

## Épissage en trans du proto-oncogène *c-myb* et différenciation hématopoïétique

Le gène *c-myb* code pour une famille de produits jouant un rôle important dans la prolifération et la différenciation des cellules hématopoïétiques normales. Les propriétés biologiques des différents produits de *c-myb* dépendent probablement de leur liaison à des séquences d'ADN spécifiques et sont modulées par plusieurs domaines régulateurs. L'activité transactivatrice de *c-myb* est contrôlée négativement par un domaine aminoproximal codé par des séquences (ET) qui ne sont pas situées sur le même chromosome que celui qui porte le locus *c-myb*, aussi bien chez la poule que chez l'homme. Les ARNm qui contiennent ET pourraient donc provenir d'un épissage en *trans* d'ARN précurseurs exprimés à partir d'unités transcriptionnelles physiquement distinctes. Le brin complémentaire du locus ET est codant et dirige la synthèse d'un facteur d'épissage qui pourrait participer directement au processus de recombinaison intermoléculaire. Ce faisant, les deux brins du locus ET interviendraient dans la modulation des propriétés transactivatrices de *c-myb* et seraient impliqués dans la régulation de la différenciation hématopoïétique.

---

**Bernard Perbal**

---

### ADRESSE

B. Perbal : professeur, docteur ès sciences, directeur du laboratoire. Laboratoire d'oncologie virale et moléculaire, bâtiment 110, Institut Curie, 91405 Orsay et UFR de biochimie, université Paris-VII, 2, place Jussieu, 75251 Paris cedex 05, France.

**L**a découverte de séquences introniques dans la grande majorité des gènes d'eucaryotes supérieurs a immédiatement posé toute une série de questions concernant la nature et la régulation des mécanismes assurant l'épissage des ARN pré-messagers.

Les différentes étapes de la maturation des ARN pré-messagers étant décrites dans plusieurs revues récentes [1-4], seules les grandes lignes de ce processus sont décrites ici (figure 1, p. 551).

Les réactions d'épissage se déroulent dans le noyau des cellules et font appel à la fois à des signaux nucléotidiques et à des édifices ribonucléoprotéiques complexes, les splicéosomes [5-7], constitués par de petits ARN nucléaires de la série U et toute une batterie de facteurs protéiques.

### Signaux et facteurs nécessaires à l'épissage

Les signaux requis en *cis* peuvent être subdivisés en trois groupes : d'une

part, les deux jeux de séquences permettant l'identification des extrémités 5' et 3' - proximales de l'intron et, d'autre part, les séquences nécessaires à la reconnaissance du « point de branchement » impliqué dans la formation de l'intermédiaire d'épissage en lasso. Notre connaissance des séquences nucléotidiques délimitant la jonction exon/intron a essentiellement résulté de la comparaison des sites d'épissage d'un grand nombre d'exons [8, 9], la dernière publication en date [10] faisant état d'une étude portant sur plusieurs milliers de sites d'épissage. De telles approches ont permis d'établir des séquences consensus pour les sites 5' donneur et 3' accepteur d'épissage (Tableau I) et de mettre en évidence l'importance d'un motif riche en pyrimidines, situé en amont du site accepteur. Cependant, un examen récent des séquences contenues dans les banques de données a montré qu'un certain nombre d'introns ne contiennent pas les séquences consensus à leurs extrémités 5' et 3' [11], ce qui suggère que d'autres paramètres soient certainement impliqués dans la reconnaissance de ces sites.

La sélection des sites d'épissage requiert également l'intervention de plusieurs autres facteurs protéiques [12] qui ne sont pas engagés dans des complexes ribonucléoprotéiques [13-17] et qui ont pu être identifiés grâce à la mise au point de systèmes d'épissage *in vitro* à partir d'extraits nucléaires de cellules HeLa [18].

L'un de ces facteurs, U2AF (*U2-auxiliary factor*) est nécessaire pour que la ribonucléoprotéine contenant l'ARN U2 (U2 snRNP) puisse se lier au point de branchement situé en amont du site d'épissage 3' [16-19]. Le facteur U2AF est constitué de deux polypeptides dont les poids moléculaires apparents sont 35 et 65 kDa. Seul celui de 65 kDa semble nécessaire pour la liaison à la ribonucléoprotéine U2 et l'épissage *in vitro* [20]. Le rôle du polypeptide de 35 kDa n'est pas élucidé.

Le facteur IBP (*intron binding protein*) serait associé à l'ARN U5 et pourrait se lier spécifiquement au site d'épissage 3' lors de l'assemblage du spliceosome. C'est une protéine dont le poids moléculaire est de 70 ou 100 kDa [21, 22].

Enfin, le facteur initialement dénommé ASF (*alternative splicing factor*) et SF2 (*splicing factor 2*) par deux groupes indépendants existe sous différentes formes dont le poids moléculaire varie entre 30 et 35 kDa [23, 24]. Il est requis lors de la coupure du site 5' et de la formation de la structure en lasso [25, 26].

### Épissage alternatif et épissage en trans

Si l'épissage ordonné des multiples introns contenus dans certains gènes nécessite la mise en œuvre de mécanismes extrêmement précis pour la reconnaissance des sites d'excision et de ligature, il est très vite apparu que des modes alternatifs d'épissage sont utilisés chez tous les métazoaires. Bien que l'on ne connaisse pas encore les règles qui déterminent l'ordre

dans lequel la ligature des différents exons est effectuée, il est maintenant bien établi que la production de plusieurs produits à partir d'un gène unique peut constituer un niveau supplémentaire de régulation de l'expression génétique. Ainsi, l'épissage alternatif d'ARN pré-messagers, gouverne l'expression spécifique de tissu de nombreux gènes au cours de la différenciation cellulaire et de l'embryogenèse [4, 27-29]. Un des exemples les mieux étudiés actuellement concerne la biosynthèse de protéines régulatrices à des stades précis de la détermination du sexe chez la drosophile [30-35] (*figure 2, p. 552*). Il semble que tous les moyens aient été utilisés dans les cellules d'eucaryotes supérieurs pour engendrer différents types d'ARN messagers à partir de séquences uniques [4]. Outre l'existence de donneurs et accepteurs

