des sites d'épissage du gène (Goguel *et al.*, 1993). Pour ETS-1, cette hypothèse est difficilement abordable, du fait de la grande taille des introns et l'impossibilité de travailler avec un locus de plus de 60 kb. Une étude approfondie de la spécificité d'épissage inversé, de la structure et de l'expression des gènes DCC et *limb deformity*, permettra d'affirmer ou d'infirmer que ces différents éléments: épissage alternatif, ordre d'épissage, grands introns et séquences complémentaires, sont impliqués indépendamment ou de façon associée pour faciliter les réactions d'épissage inversé.

La mise en évidence de jonctions exoniques inversées entre des exons appartenant au même locus est le résultat d'un mécanisme d'épissage soit intramoléculaire soit intermoléculaire. Nous avons eu rapidement la conviction que les transcrits comprenant ces jonctions exoniques particulières sont issus d'un mécanisme intramoléculaire et de plus qu'ils ont une structure circulaire (figure 5, page 66). En effet, plusieurs résultats nous ont permis de conclure dans ce sens: (i) les expériences de PCR montrent que les jonctions exoniques inversés sont présentes dans la fraction d'ARN non polyadénylés et absentes dans la fraction d'ARN polyadénylés (Figure 2b, page 57), (ii) les jonctions exoniques inversées sont observées dans les fractions de petites tailles moléculaires (300 et 900 bases) après séparation d'ARN non polyadénylé sur gradient de saccharose (Figure 4, page 65) et (iii) les expériences de PCR d'une part et les expériences de migration en gel dénaturant d'acrylamide d'autre part, ont confirmé que les transcrits d-a1 et c-a1 ont bien une structure circulaire et ne contiennent que les séquences exoniques entre les deux sites d'épissage inversé (Figure 2, page 64; Figure 1, page 89).

La mise en évidence de transcrits circulaires dans les cellules eucaryotes supérieures, comme cela a été montré pour les gènes ETS-1, SRY, DCC et *limb deformity*, amène la question du rôle éventuel d'un tel mécanisme et de ces ARN circulaires dans la cellule.

Peu d'exemples de molécules d'ARN circulaires sont connus *in vivo*. La structure

circulaire est une caractéristique commune des viroïdes et des virosoïdes, qui sont de petits ARN pathogènes chez les plantes (Symons, 1989). Chez l'Homme, le virus delta de l'hépatite B est également un petit ARN circulaire. Ces auteurs ont montré que ce virus était phylogénétiquement proche des viroïdes (Elena *et al.*, 1991). Enfin, les introns de groupe I qui sont excisés par auto-épissage peuvent se circulariser *in vivo* (Cech, 1993). Il est intéressant de noter que les petits ARN circulaires des viroïdes et des virosoïdes ainsi que du virus delta de l'hépatite B ont les mêmes propriétés catalytiques que les introns de groupe I (Elena *et al.*, 1991). De nombreuses spéculations ont été faites quant à la relation phylogénétique entre les viroïdes et les introns de groupe I, dans le cadre du «monde de l'ARN» à l'origine de l'évolution précellulaire (Diener, 1989). Il s'avère ainsi qu'une relation phylogénétique pourrait exister entre ces différentes molécules d'ARN circulaires et les transcrits maturés par épissage inversé. Cette réaction d'épissage inversé pourrait être considérée comme le reliquat d'un mécanisme d'épissage circularisant les ARN. Dans le même ordre d'idées, il se peut que l'épissage inversé soit un mécanisme participant à l'évolution des gènes. L'hypothèse de l'*exon shuffling* de Walter Gilbert veut que les exons représentent les domaines fonctionnels des protéines et que la recombinaison de ces exons, grâce à la présence des régions introniques, crée des combinaisons diverses de ces domaines protéiques (Gilbert, 1978). L'épissage inversé a pu ou pourrait participer à l'*exon shuffling* en excisant les exons comme dans le modèle de Jarrell (Jarrell, 1993). Une réaction réversible pourrait échanger des séquences exoniques au sein de l'ARN prémessager et une transcription inverse suivie d'une intégration dans le génome créerait ainsi de nouveaux gènes.

L'expression des transcrits circulaires pour le gène ETS-1 est très faible par rapport au transcrit linéaire. En effet, les expériences de protection aux ribonucléases montrent que le transcrit d-a1 représente 1 % du taux de transcrit linéaire de ETS-1 (Figure 3, page 58; Figure 4, page 92). La variation de l'expression des transcrits circulaires dans les différentes lignées est corrélée au taux de transcription du locus. Nous en concluons que dans le cas de ETS-1, la réaction d'épissage inversé est probablement associée à une erreur de la machinerie cellulaire d'épissage, à moins que cette faible expression qui reflète le taux de transcription du locus soit suffisante pour une action biologique inconnue à ce jour. Cependant, l'étude non effectuée de

l'expression du transcrite circulaire dans les tissus adultes ou embryonnaires pourrait révéler un taux plus élevé, comme cela est le cas pour les transcrits circulaires de SRY dans le testicule (Capel *et al.*, 1993). Certains résultats vont dans ce sens. En effet, l'étude de l'expression des transcrits circulaires de DCC dans le Rat adulte montre des variations importantes suivant les tissus étudiés (Nigro *et al.*, 1991) et les expériences de protection aux ribonucléases réalisées pour le gène *limb deformity* semblent montrer une forte expression des transcrits circulaires dans le cerveau (P. Leder, communication personnelle).

Nous avons montré que les transcrits circulaires d-a1 et c-a1 sont stables puisque nous pouvons les observer après 48 heures de traitement à l'actinomycine D (Figure 1C, page 63). De plus, les expériences réalisées par Marielle Cavrois à l'aide de la technique de protection aux ribonucléases ont démontré que la demi-vie du transcrite d-a1 est d'environ neuf heures (Figure 5, page 93). Il est évident que la stabilité des transcrits circulaires ne peut être due à la queue de polyadénylation, comme cela a été démontré pour d'autres transcrits (Hentze, 1991). La stabilité pourrait s'expliquer par la perte d'extrémités, rendant ces molécules circulaires résistantes aux exonucléases et de plus, il se peut qu'à cause de leur forme circulaire compacte, ces molécules offrent une résistance accrue aux endonucléases. Une telle stabilité pour des molécules d'ARN a déjà été observée lors d'injections d'introns de groupe I circulaires dans des oocytes de batraciens (Hentze, 1991). Par ailleurs, il a déjà été démontré que des transcrits ne comportant pas de cadre de lecture correct à la suite par exemple d'une mutation, d'un épissage aberrant ou de la conservation d'une séquence intronique, étaient dégradés rapidement, avant même leur transport complet vers le cytoplasme, cela par un mécanisme «de lecture d'épreuve» («*Proofreading*», Moore *et al.*, 1993). Par conséquent, la présence et la stabilité des molécules d'ARN circulaires malgré l'éventuelle présence d'un tel mécanisme semble donc être en faveur de l'existence d'un rôle biologique pour ces transcrits.

Des expériences de PCR ont montré que les transcrits circulaires d-a1 et c-a1 de ETS-1 ont une localisation cytoplasmique (Figure 2b, page 64). Une même localisation a pu être mise en évidence pour les transcrits des gènes DCC et SRY (Nigro *et al.*, 1991; Capel *et al.*, 1993). Le transport au cytoplasme des molécules circulaires est un problème non résolu, et ce d'autant

plus qu'il a été montré que la coiffe à l'extrémité 5' des ARN messagers était impliquée dans ce transport (Hamm et Mattaj, 1990). Toutefois, nous ne savons pas à quel moment se produit la réaction d'épissage inversé au cours de la maturation de l'ARN prémessager de ETS-1. On peut penser qu'une réaction d'épissage inversé se produisant tardivement, lorsque la molécule d'ARN est proche du complexe de pore nucléaire, faciliterait le passage dans le cytoplasme.

Nous nous sommes posé la question de la traduction de ces ARN circulaires à localisation cytoplasmique. Puisque ces molécules n'ont pas de coiffe en 5', l'initiation de la traduction ne peut s'effectuer selon le modèle parfaitement admis de démarrage par la coiffe en 5'. Une initiation interne de la traduction se révèle nécessaire pour les transcrits circulaires. Une telle initiation a déjà été observée pour des ARN messagers de picornavirus et de cellules eucaryotes (Jackson *et al.*, 1990; Macejak *et al.*, 1991). Cependant l'interprétation des résultats obtenus par ces auteurs est fortement controversée (Kozak, 1992). De plus, plusieurs auteurs ont montré que les ribosomes d'eucaryotes ne pouvaient pas se lier à des ARN messagers circulaires (Konarska *et al.*, 1981; Kozak, 1979), ce qui semble être confirmé par Capel et coll. qui ont montré que les transcrits circulaires de SRY n'étaient pas liés aux polysomes (Capel *et al.*, 1993). Il est surprenant de constater que les jonctions exoniques d-a1 et c-a1 conservent le cadre de lecture du transcrit linéaire de ETS-1. A l'opposé, plusieurs des jonctions exoniques inversées pour DCC font perdre le cadre de lecture (Nigro *et al.*, 1991).