Tableau I	
SIGNAUX CONSENSUS D'ÉPISSAGE	
<b>Signaux définissant le site d'épissage 5'-proximal</b>	
vertébrés :	A <sub>33</sub> /C <sub>40</sub> A <sub>63</sub> G <sub>79</sub> G <sub>100</sub> T <sub>100</sub> A <sub>63</sub> A <sub>73</sub> G <sub>80</sub> T <sub>49</sub> <sup>*</sup>
dicotylédones :	A <sub>39</sub> /C <sub>31</sub> A <sub>63</sub> G <sub>79</sub> G <sub>100</sub> T <sub>100</sub> A <sub>63</sub> A <sub>73</sub> G <sub>80</sub> T <sub>49</sub> <sup>*</sup>
monocotylédones :	A <sub>42</sub> A <sub>63</sub> G <sub>79</sub> G <sub>100</sub> T <sub>100</sub> A <sub>63</sub> A <sub>73</sub> G <sub>80</sub> T <sub>49</sub> <sup>*</sup>
levures :	G <sub>100</sub> T <sub>100</sub> A <sub>100</sub> T <sub>86</sub> G <sub>100</sub> T <sub>95</sub> <sup>*</sup>
<b>Signaux définissant le site d'épissage 3'-proximal</b>	
vertébrés :	(Py) <sub>n</sub> NC <sub>77</sub> A <sub>100</sub> G <sub>100</sub> G <sub>55</sub> <sup>*</sup>
dicotylédones :	(T) <sub>n</sub> GPyA <sub>100</sub> G <sub>100</sub> G <sub>57</sub> T <sub>47</sub> <sup>*</sup>
monocotylédones :	(Py) <sub>n</sub> G <sub>48</sub> C <sub>84</sub> A <sub>100</sub> G <sub>100</sub> G <sub>59</sub> T <sub>40</sub> <sup>*</sup>
levures :	T <sub>100</sub> A <sub>100</sub> G <sub>100</sub> <sup>*</sup>

Les nombres indiquent la fréquence avec laquelle les nucléotides sont trouvés à cette position.

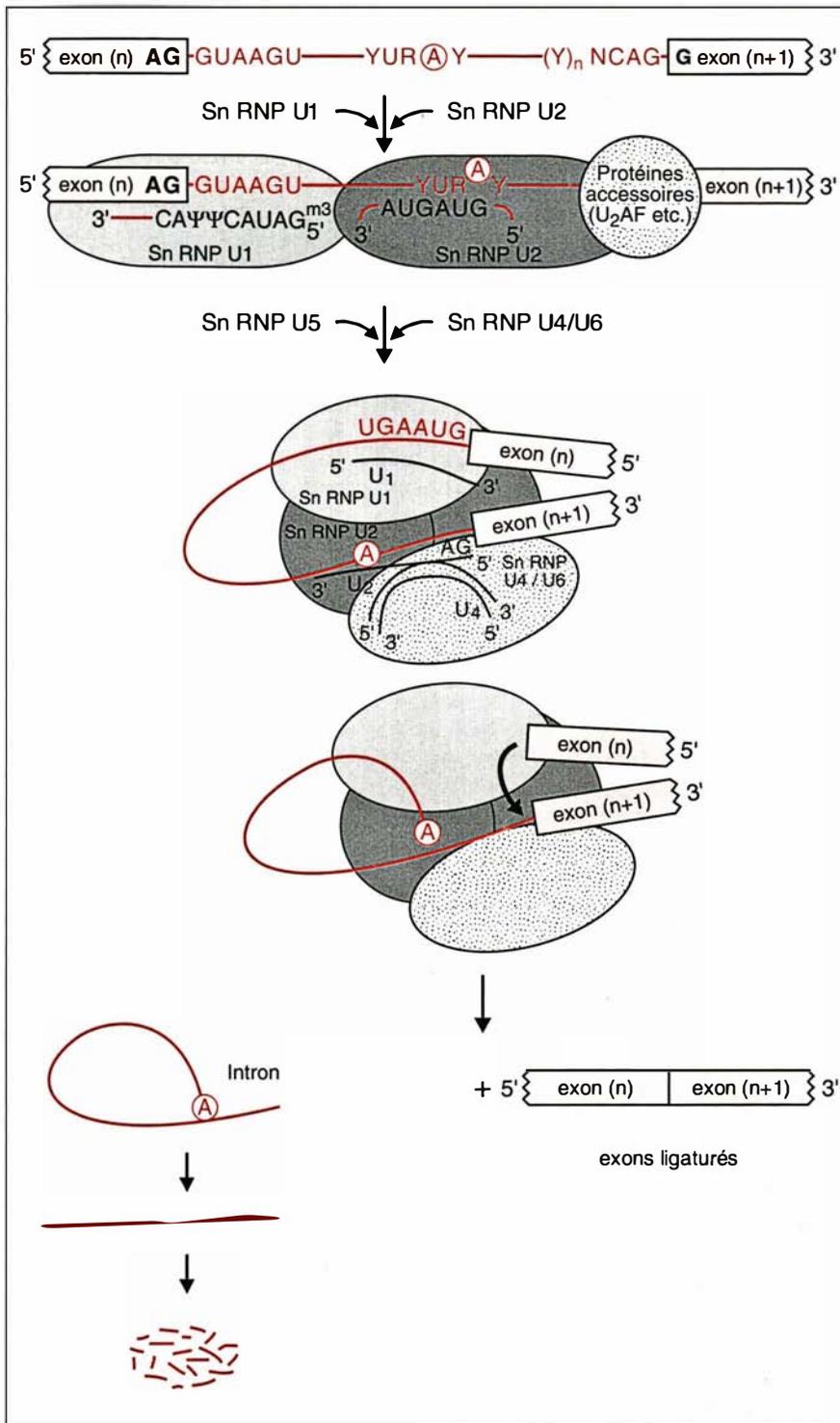
Le G surmonté d'un astérisque indique les bornes de l'intron.