Les transcrits circulaires ne semblent donc pas faire l'objet d'une traduction. Il se peut que l'épissage inversé par la structure des transcrits formés permette de prévenir à 100% la traduction de gènes en protéines, soit lorsque la transcription du locus s'effectue à un moment inopportun ou/et lorsque l'expression du gène a une fonction en terme de molécules d'ARN. Concernant SRY, il est intéressant de remarquer que pour la formation des transcrits circulaires dans le testicule adulte de souris, un promoteur alternatif est utilisé et que les sites d'épissage présents aux deux extrémités du gène sont conservés dans plusieurs souches de souris (Capel *et al.*, 1993). Il n'en est pas de même chez l'homme puisque le transcrit est linéaire; cependant aucune preuve directe n'a permis de démontrer la synthèse de protéine au niveau du testicule adulte humain. Le gène SRY a-t-il une fonction ARN et/ou l'expression de ce gène est-elle inappropriée lors

de la maturation du testicule? Dans ce dernier cas, un autre mécanisme que la circularisation de l'ARN doit prévenir la synthèse protéique du transcrit linéaire SRY chez l'homme. De plus, l'expression en protéines d'une quantité très importante de gènes dans des tissus comme le cerveau et le testicule est-elle réellement nécessaire? En effet, il est parfaitement admis que ces tissus expriment au niveau ARN un nombre très important de gènes.

A ce jour, plusieurs exemples d'ARN messagers ne semblant pas être traduits en protéine et ayant apparemment un rôle important au sein de la cellule ont été décrits. C'est notamment le cas de l'ARNm du gène Xist, qui a une localisation nucléaire et semble être responsable de l'inactivation du chromosome X (Kay *et al.*, 1993). Deux autres exemples impliquent des ARN messagers à localisation cytoplasmique. L'un concerne les séquences 3' non traduites de certains ARNm qui sont capables d'induire la différenciation de myoblastes (Rastinejad et Blau, 1993). L'autre, plus étudié du fait de l'existence d'une empreinte parentale sur ce gène (Bartolomei *et al.*, 1991), joue très probablement un rôle en tant que molécule d'ARN. En effet, ce gène, appelé H19, ne possède pas de cadre de lecture et est suspecté avoir une fonction anti-oncogène (Hao *et al.*, 1993).

Bon nombre d'hypothèses peuvent être avancées quant au rôle éventuel des transcrits circulaires dans les cellules d'eucaryotes supérieurs. Il est possible que de telles molécules se lient à des facteurs protéiques ou autres et jouent un rôle, directement ou indirectement, au niveau de la traduction des messagers normaux, du transport de certains facteurs, ou de mécanismes participant aux modifications épigénétiques du génome.

Il est fortement probable qu'il existe un grand nombre de molécules d'ARN circulaires dans la cellule. Le fait que l'existence de ces molécules ait été ignorée jusqu'à ces dernières années, montre les limites des techniques de clonage qui ne sont plus adaptées lorsque l'on est en présence d'ARN très structurés.

**CONCLUSIONS
ET
PERSPECTIVES**

La mise en évidence de transcrits circulaires, issus d'un mécanisme d'épissage inversé, pose le problème du rôle de ce mécanisme et de la fonction biologique de ces molécules d'ARN.

Le taux d'expression des transcrits circulaires pour le gène ETS-1 est faible par rapport à la maturation normale. Il est tentant de penser que cette expression est le résultat d'une erreur de la machinerie d'épissage. Cependant, l'expression majoritaire du gène SRY dans les testicules de souris sous forme de transcrits circulaires laisse envisager un rôle important pour le mécanisme d'épissage inversé et/ou les molécules d'ARN. Ce mécanisme peut soit empêcher la synthèse de protéines d'un gène transcriptionnellement actif, soit produire des transcrits ayant une fonction en tant que molécule d'ARN, comme cela semble exister pour d'autres ARN messagers. Pour le gène ETS-1, la faible expression de transcrits circulaires qui n'affecte pas la synthèse protéique, a-t-elle un rôle biologique?

Il est probable que la réaction d'épissage inversé nécessite des éléments ou des conditions particulières. Nous avons observé une spécificité de la réaction d'épissage inversé au niveau du locus de ETS-1, que nous avons corrélé à l'existence de grandes régions introniques et d'exons alternatifs. La formation de transcrits circulaires SRY est facilitée par des séquences répétées.

Le mécanisme d'épissage inversé montre de nouvelles potentialités de la machinerie d'épissage des ARN pré-messagers nucléaires pour moduler l'expression d'un gène, comme l'avait déjà révélé l'épissage alternatif. En effet, la découverte du mécanisme d'épissage inversé ouvre une nouvelle voie de réflexion à propos de la régulation de l'expression génétique et des fonctions assurées par les produits du gène.

Il existe probablement d'autres gènes exprimés préférentiellement sous forme d'ARN circulaires. La découverte et l'étude de ces gènes permettrait d'identifier les éléments requis pour le mécanisme d'épissage inversé. Le mécanisme circularisant l'ARN est de plus un modèle tout à fait intéressant dans l'étude de l'appariement des sites d'épissage au cours de la réaction d'épissage.

BIBLIOGRAPHIE

- Abelson J. (1992). Recognition of tRNA Precursors: A role for the Intron. *Science*, 255:1390-1408.
- Aebi M. & Weissmann C. (1987). Precision and orderliness in splicing. *TIG*, 3:102-107.
- Akao Y., Tsujimoto Y., Finan J., Nowell P. C. & Croce C. M. (1990). Molecular characterisation of a t(11;14)(q23;q32) chromosome translocation in a B-cell lymphoma. *Cancer Research*, 50:4856-4859.
- Aloni Y., Dhar R., Laub O., Horowitz M. & Khoury G. (1977). Novel mechanism for RNA maturation: the leader sequences of simian virus 40 mRNA are not transcribed adjacent to the coding sequences. *Proc. Natl. Acad. Sci. USA*, 75:3686-3690.
- Anderson G. J., Bach M., Luhrmann R. & Beggs J. D. (1989). Conservation between yeast and man of a protein associated with U5 small nuclear ribonucleoprotein. *Nature*, 342:819-821.
- Anton I. A. & Frampton J. (1988). Tryptophans in myb proteins. *Nature*, 336:719.
- Arenas J. & Hurwitz J. (1987). Purification of an RNA debranching activity from HeLa cells. *J. Biol. Chem.*, 262:4274-4279.
- Aurigemma R. E., Blair D. G. & Ruscetti S. K. (1992). Transactivation of erythroid transcription factor GATA-1 by myb-ets containing retrovirus. *J. Virol.*, 66:3056-3061.
- Bach M., Winkelmann G. & Luhrmann R. (1989). 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. *Proc. Natl. Acad. Sci. USA*, 86:6038-6042.
- Bartolomei M. S., Zemel S. & Tilghman S. M. (1991). Parental imprinting of the mouse H19 gene. *Nature*, 351:153-155.
- Baserga S. J. & Steitz J. A. (1993). The diverse world of small ribonucleoproteins. *The RNA World*, ed. Gesteland R. F. & Atkins J. F., Cold Spring Harbor Laboratory Press.
- Bennett M., Pinol-Roma S., Stakins D., Dreyfuss G. & Reed R. (1992). Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. *Mol. Cell. Biol.*, 12:3165-3175.

- Berger R., Bermheim H., Weh H. J. & Flandrin G. (1980). Cytogenetic studies on acute monocytic leukemia. *Leuk. Res.*, 4:119-127.
- Berger R., Bloomfield C. D. & Sutherland G. (1985). Report of the committee on chromosome rearrangements in neoplasia and on fragile sites. *Cytogenet. Cell Genet.*, 40:490-535.
- Bergeron D., Poliquin L., Kazak C. A. & Rassart E. (1991). Identification of a common viral integration region in Cas-Br-E murine leukemia virus induced non-T, non-B-cell lymphomas. *J. Virol.*, 65:7-15.
- Berget S. M., Moore C. & Sharp P. A. (1977). Spliced RNA segments at the 5' terminus of late adenovirus 2 mRNA. *Proc. Natl. Acad. Sci. USA*, 74:3171-3175.
- Bernards A., Rubin C. M., Westbrook C. A., Paskind M. & Baltimore D. (1987). The first intron in the human c-abl gene is at least 200 kilobases long and is a target for translocations in chronic myelogenous leukemia. *Mol. Cell. Biol.*, 7:3231-3236.
- Berta P., Hawkins J. R., Sinclair A. H., Taylor A., Griffiths B. L., Goodfellow P. N. & Fellous M. (1990). Genetic evidence equating Sry and the male sex determining gene. *Nature*, 348:248-250.
- Beug H., Lentz A., Kahn P. & Graf T. (1984). TS mutants of E26 leukemia virus allow transformed myeloblast but not erythroblasts or fibroblasts, to differentiate at the non permissive temperature. *Cell*, 39:579-588.
- Beyer A. L. & Osheim Y. N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.*, 2:754-765.
- Bhat N. K., Fischer R. J., Fujiwara S., Ascione R. & Papas T. S. (1987). Temporal et tissue specific expression of mouse ets genes. *Proc. Natl. Acad. Sci. USA*, 84:3161-3165.
- Bhat N. K., Komschlies K. L., Fujiwara S., Fisher R. J., Mathieson B. J., Gregorio T. A., Young H. A., Wasik J. W., Ozato K. & Papas T. S. (1989). Expression of ets genes in mouse thymocytes subsets and T cells. *J. Immunol.*, 142:672-678.