Dans tous les exemples utilisés pour définir ces consensus, les introns sont encadrés au niveau de l'ADN par les dinucléotides GT et AG.

## RÉFÉRENCES

1. Krainer AR, Maniatis T. RNA splicing. In : Hames BD, Glover DM, eds. *Transcription and Splicing*. Oxford : IRL Press, 1988 : 131-206.
2. Steitz JA, Black DL, Gerke V, et al. Functions of the abundant U-snRNPs. In : Birnstiel ML, ed. *Structure and Functions of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Berlin : Springer-Verlag, 1988 : 115-54.
3. Guthrie C, Patterson B. Spliceosomal snRNAs. *Annu Rev Genet* 1988 ; 22 : 387-419.
4. Smith CWJ, Patton JG, Nadal-Ginard B. Alternative splicing in the control of gene expression. *Annu Rev Genet* 1989 ; 23 : 527-77.
5. Frendeway D, Keller W. The stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell* 1985 ; 42 : 355-67.
6. Grabowski PJ, Seller SR, Sharp PA. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* 1985 ; 42 : 345-53.
7. Brody E, Abelson J. The « spliceosome » : yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* 1985 ; 228 : 963-7.
8. Mount SM. A catalogue of splice junction sequences. *Nucleic Acids Res* 1982 ; 10 : 459-72.
9. Shapiro MB, Senepathy P. RNA splice junctions of different classes of eukaryotes : sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 1987 ; 15 : 7155-74.
10. Senepathy P, Shapiro MB, Harris NL. Splice junction branch point site and exon : sequence statistics, identification and applications to genome projects. *Methods Enzymol* 1990 ; 183 : 252-78.
11. Jackson IJ. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res* 1991 ; 19 : 3795-8.
12. Hernandez N, Keller W. Splicing of *in vitro* synthesized messenger RNA precursors in HeLa extracts. *Cell* 1983 ; 35 : 89-99.
13. Krainer AR, Maniatis T. Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing *in vitro*. *Cell* 1985 ; 42 : 725-36.
14. Perkins KK, Furneaux HM, Hurwitz I. RNA splicing products formed with isolated fractions from HeLa cells are associated with fast-sedimenting complexes. *Proc Natl Acad Sci USA* 1986 ; 83 : 2387-92.
15. Kramer A. Presplicing complex formation requires two proteins and U2 snRNP. *Genes Dev* 1988 ; 2 : 1155-67.
16. Zamore PD, Green MR. Biochemical characterization of U2 snRNP auxiliary factor : an essential pre-mRNA splicing factor with a novel intranuclear distribution. *Proc Natl Acad Sci USA* 1989 ; 86 : 9243-7.
17. Fu XD, Maniatis T. Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* 1990 ; 343 : 437-41.

Figure 1. **Schéma récapitulatif de l'épissage en cis.** Au cours d'une première réaction de trans-estérification, l'extrémité 5'-phosphate qui est située au début de l'intron devient engagée dans une liaison ester avec le groupement 2'-hydroxyle d'une adénosine (le plus fréquemment) située à l'intérieur de l'intron. Au niveau du point de branchement, l'adénosine réactive est engagée dans trois liaisons phosphoesters (5', 3' et 2'), ce qui donne une structure intronique qualifiée de « lasso ». Dans une deuxième réaction de trans-estérification, l'extrémité 3'-hydroxyle de l'exon effectue une attaque nucléophile au niveau de l'extrémité 3' de l'intron et la ligature des exons est effectuée de manière concomitante à l'élimination de l'intron possédant la structure en lasso. C'est au cours d'une troisième étape, apparemment indépendante de la réaction d'épissage proprement dite, que la structure en lasso est débranchée et l'intron détruit. La formation des spliceosomes nécessite une succession précise d'événements impliquant des interactions ARN-ARN, ARN-protéines et protéines-protéines. La reconnaissance des sites d'épissage 5' et 3' fait intervenir les ARN U1 et U2 dont la séquence nucléotidique est partiellement complémentaire des sites consensus auxquels ils s'associent [1]. Ces petits ARN nucléaires (snRNA) n'existent pas à l'état d'ARN nu, mais sont complexés à des protéines spécifiques (composant 70K spécifique de la snRNP U1, par exemple) dont l'expression, qui repose elle-même sur des réactions d'épissage, est très probablement contrôlée par des « boucles » d'autorégulation [4]. Bien que cette complémentarité soit un élément essentiel dans le processus d'assemblage du spliceosome, elle n'est pas suffisante pour rendre compte de la spécificité du processus, qui est indispensable à la discrimination des sites d'épissage authentiques et assure la non-utilisation de sites dont la séquence primaire est pourtant voisine du consensus. Des protéines accessoires sont requises pour la reconnaissance de l'extrémité 3'-proximale. Les réactions de trans-estérification aboutissent à la formation de la structure en lasso, puis à la ligature des exons (indiquée par la flèche). La structure en lasso est débranchée avant que l'intron soit dégradé. Bien qu'il n'y ait apparemment pas de restriction quant à la taille minimale des exons, on en connaît qui contiennent seulement 3 ou 6 nucléotides [85, 86], il semble cependant que l'épissage des très petits exons ne soit pas très efficace. Un encombrement stérique résultant de la mise en place des facteurs constitutifs des spliceosomes pourrait être responsable d'une interférence négative [4]. De plus, il a été montré récemment que la longueur des petits exons peut influencer sur la sélection des sites d'épissage et que l'inclusion d'exons de courte taille dans des ARN messagers mûrs peut être affectée par des facteurs spécifiques de tissu et par la composition en bases de la région riche en pyrimidines en amont du site 3' [87]. Chez les mammifères, les distances minimales requises entre le point de branchement et les sites d'épissage 5' et 3' sont respectivement de 50 et 18 nucléotides [4], ce qui reflète sans doute les contraintes stériques imposées par la mise en place de la machinerie d'épissage. Il est intéressant de remarquer que les introns de 31 nucléotides qui existent chez les nématodes et les drosophiles ne sont pas correctement excisés par les cellules de mammifères [88]. Il est également bien établi que le contexte dans lequel se situent les signaux d'épissage peut affecter le choix des sites utilisés, et il est fort probable que la liaison des ribonucléoprotéines aux sites d'épissage nécessite des structures secondaires ribonucléiques particulières en plus des séquences consensus qui ont été définies jusqu'à présent. De telles structures constitueraient un critère décisif dans la reconnaissance des sites authentiques d'épissage en favorisant l'interaction des facteurs protéiques dont les empreintes débordent largement les signaux consensus de liaison [4]. La position de l'ARN pré-messenger dans le spliceosome n'est pas établie avec précision. Les ARN U2 et U6 pourraient interagir grâce aux séquences 5'CGCUUCUCG3' et 5'CGUGAAGCG3' qui sont situées dans les régions 5' de U2 et 3' de U6. L'interaction des ARN U6 et U4 fait intervenir des complémentarités de structure secondaire. Les tailles relatives des snRNP et des séquences nucléiques qu'elles protègent ne sont pas respectées sur le schéma.



d'épissage alternatifs, on connaît des exemples d'utilisation optionnelle de promoteurs et de signaux de polyadénylation. Les exons alternatifs peuvent être mutuellement exclusifs, ou constituer un bloc qui est ajouté ou retiré des séquences codantes, généralement sans en modifier la phase de lecture. Enfin, des séquences introniques peuvent être maintenues dans la forme mûre des ARN pré-messagers, avec conservation ou non d'une phase de lecture ouverte. Avec les différents exemples d'épissage en *trans* qui ont été décrits initialement chez les trypanosomes [36, 37] et les nématodes [38], on pouvait penser qu'un degré supérieur de complexité dans la régulation des phénomènes d'épissage alternatifs était atteint puisque la machinerie responsable de ces réactions doit effectuer la ligature adéquate de deux molécules non colinéaires qui ne sont pas nécessairement à proximité immédiate l'une de l'autre.

Chez les trypanosomes, tous les ARNm possèdent une même séquence 5'-proximale non codante de 35 nucléotides (*leader*) qui leur est ajoutée par un mécanisme d'épissage en *trans*, à partir d'un ARN précurseur de 137 nucléotides. Un site donneur consensus est localisé en 3' des 35 nucléotides *leader* dans le précurseur de 137 nucléotides et un site accepteur consensus est toujours retrouvé immédiatement en amont de chacun des ARN pré-messagers auxquels les séquences *leader* sont ajoutées [39-41]. De la même manière, trois des quatre types d'ARNm d'actine et l'un des trois types d'ARN codant pour la glyceraldéhyde 3-phosphate déshydrogénase (GAPDH) de nématodes contiennent une séquence *leader* qui dérive d'ARN précurseurs (ARN SL1 et SL2) d'une centaine de nucléotides, exprimés à partir de *loci* indépendants de ceux qui codent pour les pré-ARNm d'actine et de GAPDH (figure 3, p. 553) [38, 42]. Là aussi, des signaux consensus identiques à ceux qui sont utilisés pour l'épissage en *cis*, sont localisés aux sites réactifs, en 3' et en 5' des ARN *leader* et pré-messagers respectivement.

Chez les nématodes, les mêmes cellules musculaires peuvent aussi bien effectuer des épissages en *cis* et en

## RÉFÉRENCES

18. Dignam JD, Lebovitz RM, Roeder RG. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated nuclei. *Nucleic Acids Res* 1983 ; 11 : 1475-89.
19. Ruskin B, Zamore PD, Green MR. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* 1988 ; 52 : 207-19.
20. Zamore PD, Green MR. Biochemical characterization of U2 snRNP auxiliary factor : an essential pre-mRNA splicing factor with a novel intranuclear distribution. *EMBO J* 1991 ; 10 : 207-14.
21. Gerke V, Steitz JA. A protein associated with small nuclear ribonucleoprotein particles recognizes the 3' splice site of pre-messenger RNA. *Cell* 1986 ; 47 : 973-84.
22. Tazi T, Alibert C, Temsamani J, et al. A protein that specifically recognizes the 3' splice site of mammalian pre-mRNA introns is associated with a small nuclear ribonucleoprotein. *Cell* 1986 ; 47 : 755-66.
23. Ge H, Zuo P, Manley JL. Primary structure of the human splicing factor ASF reveals similarities with *Drosophila* regulators. *Cell* 1991 ; 66 : 373-82.
24. Krainer AR, Mayeda A, Kozak D, Binns G. Functional expression of cloned human splicing factor SF2 : homology to RNA-binding proteins, U1 70K, and *Drosophila* splicing regulators. *Cell* 1991 ; 66 : 383-94.
25. Krainer AR, Conway GC, Kozak D. The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell* 1990b ; 62 : 35-42.
26. Ge H, Manley JL. A protein factor, ASF, controls alternative splicing of SV40 early pre-mRNA *in vitro*. *Cell* 1990 ; 62 : 25-34.
27. Maniatis T. Mechanisms of alternative pre-mRNA splicing. *Science* 1991 ; 251 : 33-4.
28. Chou TB, Zachar Z, Bingham PM. Developmental expression of a regulatory gene is programmed at the level of splicing. *EMBO J* 1987 ; 13 : 4095-104.
29. Ayane M, Preuss U, Kohler G, Nielsen PJ. A differentially expressed murine RNA encoding a protein with similarities to two types of nucleic acid binding motifs. *Nucleic Acids Res* 1991 ; 19 : 1273-8.
30. Nöthiger R, Steinmann-Zwicky M. Sex determination in *Drosophila*. *Trends Genet* 1985 ; 1 : 209-15.
31. Baker BS, Nagoshi RN, Burtis KC. Molecular genetic aspects of sex determination in *Drosophila*. *Bioessays* 1987 ; 6 : 66-70.
32. Boggs RT, Gregor P, Suhair I, Belote JM, McKeown M. Regulation of sexual differentiation in *D. melanogaster* via alternative splicing of RNA from the transformer gene. *Cell* 1987 ; 50 : 739-47.
33. Amrein H, Gorman M, Nöthiger R. The sex-determining gene *tra-2* of *Drosophila* encodes a putative RNA binding protein. *Cell* 1988 ; 55 : 1025-35.
34. Hoshijima K, Inoue K, Higuchi I, Sakamoto H, Shimura Y. Control of *doublesex* alternative splicing by transformer and transformer-2 in *Drosophila*. *Science* 1991 ; 252 : 833-6.