- Bhat N. K., Thompson C. B., Lindsten T., June C. H., Fujiwara S., Koizumi S., Fischer R. J. & Papas T. S. (1990). Reciprocal expression of human Ets1 and Ets2 genes during T-Cell activation : regulatory role for the proto-oncogene Ets1. *Proc. Natl. Acad. Sci. USA*, 87:3723-3727.
- Bingham P. M., Chou T. B., Mims I. & Zachar Z. (1988). On/off regulation of gene expression at level of splicing. *T I G*, 4:134-138.
- Blencowe B. J., Sproat B. S., Ryder U., Barabino S. & Lamond A. I. (1989). Antisense probing of the human U4/U6 snRNP with biotinylated 2'-OMe RNA oligonucleotides. *Cell*, 59:531-539.
- Bonen L. (1993). Trans-splicing of pre-mRNA in plants, animals, and protists. *FASEB J.*, 7:40-46.
- Bordonne R., Banroques J., Abelson J. & Guthrie C. (1990). Domains of yeast U4 spliceosomal RNA required for PRP4 protein binding, snRNP-snRNP interactions and pre-mRNA splicing in vivo. *Genes Dev.*, 4:1185-1196.
- Bosselut R., Duvall J. F., Gegonne A., Baily L., Hemar A., Brady J. & Ghysdael J. (1990). The product of the c-ets-1 proto-oncogene and the related Ets-2 protein act as transcriptional activators of the long terminal repeat of human T cell leukemia virus HTLV-1. *Embo. J.*, 9:3137-3144.
- Boulukos K. E., Pognonec P., Rabault B., Begue A. & Ghysdael J. (1989). Definition of an ets 1 protein domain required for nuclear localization in cells and DNA-binding activity in vitro. *Mol. Cell. Biol.*, 9:5718-5721.
- Bovenberg R. A. L., Van de Meerendonk W. P. M., Baas P.D., Steenbergh P. H., Leeps C. J. M. & Jansz H. S. (1986). Model for alternative RNA processing in human calcitonin gene expression. *Nucleic Acids Res.*, 14:8785-8803.
- Brannan C. I., Dees E. C., Ingram R. S. & Tilghman S. M. (1990). The product of the H19 gene may function as an RNA. *Mol. Cell. Biol.*, 10:28-36.
- Breathnach R., Mandel J. L. & Chambon P. (1977). Ovalbumin gene is split in chicken DNA. *Nature*, 270:314-319.

- Breathnach R., Benoist C., O'Hare K., Gannon F. & Chambon P. (1978). Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. USA*, 75:4853-4857.
- Breitbart R. E., Andreadis A. & Nadal-Ginard B. (1987). Alternative splicing: a ubiquitous mechanism for generation of multiple protein isoforms from single genes. *Ann. Rev. Biochem.*, 56:467-495.
- Brockdorff N., Ashworth A., Kay G. F., McCabe V. M., Norris D. P., Cooper P.J., Swift S. & Rastan S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*, 71:515-526.
- Brody E. & Abelson J. (1985). The «spliceosome»: yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science*, 228:963-967.
- Brown C. J., Hendrich B. D., Rupert J. L., Lafrenière R. G., Xing Y., Lawrence J. & Willard H. F. (1992). The human Xist gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*, 71:527-542.
- Brown J. D., Plumpton M. & Beggs J. D. (1992). The genetics of nuclear pre-mRNA splicing: a complex story. *Antonie van Leeuwenhoek*, 62:35-46.
- Brunkow M. E. & Tilghman S. M. (1991). Ectopic expression of the H19 gene in mice causes prenatal lethality. *Gen. Dev.*, 5:1092-1101;
- Bruzik J. P. & Maniatis T. (1992). Spliced leader RNAs from lower eukaryotes are trans-spliced in mammalian cells. *Nature*, 360:692-695.
- Campbell D. A., Thornton D. A. & Boothroyd J. C. (1984). Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes. *Nature*, 311:350-355.
- Capel B., Swain A., Nicolis S., Hacker A., Walter M., Koopman P., Goodfellow P. & Lovell-Badge R. (1993). Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell*, 73:1019-1030.
- Carter K. C., Bowman D., Carrington W., Fogarty K., McNeil J. A., Fay F. S. & Lawrence J. B. (1993). A three dimensional view of precursor messenger RNA metabolism within the mammalian nucleus. *Science*, 259:1330-1335.

- Cavrois M., Cocquerelle C., Hetuin D. & Bailleul B. (1993). Circular RNA transcripts formation by splicing. *Mol. Cell. Biol.*, soumis à publication.
- Cech T. (1986). The Generality of Self-Splicing RNA: Relationship to Nuclear mRNA Splicing. *Cell*, 44:207-210.
- Cech T. R. (1993). Structure and mechanism of the large catalytic RNAs: Group I and group II introns and ribonuclease P. *The RNA World*, ed. Gesteland R. F. & Atkins J. F., Cold Spring Harbor Laboratory Press.
- Chapman K. B. & Boeckle J. D. (1991). Isolation and characterization of the gene encoding yeast debranching enzyme. *Cell*, 65:483-492.
- Chen J. H. (1985). The proto-oncogene c-ets is preferentially expressed in lymphoid cells. *Mol. Cell. Biol.*, 5:2993-3000.
- Chen Z. Q., Kan N. C., Pribyl L., Lautenberger J. A., Moudrianaski E. & Papas T. S. (1988). Molecular cloning of the ets proto-oncogene of the sea urchin and analysis of its developmental expression. *Dev. Biol.*, 125:432-440.
- Choi Y. D., Grabowski P. J., Sharp P. A. & Dreyfuss G. (1986). Heterogeneous nuclear ribonucleoproteins : role in RNA splicing. *Science*, 231:1534-1539.
- Chow L. T., Gelinas R. E., Broker T. R. & Roberts R. I. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, 12:1-8.
- Cimino G., Moir D. T., Canaani O., Crist W. W. M., Katzav S., Cannizzaro L., Lange B., Nowell P. C., Croce C. M. & Canaani E. (1991). Cloning of ALL-1, the locus involved in leukemias with the t(4;11)(q21;q23), t(9;11)(p22;q23), and t(11;19)(q23;p13) chromosome translocations. *Cancer Research*, 51:6712-6714.
- Clouet d'Orval B., d'Aubenton-Carafa Y, Sirand-Pugnet P., Gallego M. E., Brody E. & Marie J. (1991). «RNA secondary structure represses utilization of a muscle specific exon in HeLa cell nuclear extracts». *Science*, 252:1823-1829.
- Cocquerelle C. (1990). Existence d'un transcrit anormalement mûré dans le locus du proto-oncogène ETS-1. DEA Lille I.

- Cocquerelle C., Daubersies P., Majérus M. A., Kerckaert J. P. & Bailleul B. (1992). Splicing with inverted order of exons occurs proximal to large introns. *Embo J.*, 11:1095-1098.
- Cocquerelle C., Mascrez B., Hetuin D. & Bailleul B. (1993). Mis-splicing yields circular RNA molecules. *FASEB J.*, 7:155-159.
- Collyn d'Hooghe M., Galiegue-Zouitina S., Szymiczek D., Lantoiné D., Quief S., Loucheux-Lefebvre M. H. & Kerckaert J. P. (1993). Quantitative and qualitative variation of ETS-1 transcripts in hematologic malignancies. *Leukemia*, sous presse.
- Company M., Arenas J. & Abelson J. (1991). Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature*, 349:387-393.
- Crick F. H. C. (1967). Origin of the genetic code. *Nature (London)*. 213:119.
- Dalrymple M. A., Petersen-Bjorn S., Friesen J. D. & Beggs J. D. (1989). The product of the PRP4 gene of *S. cerevisiae* shows homology to β subunits of G proteins. *Cell*, 58:811-812.
- Davies R. W., Waring R. B., Ray J. A., Brown T. A. & Scazzocchio C. (1982). Making ends meet: a model for RNA splicing in fungal mitochondria. *Nature*, 300:719-724.
- Degnan B. M., Degnan S. M., Naganuma T. & Morse D. E. (1993). The ets multigene family is conserved throughout the metazoa. *Nucleic Acids Res.*, 21:3479-3484.
- De Lange T., Berkuens T. M., Veerman H. J. G., Frasch A. C., Barry J. D. & Borst P. (1984). Comparison of genes coding the common 5' terminal sequence of messenger RNA in three trypanosome species. *Nucl. Acids Res.* 12:4431-4443.
- Desbiens X., Quéva C., Jaffredo T., Stehelin D. & Vandebunder B. (1991). The relationship between cell proliferation and the transcription of the nuclear oncogenes c-myc, c-myb and c-ets-1 during feather morphogenesis in the chick embryo. *Development*, 111:699-713.
- Diaz M. O., Le Beau M. M., Pitha P. & Rowley J. D. (1986). Interferon and c-ets-1 genes in the translocation t(9;11)(p22;q23) in human acute monocytic leukemia. *Science*, 296:265-267.

- Diener T. O. (1989). Circular RNAs: Relics of precellular evolution? *Proc. Natl. Acad. Sci. USA*, 86:9370-9374.
- Djabali M., Selleri L., Parry P., Bower M., Young B. D. & Evans G. A. (1992). A trithorax like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. *Nature Genetics*, 2:113-118.
- Domenget C., Leprince D., Pain B., Peyrol S., Li R. P., Stehelin D., Samarut J. & Jurdic P. (1992). The various domains of v-myb and v-ets oncogenes of E26 retrovirus contribute differently, but cooperatively, in transformation of hematopoietic lineages. *Oncogene*, 7:2231-2241.
- Dreyfuss G., Swanson M. S. & Pinol-Roma S. (1988). Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *TIBS.*, 13: 86-90.
- Dreyfuss G., Matunis M. J., Pinol-Roma S. & Burd C. G. (1993). hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.*, 62:289-321.
- Dunn J. M., Phillips R. A., Zhu X., Becker A. & Gallie B.L. (1989). Mutations in the RB1 gene and their effects on transcription. *Mol. Cell. Biol.*, 9:4596-4604.
- Duterque-Coquillaud M., Leprince D., Flourens A., Henry C., Ghysdael J., Debuire B. & Stehelin D. (1988). Cloning and expression of chicken p54c-ets cDNA : the first p54c-ets coding exon is located into the 400 Kbp genomic domain unrelated to v-ets. *Oncogene Res.*, 2:335-344.
- Elena S. F., Dopazo J., Flores R., Diener T. O. & Moya A. (1991). Phylogeny of viroids, viroidlike satellite RNAs, and the viroidlike domain of hepatitis delta virus RNA. *Proc. Natl. Acad. Sci. USA*, 88:5631-5634.
- Evans R. M. & Hollenberg S. M. (1988). Zinc fingers: guilt by association. *Cell*, 52:1-3.
- Fakan S., Leser G. & Martin T.E. (1986). Immunoelectron microscope visualisation of nuclear ribonucleoprotein antigens within spread transcription complexes. *J. Cell Biol.*, 103:1153-1157.

- Fearon E. R., Cho K. R., Nigro J. M., Kern S. E., Simons J. W., Ruppert J. W., Hamilton S. R., Preisinger A. C., Thomas J., Kinzler K. W. & Vogelstein B. (1990). Identification of a chromosome 18q gene that is altered in colorectal cancers. *Science*, 247:49-56.
- Feener C.A., Koenig M. & Kunkel L. M. (1989). Alternative splicing of human dystrophin mRNA generates isoforms at carboxy terminus. *Nature*, 338:509-511.
- Ferrari S., Harley V. R., Pontiggia A., Goodfellow P. N., Lovell-Badge R. & Bianchi M. E. (1992). SRY, like HMG1, recognizes sharp angles in DNA. *Embo. J.*, 11:4497-4506.
- Fisher R. J., Koizumi S., Kondoh A., Mariano J. M., Mavrothalassitis G., Bhat N. K. & Papas T.S. (1992). Human ETS-1 oncoprotein. Purification, isoforms, -SH modification, and DNA sequence-specific binding. *J. Biol. Chem.*, 25:17957-17965.
- Fleischman L. F., Pilaro A. M., Murakami K., Kondoh A., Fisher R. J. & Papas T. S. (1993). c-ets-1 protein is hyperphosphorylated during mitosis. *Oncogene*, 8:771-780.
- Foster J. W., Brennan F. E., Hampikian G. K., Goodfellow P. N., Sinclair A. H., Lovell-Badge R., Selwood L., Renfree M. B., Cooper D. W. & Marshal Graves M. (1992). Evolution of sex determination and the Y chromosome : SRY-related sequences in marsupials. *Nature*, 359:531-533.
- Fu X. D. & Maniatis T. (1990). Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature*, 343:437-441.
- Fu X. D. (1993). Specific commitment of different pre-messenger RNAs to splicing by single SR proteins. *Nature*, 365:65-82.
- Fujiwara S., Fisher R. J., Bhat N. K., Diaz de la Espina S. M. & Papas T. S. (1988). A short-lived nuclear phosphoprotein encoded by the human ets-2 proto-oncogene is stabilized by activation of protein kinase C. *Mol. Cell. Biol.*, 8:4700-4706.
- Garcia-Blanco M. A., Jamison G. J. & Sharp P. A. (1989). Identification of a 62,000 dalton protein which binds specifically to the polypyrimidine tract of introns. *Genes Dev.*, 3:1874-1886.

- Gattoni R., Keohavong P. & Stévenin J. (1986). Splicing of E2A premessenger RNA of adenovirus serotype 2. *J. Mol. Biol.*, 187:379-397.
- Ge H. & Manley J. L. (1990). A protein factor, ASF, controls cell specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell*, 62:25-34.
- Gegonne A., Leprince D., Duterque-Coquillaud M., Vandebunder B., Flourens A., Ghysdael J., Debuire B. & Stehelin D. (1987a). Multiple domains for the chicken cellular sequences homologous to the v-ets oncogene of the E26 retrovirus. *Mol. Cell. Biol.*, 7:806-812.
- Gegonne A., Leprince D., Pognonec P., Denis D., Raes M. B., Stehelin D. & Ghysdael J. (1987b). The 5' extremity of the v-ets oncogene of avian leukemia virus E26 encodes amino acid sequences not derived from the major c-ets-encoded cellular proteins. *Virology*, 156:177-180.
- Gerke V. & Steitz J. A. (1986). A protein associated with small ribonuclearprotein particules recognizes the 3' splice site of pre-messenger RNA. *Cell*, 47:973-984.
- Ghetti A., Pinol-Roma S., Michael C., Morandi C. & Dreyfuss G. (1992). hnRNP I the polypyrimidine tract-binding protein : distinct nuclear localization and association with hnRNAs. *Nucleic Acids Res.*, 20:3671-3678.
- Ghysdael J., Gegonne A., Pognonec P., Denis D., Leprince D., Stehelin D. (1986). Identification and preferential expression in thymic and bursal lymphocytes of a c-ets oncogene-encoded Mr 54,000 cytoplasmic protein. *Proc. Natl. Acad. Sci. USA*, 83:1714-1718.
- Gil A., Sharp P. A., Jamison S. F. & Garcia-Blanco M. A. (1991). Characterization of cDNAs encoding the polypyrimidine tract binding protein. *Genes Dev.*, 5:1224-1236.
- Gilbert W. (1978). Why genes in pieces? *Nature*, 271:501.
- Glover D. M. & Hogness D. S. (1977). A novel arrangement of the 18S and the 28S sequences in a repeating unit of *Drosophila melanogaster*. *Cell*, 10:167-176.
- Goguel V. & Rosbash M. (1993). Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base-pairing in yeast. *Cell*, 72:893-901.

- Golay J., Introna M. & Graf T. (1988). A single point mutation in the v-ets oncogene affects both erythroid and myelomonocytic cell differentiation. *Cell*, 55:1147-1158.
- Gonda T. J. & Bishop J. M. (1983). Structure and transcription of the cellular homolog (c-myb) of the avian myeloblastosis virus transforming gene (v-myb). *J. Virol.*, 46:212-220.
- Goodman H. M., Olson M. V. & Hall B. D. (1977). Nucleotide sequence of a mutant eukaryotic gene: the yeast tyrosine inserting ochre suppressor SUP4-0. *Proc. Natl. Acad. Sci. USA*, 74:5453-5457.
- Grabowski P. J., Seller S. R. & Sharp P. A. (1985). A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell*, 42:345-353.
- Graf T., Oker-Blom N., Todorov T. G. & Beug H. (1979). Transforming capacities and defectiveness of avian leukemia virus OK10 and E26. *Virology*, 99:431-436.
- Green M. R. (1991). Biochemical mechanisms of constitutive and regulated pre-mRNA Splicing. *Annu. Rev. Cell Biol.*, 7:559-599.
- Gubbay J., Collignon J., Koopman P., Capel B., Economou A., Münsterberg A., Vivian N., Goodfellow P. & Lovell-Badge R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature*, 346:245-250.
- Gubbay J., Vivian N., Economou A., Jackson D., Goodfellow P. & Lovell-Badge R. (1992). Inverted repeat structure of the Sry locus in mice. *Proc. Natl. Acad. Sci. USA*, 89:7953-7957.
- Gudas J. M., Knight G. B. & Pardee A. B. (1990). Ordered splicing of thymidine kinase pre-mRNA during the S phase of the cell cycle. *Mol. Cell. Biol.*, 10:5591-5596.
- Guerrier-Takada C. K., Gardiner T., Marsh N., Pace N. & Altman S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35:849-857.
- Guthrie C. (1991). Messenger RNA Splicing in Yeast: Clues to Why the Spliceosome Is a Ribonucleoprotein. *Science*, 253:157-163.

- Gutman A. & Wasylyk B. (1991). Nuclear targets for transcription regulation by oncogens. *TIG*, 7:49-54.
- Gunther C. V., Nye J. A., Bryner R. S. & Graves B. J. (1990). Sequence-specific DNA binding of the proto-oncogene *ets-1* defines a transcriptional activator sequence within the long terminal repeat of the Moloney murine sarcoma virus. *Genes Dev.*, 4:667-679.
- Guyaux M., Cornelissen A. W. C. A., Pays E., Steinert M. & Borst P. (1985). *Trypanosoma brucei*: a surface antigen mRNA is discontinuously transcribed from two distinct chromosomes. *EMBO J.*, 4:995-998.
- Hamm J. & Mattaj J. W. (1990). Monomethylated Cap Structures Facilitate RNA Export from the Nucleus. *EMBO J.*, 63:109-118.
- Hao Y., Crenshaw T., Moulton T., Newcomb E. & Tycko B. (1993). Tumour-suppressor activity of H19 RNA. *Nature*, 365:764-767.
- Harley V. R., Jackson D., Hextall P., Hawkins J. R., Berkovitz G. D., Sockanathan S., Lovell-Badge R. & Goodfellow P. N. (1992). DNA binding activity of recombinant SRY from normal males and XY females. *Science*, 255:453-456.
- Hartwell L. H. (1967). Macromolecule synthesis in temperature-sensitive mutants of yeast. *J. Bacteriol.*, 93:1662-1670.
- Hatzoglou M., Sekeris C. E. & Hanson R. W. (1985). Processing of phosphoenolpyruvate carboxykinase (GTP) RNA in vivo. *Proc. Natl. Acad. Sci. USA*, 82:4346-4350.
- Hawkins J. D. (1988). A survey on intron and exon lengths. *Nucleic Acids Res.*, 16:9893-9908.
- Hawkins J. R., Taylor A., Berta P., Levilliers J., Van Der Auwere B. & Goodfellow P. N. (1992a). Mutational analysis of SRY : non-sense and missense mutations in XY sex reversal. *Hum. Genet.*, 88:471-474.
- Hawkins J. R., Taylor A., Goodfellow P. N., Migeon C. J., Smith K. D. & Berkovitz G. D. (1992b). Evidence for increased prevalence of SRY mutations in XY females with complete rather than partial gonadal dysgenesis. *Am. J. Hum. Genet.*, 51:979-984.