*trans*, de manière exclusive ou concomitante sur une même molécule d'ARN pré-messager et les règles qui dictent le choix des séquences SL (1 ou 2) ajoutées spécifiquement à certains types d'ARN lors de leur maturation ne sont pas encore élucidées. Il est possible que ces séquences aient une influence sur la stabilité ou l'adressage intracellulaire des ARN qui les possèdent. L'épissage *in trans* de séquences codantes a été décrit dans le cas de la synthèse de la protéine ribosomique S12 de chloroplastes [43]. Les exons codant pour cette protéine dans l'ARN messager mûr sont exprimés à partir de *loci* distants l'un de l'autre et sur des brins d'ADN de polarité opposée. On retrouve, aux bornes des introns correspondants, des séquences

nucléotidiques analogues aux signaux consensus spécifiques des introns de type II qui sont éliminés de manière autocatalytique. Dans ce cas, on peut envisager que les ARN pré-messagers participent eux-mêmes à la formation d'une structure secondaire essentielle à la ligature des exons. Des expériences réalisées *in vitro* avaient déjà montré que des molécules d'ARN pouvaient être recombinaisonnées *in trans*, même en absence de complémentarité de structure primaire [44, 45]. La découverte d'épissage *in trans*, *in vivo*, renforçait la possibilité que de tels mécanismes aient pu jouer un rôle extrêmement important au cours de l'évolution. On ne connaît pratiquement rien quant à la nature des facteurs et des signaux qui président à la sélection

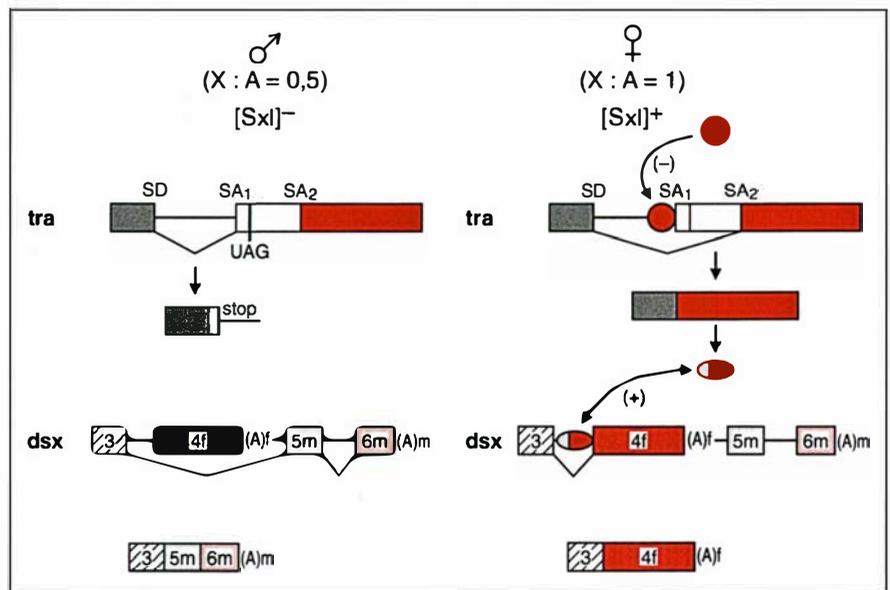


Figure 2. **Régulation de l'épissage et détermination du sexe chez la drosophile.** L'expression du gène sexlethal (*sxl*) est le point clé d'une cascade d'expressions différentielles des gènes transformer (*tra*), transformer 2 (*tra2*), intersex (*ix*) et double sex (*dsx*) aboutissant à l'expression des phénotypes mâle ou femelle [30, 31]. L'expression de ces gènes est réglée par des mécanismes d'épissage alternatifs. La proportion relative du nombre des chromosomes autosomiques et X (rapport X : A) gouverne la production de la protéine *Sxl* chez les femelles. Chez les mâles, l'utilisation du site accepteur d'épissage SA1 conduit à l'interruption de la phase de lecture de l'ARN *tra*. Chez les femelles, la protéine *Sxl* interagit avec SA1 et conduit à l'utilisation du site SA2 qui permet la synthèse de la protéine *Tra* fonctionnelle. La protéine *Sxl* a donc un effet régulateur négatif. Chez les femelles, la protéine *Tra* permet à son tour l'utilisation du site accepteur d'épissage peu efficace qui précède l'exon 4f. L'absence de *Tra* chez les mâles ne permet pas l'utilisation de ce site et conduit à l'épissage des exons 5m et 6m. Ici, la protéine *Tra* exerce un effet régulateur positif, en facilitant la reconnaissance d'un site 3' accepteur d'épissage qui contient des purines dans la région habituellement riche en pyrimidines.

des sites d'épissage en *trans* et, par conséquent, au choix des molécules d'ARN pré-messagers réactives. Étant donné que des signaux consensus d'épissage en *cis* sont retrouvés aux bornes des séquences épissées en *trans*, on peut penser que certaines réactions de base sont communes aux deux types de systèmes. L'épissage en *trans* pourrait requérir l'intervention de structures secondaires particulières

ou de facteurs spécifiques responsables d'interactions de type ARN-protéine ou protéine-protéine, qui favoriseraient l'interaction des séquences nucléotidiques réagissant en *trans*. Jusqu'à présent, aucun motif structural consensus ou facteur protéique spécifique d'épissage en *trans* n'a pu être identifié. Sur la base de quelques caractéristiques qui semblent communes aux différents systèmes

d'épissage en *trans*, il a été proposé que ce phénomène puisse se produire dans de nombreuses espèces, y compris chez les vertébrés.