- Hentze M. W. (1991). Determinants and regulation of cytoplasmic mRNA stability in eucaryotic cells. *Biochem. Biophys. Acta*, 1090:281-292.
- Ho I. C., Bhat N. K., Gottschalk L. R., Lindsten T., Thompson C., Papas T. S. & Leiden J. M. (1990). Sequence-specific binding of human ETS-1 to the T-cell receptor α gene enhancer.
- Hoffman B. E. & Grabowski P. J. (1992). U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. *Genes Dev.*, 6:2554-2568.
- Hoffman M. (1993). The cell's nucleus shapes up. *Science*, 259:1257-1259.
- Huang X. Y. & Hirsh D. (1989). A second trans-spliced RNA leader sequence in the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA*, 86:8640-8644.
- Hunter T. (1991). Cooperation between oncogenes. *Cell*, 64:249-270.
- Inoue K., Ohno M., Sakamoto H. & Shimura Y. (1989). Effect of the cap structure on pre-mRNA splicing in *Xenopus* oocyte nuclei. *Genes Dev.*, 3:1472-1479.
- Jackson-Grusby L. & Kuo A. & Leder P. (1992). A variant limb deformity transcript expressed in the embryonic mouse limb defines a novel formin. *Genes Dev.*, 6:29-37.
- Jackson I. J. (1991). A reappraisal of non-consensus mRNA splice sites. *Nucl. Acids Res.*, 19:3795-3798.
- Jackson R. J., Howell M. T. & Kaminski A. (1990). The novel mechanism of initiation of picornavirus RNA translation. *TIBS*, 15:477-483.
- Jackson R. J. (1993). Cytoplasmic regulation of mRNA function: the importance of the 3' untranslated region. *Cell*, 74:9-14.
- Jarrell K. A. (1993). Inverse splicing of a group II intron. *Proc. Natl. Acad. Sci. USA*, 90:8624-8627.

- Jeffreys J. & Flavell R. A. (1977). The rabbit β -globin gene contains a large insert in the coding sequence. *Cell*, 12:1097-1108.
- Jeppesen P. & Turner B. M. (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell*, 74:281-289.
- Jiménez-García L. F. & Spector D. L. (1993). In vivo evidence that transcription and splicing are coordinated by recruiting mechanism. *Cell*, 73:47-59.
- Jorcyk C. L., Watson D. K., Mavrothalassitis G. J. & Papas T. S. (1991). The human ETS-1 gene: genomic structure, promoter characterization and alternative splicing. *Oncogene*, 6:523-532.
- Junien C. & van Heyningen V. (1990). Report of the committee on the genetic constitution of chromosome 11. *Cytogenet. Cell. Genet.*, 55:153-169.
- Kalderon D., Roberts B. L., Richardson W. D. & Smith A. E. (1984). A short amino acid sequence able to specify nuclear location. *Cell*, 39:499-509.
- Karim F. D., Urness L. D., Thummel C. S., Klemsz M. J., McKercher S. R., Celada A., van Beveren C., Maki R. A., Gunther C. V., Nye J. A. & Graves B. J. (1990). The Ets-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence. *Genes Dev.*, 4:1451-1453.
- Katz S. M., Gibbons B., Tilly R., Lam G., Robertson M. E., Czepulkowski B. & Chessnells J. (1988). Cellular and molecular studies on infant null acute lymphoblastic leukemia. *Blood*, 71:1438-
- Kay G. F., Penny G. D., Patel D., Ashworth A., Brockdorff N. & Rastan, S. (1993). Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. *Cell*, 72:171-182.
- Kennedy M. A., Morris C. M. & Fitzgerald P. H. (1989). The oncogene ETS-1 is distant from the chromosome 11 breakpoint in leukaemic cells with the t(11;19) (q23;p13). *Leukemia Research*, 13:459-463.

- Kerckaert J. P., Duterque-Coquillaud M., Collyn-d'Hooghe M., Morel P., Majerus M. A., Lai J. L., Fenaux P., Bauters F., Debuire B. & Loucheux-Lefebvre M. H. (1990). Polymorphism of the proto-oncogene ETS-1 in hematological malignancies. *Leukemia*, 4:16-19.
- Kessler O., Jiang Y. & Chasin L. A. (1993). Order of intron removal during splicing endogeneous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. *Mol. Cell. Biol.*, 13:6211-6222.
- Kingsley D.M., Jenkins N.A. & Copeland N. G. (1989). A molecular genetic linkage map of mouse chromosome 9 with regional localizations for the Gsta, T3g, Ets-1 and Ldlr loci. *Genetics*, 123:165-172.
- Klempnauer K. H., Gonda T. J. & Bishop J. M. (1982). Nucleotide sequence of the retroviral leukemia gene v-myb and its cellular progenitor c-myb: the architecture of a transduced oncogene. *Cell*, 31:453-463.
- Klempnauer K. H., Symonds G., Evans, G. I. & Bishop J. M. (1984). Subcellular localization of proteins encoded by oncogenes of avian myeloblastosis virus and avian leukemia virus E26 and by the chicken c-myb gene. *Cell*, 37:537-547.
- Koizumi S., Fisher R., Fujiwara S., Jorcyk C., Bhat N., Seth A. & Papas T. (1990). Isoforms of the human ets-1 protein: generation by alternative splicing and differential phosphorylation. *Oncogene*, 5:675-681.
- Konarska M., Filipowicz W., Domdey H. & Gross H. J. (1981). Binding of ribosomes to linear and circular forms of the 5'-terminal leader fragment of tobacco mosaic virus RNA. *Eur. J. Biochem.*, 114:221-227.
- Konarska M. & Sharp P. A. (1986). Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell*, 46:845-855.
- Kooter J. M., De Lange T. & Borst P. (1984). Discontinuous synthesis of mRNA in trypanosomes. *EMBO J.*, 3:2387-2392.
- Koopman P., Münsterberg A., Capel B., Vivian N. & Lovell-Badge R. (1990). Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature*, 348:450-452.

- Koopman P., Gubbay J., Vivian N., Goodfellow P. & Lovell-Badge R. (1991). Male development of chromosomally female mice transgenic for SRY. *Nature*, 351:117-121.
- Kourilsky P. & Chambon P. (1978). The ovalbumin gene: an amazing gene in 8 pieces. *TIBS*, 3:244-
- Kozak M. (1979). Inability of circular mRNA to attach to eucaryotic ribosomes. *Nature*, 280:82-85.
- Kozak M. (1992). A consideration of alternative models for the initiation in eucaryotes. *Crit. Rev. Bioch. Mol. Biol.*, 27:385-402.
- Krainer A. R., Conway G.C. & Kosack D. (1990). The essential pre-mRNA splicing factor SF2 influences 5' selection by activation proximal sites. *Cell*, 62:35-42.
- Krause M. & Hirsh D. (1987). A trans-spliced leader sequence on actin mRNA in *C. elegans*. *Cell*, 49:753-761.
- Krawczak M., Reiss J. & Cooper D.N. (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet.*, 90:41-54.
- Kruger K., Grabowski P. J., Zaug A. J., Sands J., Gottschling D. E. & Cech T. R. (1982). Self-splicing RNA : autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, 31:147-157.
- Kuhne T., Wieringa B., Reiser J. & Weissmann C. (1983). Evidence against a scanning model for RNA splicing. *EMBO J.*, 2:727-233.
- Lamm G. M., Blencowe B. J., Sproat B. S., Iribarren A. M., Ryder U. & Lamond A. I. (1991). Antisense probes containing 2-aminoadenosine allow efficient depletion of U5 snRNP from HeLa splicing extracts. *Nucl. Acids Res.*, 19:3193-3198.
- Lamond A. I. & Gibson T. J. (1990). Catalytic RNA and the origin of genetic systems. *TIG.*, 6:145-149.

- Landschulz W. H., Johnson P. F. & McKnight S. L. (1988). The leucine zipper : a hypothetical structure common to a new class of DNA binding proteins. *Science*, 240:1759-1764.
- Lassar A. B., Buskin J. N., Lockshon D. L. & Weintraub H. (1989). MyoD is a sequence specific DNA-binding protein requiring a region of myc homology to bind to the muscle creatine kinase enhancer. *Cell*, 58:823-831.
- Laudet V., Niel C., Duterque-Coquillaud M., Leprince D. & Stehelin D. (1993). Evolution of the ets gene family. *Biochem. Biophys. Res. Communications*, 190:8-14.
- Lautenberger J. A., Burdett L. A., Gunnell M. A., Qi S., Watson D. K., O'Brien S. J. & Papas T. S. (1992). Genomique dispersal of ets gene family during metazoan evolution. *Oncogene*, 7:1713-1719.
- Laughon A. & Scott M. P. (1984). Sequence of a drosophila segmentation gene: protein structure homology with DNA-binding proteins. *Nature*, 310:25-31.
- Legrain P., Séraphin B. & Rosbash M. (1988). Early commitment of yeast pre-mRNA to the spliceosome pathway. *Mol. Cell. Biol.*, 8:3755-3760.
- Legrain P. & Choulika A. (1990). The molecular characterization of PRP6 and PRP9 yeast genes reveals a new cysteine/histidine motif common to several splicing factors. *EMBO J.*, 9:2775-2781.
- Leprince D., Gegonne A., de Taisne C., Schnuberg A., Lagrou C. & Stehelin D. (1983). A putative second cell-derived oncogene of the avian leukemia retrovirus E26. *Nature*, 306:395-397.
- Leprince D., Duterque-Coquillaud M., Li R. P., Henry C., Flourens A., Debuire B. & Stehelin D. (1988). Alternative splicing within the chicken c-ets-1 locus: implications for transduction within the E26 retrovirus of the c-ets proto-oncogene. *J. Virol.*, 62:3233-3241.
- Leprince D. (1989). A single amino-acid substitution in the DNA-binding domain of the myb oncogene confers a thermolabile phenotype to E26-transformed myeloid cells. *Oncogene Res.*, 5:137-141.

- Leprince D., Guesquiere J. C. & Stehelin D. (1990). The chicken cellular progenitor of the v-ets oncogene, p68c-ets-1, is a nuclear DNA-Binding protein not expressed in lymphoid cells of the spleen. *Oncogene Res.*, 5:255-265.
- Leprince D., Crepieux P. & Stehelin D. (1992). c-ets-1 DNA binding to the PEA3 motif is differentially inhibited by all the mutations found in v-ets. *Oncogene*, 7:9-17.
- Leprince D., Crepieux P., Laudet V., Flourens A. & Stehelin D. (1993). A New Mechanism of Oncogenic Activation: E26 Retroviral v-ets Oncogene Has Inverted the C-Terminal End of the Transcription Factor c-ets-1. *Virology*, 194:855-857.
- Lewin B. (1980). Alternatives for splicing : recognizing the ends of introns. *Cell*, 22:324-326.
- Liao X. C., Colot H. V., Wang Y. & Rosbash M. (1992). Requirements for U2 snRNP addition to yeast pre-mRNA. *Nucl. Acids Res.*, 20:4237-4245.
- Lin R. J., Lustig A. & Abelson J. (1987). Splicing of yeast nuclear pre-mRNA in vitro requires a functional 40S spliceosome and several extrinsic factors. *Genes Dev.*, 1:7-18.
- Lüscher B. & Eisenman R. N. (1990). New light on Myc and Myb. Part II. Myb. *Gen. Dev.*, 4:2235-2241.
- Lustig A. J., Lin R. J. & Abelson J. (1986). The yeast RNA gene products are essential for mRNA splicing in vitro. *Cell*, 47:953-963.
- Lutz-Freyermuth C., Query C. C. & Keene J. D. (1990). Quantitative determination that one of two potential RNA binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to the stem-loop II of U1 RNA. *Proc. Natl. Acad. Sci. USA*, 87:6393-6397.
- Lygerou Z., Kandels-Lewis B. & Séraphin B. (1993). Le rôle des snRNP dans l'épissage des ARN pré-messagers. *Médecine Sciences*, 9:165-170.
- Macejak D.G. & Sarnow P. (1991). Internal initiation of translation mediated by the 5' leader of a cellular mRNA. *Nature*, 353:90-94.

- Macleod K., Leprince D & Stehelin D. (1992). The ets gene family. *TIBS*, 17:251-256.
- Majérus M. A., Bibollet-Ruche F., Telliez J.B., Wasyluk B. & Bailleul B. (1992). Serum, AP-1 and Ets-1 stimulate the human ets-1 promoter. *Nucleic Acids Res.*, 20:2699-2703.
- Mancebo R., Lo P. C. H. & Mount S. M. (1990). Structure and expression of the *Drosophila melanogaster* gene for the U1 small nuclear ribonucleoprotein particle 70K protein. *Mol. Cell. Biol.* 10:2492-2502.
- Maniatis T. (1991). Mechanisms of alternative pre-mRNA splicing. *Science*. 251:33-34
- Manley J. (1993). Question of commitment. *Nature*, 365:81-85.
- Mattaj L. W. (1993). RNA recognition a family matter? *Cell*, 73:837-840.
- Mayeda A. & Krainer A. R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell*, 69:365-375.
- McCarrey J. R. & Dilworth D. D. (1992). Expression of XIST in mouse germ cells correlates with X-chromosome inactivation. *Nature Gen.*, 2:200-203.
- McElreavey K., Vilain E, Cotinot C., Payen E. & Fellous M. (1993). Control of sex determination. *Eur. J. Biochem.*, sous presse.
- McKeown M. (1992). Alternative mRNA splicing. *Annu. Rev. Cell Biol.*, 8:133-155.
- Mermoud J. E., Cohen P. & Lamond A. I. (1992). Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. *Nucl. Acid Res.*, 20:5263-5269.
- Metz T. & Graf T. (1991a). v-myb and v-ets transform chicken erythroid cells and cooperate both in trans and cis to induce distinct differentiation phenotypes. *Genes Dev.*, 5:369-380.
- Metz T. Graf T. (1991b). Fusion of nuclear oncoprotein v-myb and v-ets is required for the leukemogenicity of E26 virus. *Cell*, 66:95-105.
- Michaud S. & Reed R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes Dev.*, 5:2534-2546.

- Michel F., Jacquier A. & Dujon B. (1982). Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie*, 64:867-881.
- Milhausen M., Nelson R. G., Sather S., Selkirk M. & Agabian N. (1984). Identification of a small RNA containing the trypanosome splicing leader: a donor of the shared 5' sequences of trypanosomatid mRNAs? *Cell*, 38:721-729.
- Moelling K., Pfaff E., Beug H., Beimling P., Bunte T., Schaller M. E. & Graf T. (1985). DNA binding activity is associated with purified Myb proteins from AMV and E26 viruses and is temperature sensitive for E26 ts mutants. *Cell*, 40:983-990.
- Moore M. J., Query C. C. & Sharp P. A. (1993). Splicing of precursors to mRNA by the spliceosome. *The RNA World*, ed. Gesteland R. F. & Atkins J. F., Cold Spring Harbor Laboratory Press.
- Moscovici C., Samarut J., Gazzolo L. & Moscovici M. G. (1981). Myeloid and erythroid nonplastic responses to avian defective leukemia viruses in chickens and in quail. *Virology*, 113:765-768.
- Moscovici M. G., Jurdic P., Samarut J., Gazzolo L., Mura C. V. & Moscovici C. (1983). Characterisation of the hematopoietic target cells for the avian leukemia virus E26. *Virology*, 129:65-78.
- Mount S. M. (1982). A catalogue of splice junction sequences. *Nucl. Acids Res.*, 10:459-472.
- Mullen M. P., Smith C. W. J., Patton J. G. & Nadal-Ginard B. (1991). α -Tropomyosin mutually exclusive exon selection : competition between branchpoint/polypyrimidine tracts determines default exon choice. *Genes Dev.*, 5:642-655.
- Murre C., McCaw P.S. & Baltimore D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding daughterless, MyoD, and myc proteins. *Cell*, 56:777-783.
- Nasim F. H., Spear P. A., Hoffman H. M., Kuo H. & Grabowski P. J. (1990). A sequential splicing mechanism promotes selection of an optional exon by repositioning a downstream 5' splice site in preprotachykinin pre-mRNA. *Genes Dev.*, 4:1172-1184.

- Nelson K. J., Haimovich J. & Perry R. P. (1983). Characterization of productive and sterile transcripts from the immunoglobulin heavy-chain locus: processing of μ_m and μ_s mRNA. *Mol. Cell. Biol.*, 3:1317-1332.
- Newman A. J. & Norman C. (1992). U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell*, 68:743-754.
- Nigro J. N., Cho K. R., Fearon E. R., Kern S. C., Ruppert J. M., Oliner J. D., Kinzler K. W. & Vogelstein B. (1991). Scrambled exons. *Cell*, 64: 607-613.
- Niwa M. & Berget S. M. (1991). Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes Dev.*, 5:2086-2095.
- Niwa M., McDonald C. C. & Berget S. M. (1992). Are vertebrate exons scanned during splice-site selection? *Nature*, 360:277-280.
- Noller H. F., Hoffarth V. & Zimniak L. (1992). Unusual resistance of peptidyl transferase to protein extraction procedures. *Science*, 246:1416-1419.
- Nunn M. F., Sieburg P. H., Moscovici C. & Duesberg P. H. (1983). Tripartite structure of the avian erythroblastis virus E26 transforming gene. *Nature*, 306:391-395.
- Nye J.A., Petersen J. M., Gunther C. V., Jonsen M. D. & Graves B.J. (1992). Interaction of murine Ets-1 with GGA binding sites establishes the Ets domain as a new DNA binding motif. *Genes Dev.*, 6:975-990.
- Ohshima Y. & Gotoh Y. (1987). Signals for the selection of a splice site in pre-mRNA. *J. Mol. Biol.*, 195:247-259.
- Ohyashiki K., Ohyashiki J. H., Sandberg A. A., Minowada J. & Kinniburgh A. J. (1987). Loss of Hu-ets-1 allele in human leukemia cell lines ML-1, -2 and -3 with a chromosome change at 11q24. *Cancer Res.*, 47:3842-3846.
- Pachnis V., Brannan C. I. & Tilghman S. M. (1988). The structure and expression of a novel gene activated in early mouse embryogenesis. *Embo J.*, 7:673-681.