La description, par notre équipe, de la recombinaison intermoléculaire de séquences codantes du proto-oncogène *c-myb* exprimées à partir de chromosomes différents, chez la poule comme chez l'homme [46], pourrait constituer le premier exemple d'épis-

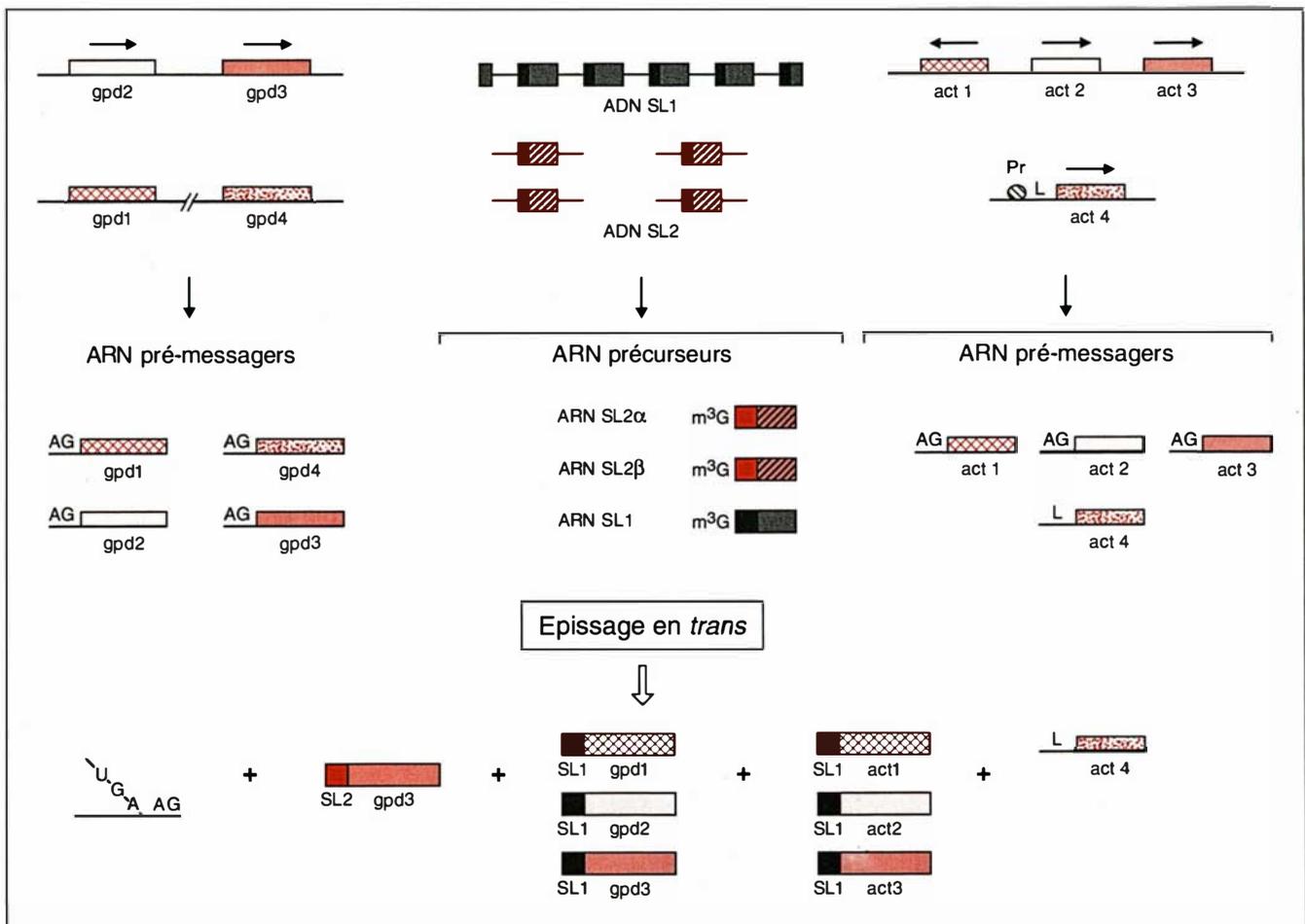


Figure 3. **Épissage en trans chez les nématodes.** L'expression des gènes *gpd1-4* codant pour la glycéraldéhyde 3-phosphate déshydrogénase (GAPDH) et *act1-3* codant pour l'actine fait intervenir l'épissage en trans de séquences leader (SL) d'au moins deux types (SL1 et SL2). Le gène *gpd1* code pour la GAPDH1 exprimée dans toutes les cellules. Le gène *gpd4* code pour la GAPDH2 qui est exprimée uniquement dans les muscles. Ces deux gènes sont espacés d'environ 250 kb sur le chromosome II. Les gènes *gpd2* et *gpd3* sont séparés par 224 paires de bases sur le chromosome X et codent pour la même isoenzyme. Les séquences leader des ARN *gpd* proviennent des gènes SL. Les gènes *act1-3* sont localisés en groupe sur le chromosome X. Les séquences leader (22 nucléotides) des ARN *act1-3* sont transcrites à partir des gènes SL. Le gène *act4* est localisé sur le chromosome X. Les séquences leader d'*act4* sont contiguës au gène qui est exprimé à partir d'un promoteur situé directement en amont. Les gènes codant pour les leaders SL1 sont groupés dans des unités de 1kb répétées 110 fois. Ces unités transcriptionnelles codent aussi, dans l'autre orientation, pour l'ARN 5S. Les séquences codant pour SL et l'ARN 5S sont espacées de 176 nucléotides. Les ARN SL1 et SL2 possèdent une coiffe constituée de triméthylguanosine (m<sup>2</sup>, 2, 7, G), typique de la plupart des ARN U. L'ARN SL1 existe sous la forme d'une ribonucléoprotéine qui pourrait jouer un rôle semblable à la ribonucléoprotéine U1 dans l'épissage en *cis* [89, 90]. Les ARN précurseurs contenant les séquences SL ne sont pas polyadénylés. Les ARN SL2 sont codés par des loci distincts et codent pour au moins deux types d'ARN SL2 ( $\alpha$  et  $\beta$ ).