- Padgett R. A., Grabowski P. J., Konarska M. M. & Sharp P. A. (1985). Splicing messenger RNA precursors: branch sites and lariat RNAs. *TIBS*, 10:154-157.
- Parker R., Siliciano P. G. & Guthrie C. (1987). Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell*, 229-239.
- Patterson B. & Guthrie C. (1987). An essential yeast snRNP with a U5-like domain is required for slicing *in vivo*. *Cell*, 49:613-624.
- Patton J. G., Mayer S. A., Tempst P. & Nadal-Ginard B. (1991). Characterization and molecular cloning of polypyrimidine tract-binding protein : a component of a complex necessary for pre-mRNA splicing. *Genes Dev.*, 5:1224-1236.
- Perkins K. K., Furneaux H. M. & Hurwitz I. (1986). RNA splicing products formed with the isolated fractions from Hela cells are associated with fast-sedimentating complexes. *Proc. Natl. Acad. Sci. USA*, 83:2387-2392.
- Perlman P. S., Peebles C. L. & Daniels C. (1990). Different types of introns and splicing mechanisms. *Intervening Sequences in Evolution and Development*, ed. Stone E. M. & Schwartz R. J., Oxford University Press.
- Piccirilli J. A., McConnell T. S., Zaug A. J., Noller H. F. & Cech T. R. (1992). Aminoacyl esterase activity of the *Tetrahymena* ribozyme. *Science*, 256:1420-1424.
- Pinto A. L. & Steitz J. A. (1989). The mammalian analogue of the yeast PRP8 protein is present in the U4/5/6 small ribonucleoprotein particle and the spliceosome.
- Pognonec P., Boulukos K. E., Gesquière J. C., Stéhelin D. & Ghysdael J. (1988). Mitogenic stimulation of thymocytes results in the calcium dependent phosphorylation of c-ets-1 proteins. *Embo J.*, 7:977-983.
- Pognonec P., Boulukos K. E. & Ghysdael J. (1989). The c-ets-1 protein is chromatin associated and binds to DNA in-vitro. *Oncogene*, 4:691-697.
- Pognonec P., Boulukos K. E., Bosselut R., Boyer C., Schmitt-Verhulst A. M. & Ghysdael J. (1990). Identification of a ETS-1 variant protein unaffected in its chromatin and in vitro DNA binding capacities by T cell antigen receptor triggering and intracellular calcium rises. *Oncogene*, 5:603-610.

- Pribyl L. J., Watson D. K., McWilliams M. J., Ascione R. & Papas T. S. (1988). The *Drosophila* ets-2 gene : molecular structure, chromosomal localization, and developmental expression. *Dev. Biol.*, 127:43-53.
- Price J., Engberg J. & Cech T. (1987). 5' exon requirement for self-splicing of the *Tetrahymena thermophila* pre-ribosomal RNA and identification of a cryptic 5' splice site in the 3' exon. *J. Mol. Biol.*, 196:49-60.
- Puttaraju M. & Been M. D. (1992). Group I permuted intron-exon (PIE) sequences self-splice to produce circular exons. *Nucl. Acid. Res.*, 20:5357-5364.
- Puttaraju M., Perrotta A. T. & Been M.D. (1993). A circular trans-acting hepatitis delta virus ribozyme. *Nucleic. Acids. Res.*, 21:4253-4258.
- Query C. C., Bentley R. C. & Keene J. D. (1989). A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell*, 57:89-101.
- Queva C., Leprince D., Stéhelin D. & Vandebunder B. (1993). p54c-ets-1 p68c-ets-1, the two transcription factors encoded by the c-ets-1 locus, are differentially expressed during the development of the chick embryo. *Oncogene*, 8:2511-2520.
- Radke K., Beug H., Kornfeld S. Graf T. (1982). Transformation of both erythroid and myeloid cells by E26, an avian leukemia virus that contains the myb gene. *Cell*, 31:643-653.
- Rastinejad F. & Blau H. M. (1993). Genetic complementation reveals a novel regulatory role for 3' untranslated regions in growth and differentiation. *Cell*, 72:903-917.
- Reedy E. S. P. & Rao V.N. (1988). Structure expression and alternative splicing of the human c-ets-1 proto-oncogene. *Oncogene Res.*, 3:239-246.
- Reves R. H., O'Hara B. F., Pavan W. J., Gearhart J. D. & Haller O. (1988). Genetic mapping of the Mx influenza virus resistance gene within the region of mouse chromosome 16 that is homologous to the human chromosome 21. *J. Virol.*, 62:4372-4375.
- Reich C. I., Hoy G. J., Porter G. L. & Wise J. A. (1992). Mutations at the 3' splice site can be suppressed by compensatory base changes in U1 snRNA in fission yeast. *Cell*, 69:1159-1169.

- Reyes A. A., Small S. J. & Akeson R. (1991). At least 27 alternatively spliced forms of the Neuronal Cell Adhesion Molecule mRNA are expressed during rat heart development. *Mol. Cell. Biol.*, 11:1654-1661.
- Richler C., Soreq H. & Wahrman J. (1992). X inactivation in mammalian testis is correlated with inactive X-specific transcription. *Nature Gen.*, 2:192-195.
- Robberson B. L., Cote G. J. & Berget S. M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.*, 10:84-94.
- Rosbash M., Harris P. K. W., Woolford J. L. & Teem J. L. (1981). The effect of temperature-sensitive mutants on the transcription products from cloned ribosomal protein genes of yeast. *Cell*, 24:679-686.
- Rosbash M. & Seraphin B. (1991). Who's on first? The U1 snRNP-5' splice site interaction and splicing. *TIBS*, 187-190.
- Rosson D. & Reddy P. (1986). Nucleotide sequence of chicken c-myc complementary DNA and implications for myc oncogene activation. *Nature*, 319:604-606.
- Ruby S. & Abelson J. (1988). An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. *Science*, 242:1028-1035.
- Ruby S. & Abelson J. (1991). Pre-mRNA splicing in yeast. *TIG*, 7:79-85.
- Ruskin B. & Green M. R. (1985). An RNA processing activity that debranches RNA lariats. *Science*, 229:135-140.
- Ruskin B., Zamore P.D. & Green M. R. (1988). A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell*, 52:207-219.
- Sacchi N., Watson D. K., Van Kessel Ad H. M. G., Hagemeyer A., Kersey J., Drabkin H. D., Patterson D. & Papas T. S. (1986). Hu-ets-1 and Hu-ets-2 genes are transposed in acute leukemias with (4;11) and (8;21) translocations. *Science*, 231:379-382.
- Sacchi N., de Klein A., Showalter S. D., Bigi G. & Papas T. S. (1988). High expression of ets-1 gene in human thymocytes and immature T leukemic cells. *Leukemia*, 2:12-18.

- Saikumar P., Murali R. & Reddy E. S. P. (1990). Role of tryptophan repeats and flanking amino acids in Myb-DNA interactions. *Proc. Natl. Acad. Sci. USA*, 87:8452-8456.
- Salido E. C., Yen P. H., Mohandas T. K. & Shapiro L. J. (1992). Expression of the X-inactivation-associated gene XIST during spermatogenesis. *Nature Gen.*, 2:196-199.
- Sass H. & Pederson T. (1984). Transcription dependent localization of U1 and U2 small nuclear ribonucleoproteins at major sites of gene activity in polytene chromosomes. *J. Mol. Biol.*, 180:911-926.
- Sassanfar M. & Szostak J. W. (1993). An RNA motif that binds ATP. *Nature*, 364:550-554.
- Sawa H. & Shimura Y. (1992). Association of U6 snRNA with the 5'-splice site region of pre-mRNA in the spliceosome. *Genes Dev.*, 6:244-254.
- Scherly D., Boelens W., van Venrooij W. J., Dathan N. A., Hamm J. & Mattaj I. W. (1989). Identification of the RNA binding segment of human U1 a protein and definition of its binding site on U1 snRNA. *Embo J.*, 8:4163-4170.
- Schmid S. R. & Linder P. (1992). D-E-A-D protein family of putative RNA helicases. *Mol Microbiol.*, 6:283-292.
- Schwer B. & Guthrie C. (1991). PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature*, 349:494-499.
- Senepathy P., Shapiro M. B. & Harris N. L. (1990). Splice junction branch point site and exon: sequence statistics, identification and applications to genome projects. *Methods Enzymol.*, 183:252-278.
- S raphin B., Kretzner L. & Rosbash M. (1988). A U1 snRNA : pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *Embo. J.*, 7:2533-2538.
- S raphin B. & Rosbash M. (1989). Identification of functional U1 snRNA pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell*, 59:349-358.

- S raphin B. & Rosbash M. (1990). Exon mutations uncouple 5' splice site selection from U1 snRNA pairing. *Cell*, 63:619-629.
- S raphin B., Abovich N. & Rosbach M. (1991a). Genetic depletion indicates a late role for U5 snRNP during in vitro spliceosome assembly. *Nucl. Acids Res*, 19:3857-3860.
- S raphin B. & Rosbash M. (1991b). The yeast branchpoint sequence is not required for the formation of a stable U1 snRNA-pre-RNA complex and is recognized in the absence of U2 snRNA. *Embo. J.*, 10:1209-1216.
- Seth A. & Papas T.S. (1990). The c-ets-1 proto-oncogene has oncogenic activity and is positively autoregulated. *Oncogene*, 5:1761-1768.
- Seth A., Ascione R., Fisher R. J., Mavrothalassitis G. J., Bhat N. K. & Papas T. S. (1992). The ets gene family. *Cell Growth & Differentiation*, 3:327-334.
- Shannon K. W. & Guthrie C. (1991). Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA-binding motifs. *Genes Dev.*, 5:773-785.
- Shapiro M. B. & Senepathy P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.*, 15:7155-7174.
- Sharp P. A., Gallimore P. H. & Flint S. J. (1974). Mapping of adenovirus 2 RNA sequences in lytically infected cells and transformed cell lines. *Cold Spring Harbor Symp. Quant. Biol.*, 39:457-474.
- Sharp P. A. (1987). Splicing of Messenger RNA Precursors. *Science*, 235: 766-771.
- Shimizu A., Nussenzweig M. C., Mizuta T. R., Leder P. & Honjo, T. (1989). Immunoglobulin double-isotype expression by trans-mRNA in a human immunoglobulin transgenic mouse. *Proc. Natl. Acad. Sci. USA*, 86:8020-8023.
- Shipley J. & Sheer D. (1989). Deletion of c-ets-1 and T3 loci from the 11q chromosome in the human monoblastic cell line U937. *Leukemia Research*, 13:445-450.

- Sinclair A. H., Berta P., Palmer M. S., Hawkins J. R., Griffiths B. L., Smith M. J., Foster J. W., Frischauf A. M., Lovell-Badge R. & Goodfellow P. N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, 346:240-244.
- Smith C. W. J., Patton J. G. & Nadal-Ginard B. (1989). Alternative splicing in the control of gene expression. *Annu. Rev. Genet.*, 23:527-577.
- Spector D. L., Fu X. D. & Maniatis T. (1991). Associations between distinct pre-mRNA splicing components and the cell nucleus. *Embo J.*, 10:3467-3481.
- Starzyk R. M. (1986). Prokaryotic mRNA processing. *TIBS*, 11:60.
- Steitz J. A., Black D. L. & Gerke V. (1988). Functions of the abundant U-snRNP. In : Birnstiel M. L., ed. *Structure and Functions of Major and Minor Small Ribonucleoprotein Particles*. Berlin : Springer-Verlag , 115-154.
- Steitz J. A. (1992). Splicing Takes a Holliday. *Science*, 257:888-889.
- Symons R. (1989). Self-cleavage of RNA in the replication of small pathogens of plants and animals. *T. I. B. S.*, 14:445-450.
- de Taisne C., Gegonne A. & Stehelin, D. (1984). Chromosomal localization of the human proto-oncogene c-ets. *Nature*, 310:581-583.
- Tazi T., Alibert C. & Temsamani J. (1986) A protein that specifically recognizes the 3' splice site of mammalian pre-mRNA introns is associated with a small ribonucleoprotein. *Cell*, 47:555-766.
- Thayer M. J., Tapscott S. J., Davis R. L., Wright W. E., Lassar A. B. & Weintraub H. (1989). Positive autoregulation of the myogenic determination of gene MyoD1. *Cell*, 58:241-248.
- Trumpp A., Blundell P. A., Luis de la Pompa J. & Zeller R. (1992). The chicken limb deformity gene encodes nuclear proteins expressed in specific cell types during morphogenesis. *Genes Dev.*, 6:14-28.

- Tsai M.J., Ting A.C., Nordstrom J.L., Zimmer W. & O'Malley B.W. (1980). Processing of high molecular weight ovalbumin and ovomucoid precursor RNAs to messenger RNA. *Cell*, 22:219-230.
- Utans U. & Kramer A. (1990). Splicing factor SF4 is dispensable for the assembly of a functional splicing complex and participates in subsequent steps of the splicing reaction. *EMBO J.*, 9:4119-4126.
- Utans U., Behrens S. E., Lührmann R, Kole R. & Krämer A. (1992). A splicing factor that is inactivated during in vivo heat shock is functionally equivalent to the (U4/U6.U5) triple snRNP-specific proteins. *Genes Dev.*, 6: 631-641.
- Vandenbunder B., Pardanaud L., Jaffredo T., Mirabel M. A. & Stehelin D. (1989). Complementary patterns of expression of c-ets1, c-myb and c-myc in the blood-forming system of the chick embryo. *Development*, 106:265-274.
- Vellard M., Soret J. & Viegas-Péquignot E. (1991). c-myb proto-oncogene: evidence for intermolecular recombination of coding sequences. *Oncogene*, 6:505-514.
- Vellard M., Sureau A., Soret J., Martinerie C. & Perbal B. A. (1992). A potential splicing factor is encoded by the opposite strand of the trans-spliced c-myc exon. *Proc. Natl. Acad. Sci. USA*, 89:2511-2515.
- Wassarman D. A. & Steitz J. A. (1991). Alive with DEAD proteins. *Nature*, 349:463-464.
- Wasylyk C., Wasylyk B., Heidecker G., Huleihel M. & Rapp U. R. (1989). Expression of rat oncogene activates the transcription factor PEA1. *Mol. Cell. Biol.*, 9:2247-2250.
- Wasylyk B., Wasylyk C., Flores P., Begue A., Leprince D. & Stehelin D. (1990). The c-ets proto-oncogene encode transcription factors that cooperate with c-Fos and c-Jun for transcriptional activation. *Nature*, 346:191-193.
- Wasylyk C., Gutman A., Nicholson R. & Wasylyk B. (1991). The c-Ets oncoprotein activates the stromelysin promoter through the same elements as several non-nuclear oncoproteins. *Embo J.*, 10:1127-1134.

- Wasylyk C., Kerckaert J.P. & Wasylyk B. (1992). A novel modulator domain of Ets transcription factors. *Genes Dev.*, 6:965-974.
- Wasylyk B., Hahn S. L. & Giovane A. (1993). The Ets family of transcription factors. *Eur. J. Biochem.*, 211:7-18.
- Watakabe A., Tanaka K. & Shimura Y. (1993). The role of exon sequences in splice site selection. *Genes Dev.*, 7:407-418
- Watson D.K., McWilliamy-Smith M.J., Nunn M. F., Duesberg P.H., O'Brien S.J. & Papas T.S. (1985). The ets sequences from the transforming gene of avian erythroblastosis virus E26, has unique domains on human chromosome 11 and 21: Both loci are transcriptionally active. *Proc. Natl. Acad. Sci. USA*, 82:7294-7298.
- Watson D. K., McWilliams M. J., Lapis P., Lautenberger J. A., Schweinfest C. W. & Papas T. S. (1988). Mammalian ets-1 and ets-2 genes encode highly conserved proteins. *Proc. Natl. Acad. Sci. USA*, 85:7862-7866.
- Weil D., Brosset S. & Dautry F. (1990). RNA processing is a limiting step for murine tumor necrosis factor β expression in response to interleukin-2. *Mol. Cell. Biol.*, 10:5865-5875.
- Weiner A. M. (1993). mRNA splicing and autocatalytic introns : distant cousins or the products of chemical determinism? *Cell*, 72:161-164.
- Wellauer P. K. & Dawid I. B. (1977). The structural organization of ribosomal DNA in *Drosophila melanogaster*. *Cell*, 10:193-212.
- Wernert N., Raes M. B., Lassale P., Dehouck M. P., Gosselin B., Vandebunder B. & Stéhelin D. (1992). The c-ets-1 proto-oncogene is a transcription factor expressed in endothelial cells during tumor vascularisation and other forms of angiogenesis in man. *Am. J. Pathology*, 140:119-127.
- White R. L. & Hogness D. S. (1977). R loop mapping of 18S and 28S sequences in the long and short repeating units of *Drosophila melanogaster* rDNA. *Cell*, 10:177-192.
- Witkowski J. A. (1988). The discovery of "split" genes : a scientific revolution. *TIBS.*, 13:110-113.

- Woods D.B., Ghysdael J. & Owen M.J. (1992). Identification of the nucleotide preferences in DNA sequences recognised specifically by c-Ets-1 protein. *Nucl. Acids Res.*, 4:699-704.
- Woychik R. P., Maas R. L., Zeller R., Vogt T. F. & Leder P. (1990). "Forming" : proteins deduced from the alternative transcripts of the limb deformity gene. *Nature*, 346:850-855.
- Xing Y., Johnson C. V., Dobner P. R. & Lawrence J. B. (1993). Higher level organization of individual gene transcription and RNA splicing. *Science*, 259:1326-1330.
- Yunis J. J. & Brunning R. D. (1986). Prognostic significance of chromosomal abnormalities in acute leukemias and myelodysplastic syndromes. *Clin. Haematol.*, 15:597-620.
- Zahler A. M., Lane W. S., Stolk J. A. & Roth M. B. (1992). SR proteins : a conserved family of pre-mRNA splicing factors. *Genes Dev.*, 6:837-847.
- Zahler A. M., Neugebauer K. M., Lane W. S. & Roth M. B. (1993). Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science*, 260:219-222.
- Zamore P. D. & Green M. R. (1991). Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *EMBO J.*, 10:207-214.
- Zamore P. D., Patton J. G. & Green M. R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. *Nature*, 355:609-614.
- Zeitlin S. & Efstratiadis A. (1984). In vivo splicing products of the rabbit β globin pre-mRNA. *Cell*, 39:589-602.
- Zhang M., Zamore P. D., Carmo-Fonseca M., Lamond A. I. & Green M. R. (1992). Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc. Natl. Acad. Sci. USA*, 89:8769-8773.
- Zhuang Y. & Weiner A. M. (1986). A compensatory base change in U1 snRNA Suppresses a 5' splice site mutation. *Cell*, 46:827-835.

Ziemen-Van Der Poel S., McCabe N., Gill H. J., Espinosa R., Patel Y., Harden A., Rubinelli P., Smith S. D., Le Beau M. M., Rowley J. D. & Diaz M. O. (1991). Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias. *Proc. Natl. Acad. Sci. USA.*, 88:10735-10739.

Zieve G. W. & Sauterer R. A. (1990). Cell biology of the snRNP particles. *Biochem. Mol. Biol.*, 25:1-46.