## RÉFÉRENCES

35. Hedley ML, Maniatis T. Sex-specific splicing and polyadenylation of *dsx* pre-mRNA requires a sequence that binds specifically to Tra-2 protein *in vivo*. *Cell* 1991 ; 65 : 579-86.
36. De Lange T, Berkuens TM, Veerman HJG, Frasch AC, Barry JD, Borst P. Comparison of the genes coding for the common 5' terminal sequence of messenger RNA in three trypanosome species. *Nucleic Acids Res* 1984 ; 12 : 4431-43.
37. Milhausen M, Nelson RG, Sather S, Selkirk M, Agabian N. Identification of a small RNA containing the trypanosome spliced leader : a donor of the shared 5' sequences of trypanosomatid mRNAs ?. *Cell* 1984 ; 38 : 721-9.
38. Krause M, Hirsh D. A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* 1987 ; 49 : 753-61.
39. Guyaux M, Cornelissen AWCA, Pays E, Steinert M, Borst P. *Trypanosoma brucei* : a surface antigen mRNA is discontinuously transcribed from two distinct chromosomes. *EMBO J* 1985 ; 4 : 995-8.
40. Campbell DA, Thornton DA, Boothroyd JC. Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes. *Nature* 1984 ; 311 : 350-5.
41. Kooter JM, De Lange T, Borst P. Discontinuous synthesis of mRNA in trypanosomes. *EMBO J* 1984 ; 3 : 2387-92.
42. Huang XY, Hirsh D. A second *trans*-spliced RNA leader sequence in the nematode *Caenorhabditis elegans*. *Proc Natl Acad Sci USA* 1989 ; 86 : 8640-4.
43. Koller B, Fromm H, Galun E, Edelman M. Evidence for *in vivo trans*-splicing of pre-mRNAs in tobacco chloroplasts. *Cell* 1987 ; 48 : 111-9.
44. Solnick D. *Trans*-splicing of mRNA precursors. *Cell* 1985 ; 42 : 157-64.
45. Konarska MM, Padgett RA, Sharp PA ; *Trans*-splicing of mRNA precursors *in vitro*. *Cell* 1985 ; 42 : 165-71.
46. Vellard M, Soret J, Viegas-Péquignot E, et al. *c-myb* proto-oncogene : evidence for intermolecular recombination of coding sequences. *Oncogene* 1991 ; 6 : 505-14.
47. Dalla-Favera R, Franchini G, Martinotti S, Wong-Staal F, Gallo RC, Croce CM. Chromosomal assignment of the human homologues of feline sarcoma virus and avian myeloblastosis virus *onc* genes. *Proc Natl Acad Sci USA* 1982 ; 79 : 4714-7.
48. Soret J, Vellard M, Viegas-Péquignot E, Apiou F, Dutrillaux B, Perbal B. Complete organization of chicken *c-myb* proto-oncogene : chromosomal reallocation of the *c-myb* locus and splicing of 3' non-coding sequences. *FEBS Lett* 1990 ; 263 : 254-60.
49. Shen-Ong GLC. The *myb* oncogene. *Biochim Biophys Acta* 1990 ; 1032 : 39-52.
50. McMahon J, Howe KM, Watson RJ. The induction of Friend erythroleukaemia differentiation is markedly affected by expression of a transfected *c-myb* cDNA. *Oncogene* 1988 ; 3 : 717-20.
- sage en *trans* faisant intervenir des séquences non consensus chez les eucaryotes supérieurs. Cette découverte démontre que le phénomène d'épissage en *trans* pourrait ne pas être restreint à quelques cas particuliers de l'évolution et pourrait jouer un rôle clé dans la spécificité tissulaire d'expression d'un gène cellulaire impliqué dans la régulation de la multiplication cellulaire. Les implications de ce nouveau concept sont rendues encore plus attrayantes par le fait que le brin antisens de l'exon qui est épissé en *trans* code lui-même pour une protéine susceptible de représenter le premier exemple de facteur d'épissage en *trans* chez les vertébrés.

### Le proto-oncogène *c-myb* code pour une famille de protéines

Le proto-oncogène *c-myb* est l'homologue cellulaire de l'oncogène *v-myb* qui est transduit par les virus de la myéloblastose et de l'érythroblastose aviaires (AMV et E26). Le locus *c-myb* est localisé sur le chromosome 6 (6q22, 23) chez l'homme [47] et sur le chromosome 3 de la poule [48]. Il est constitué de 15 exons (E1 à E15) répartis sur une distance d'environ 35 kilobases (kpb) d'ADN [48] (figure 4).

En dépit de plusieurs indications suggérant que le locus *c-myb* code pour plusieurs produits distincts, il a été considéré pendant de nombreuses années que le seul produit d'expression normale du gène *c-myb* était une protéine nucléaire de 75 kDa traduite à partir d'un ARNm de 4,5 kb. Il existe maintenant toute une série de données expérimentales indiquant que cette vision restrictive est périmée et que l'expression de *c-myb*, soumise à différents processus d'épissage alternatifs, conduit à la biosynthèse de plusieurs produits susceptibles de jouer un rôle dans la régulation de la différenciation hématopoïétique et de la prolifération cellulaire.

Dans des conditions physiologiques normales, le gène *c-myb* est essentiellement exprimé dans les précurseurs immatures des lignages lymphoïdes, myéloïdes et érythroïdes, et une diminution rapide de son expression précède la différenciation terminale des granulocytes, monocytes et autres cel-

lules hématopoïétiques, y compris les lymphocytes B et T [49]. La différenciation des cellules érythroleucémiques de Friend, normalement induite *in vitro* par l'érythropoïétine ou le diméthyl sulfoxyde (DMSO), est inhibée par l'expression constitutive d'un type d'ARN *c-myb* [50-52], mais peut être accélérée lorsqu'un ARN tronqué est exprimé dans ces cellules [52]. L'altération du locus *c-myb* induite par recombinaison homologue provoque une différenciation prématurée des précurseurs hématopoïétiques chez la souris [53]. L'utilisation d'oligonucléotides antisens a permis d'établir que les produits d'expression de *c-myb* sont nécessaires à la multiplication des cellules normales ou tumorales [54, 55] et que l'inhibition de *c-myb* dans les cellules T est accompagnée d'une diminution concomitante de l'expression de l'ADN polymérase  $\alpha$  [56] et de la synthèse de l'ADN. Cette dernière observation est sans doute à rapprocher des résultats que nous avons obtenus récemment et qui suggèrent que le produit du gène *v-myb* puisse stimuler, directement ou indirectement, la réplication [57]. Les produits d'expression de *c-myb*, tout comme la protéine P48<sup>*v-myb*</sup> d'AMV, peuvent se lier de manière spécifique à l'ADN [49] et sont capables de *trans*-activer l'expression de gènes dont les promoteurs contiennent plusieurs copies des motifs nucléotidiques CC/AGTTR ou YGRCVGTTR [49, 58] qui constituent les sites consensus de liaison pour, respectivement, les protéines v-Myb et c-Myb. Le domaine de liaison à l'ADN (figure 4) est formé par trois répétitions imparfaites (R1, R2, R3) de 51-52 acides aminés codés par les exons 2-6 et localisés dans la région amino-proximale des protéines Myb [49]. La répétition R1, qui est partiellement représentée dans l'oncogène *v-myb* d'AMV, n'est pas indispensable à la liaison à l'ADN [59, 60]. Elle pourrait contribuer à la spécificité de l'interaction avec les cibles cellulaires potentielles. Une région codée par les exons 6 et 8, riche en acides aminés acides, est un élément constitutif du domaine *trans*-activateur. Un domaine de régulation négative codé par les exons 9-12 a également été identifié dans la région carboxy-

proximale de *c-myb* (figure 4). Seulement une fraction de ce domaine est conservée dans l'oncogène *v-myb*. Une quatrième région, située dans les exons 9 et 10, contient une fermeture à glissière de leucines (*leucine zippers*) potentielle qui semble également importante pour l'activité *trans*-activatrice de transcription de *v-myb* [61] (figure 4). Le phénotype négatif dominant des mutants altérés dans cette zone rappelle les observations précédemment effectuées avec les facteurs C/EBP et Id [62, 63] et suggère que le domaine carboxy-proximal qui contient également une structure du type « hélice-coude-hélice » permette aux produits de *c-myb* d'interagir avec d'autres facteurs cellulaires.

### Expression du locus *c-myb*

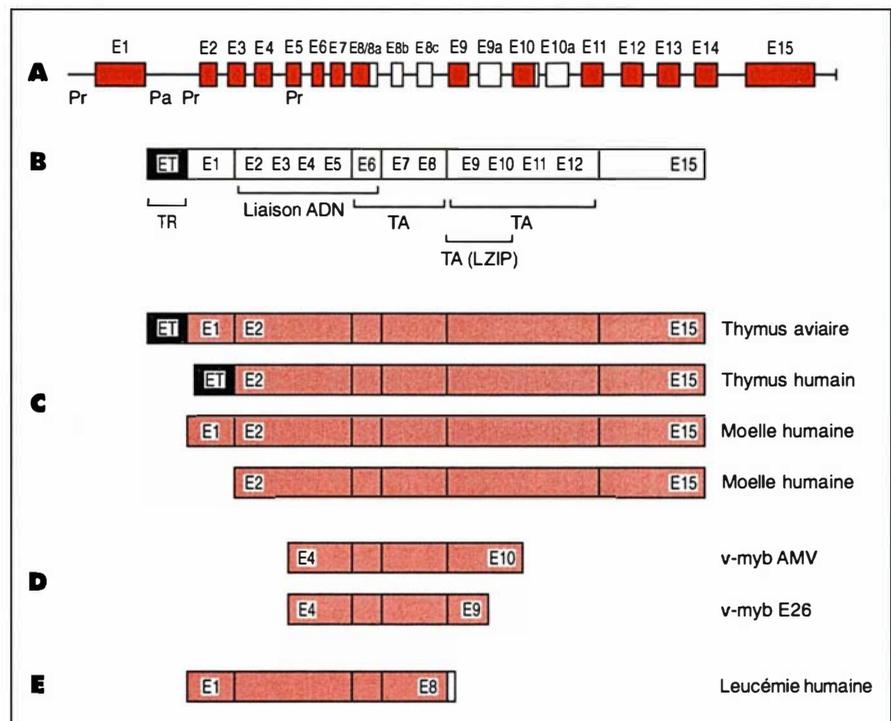
Le séquençage partiel ou total de plusieurs ADNc de *c-myb* isolés à partir de tissus normaux provenant de différentes espèces a révélé que le locus *c-myb* code pour une famille d'ARN messagers qui diffèrent par l'addition ou l'élimination d'exons alternatifs contenant des séquences internes, 3'-distales ou 5'-proximales [49]. Dans certaines cellules leucémiques, l'utilisation d'un exon alternatif interne conduit à l'interruption de la phase codante du produit de *c-myb* et à l'utilisation d'un site de polyadénylation cryptique [49, 64].

L'hétérogénéité des extrémités 5'-proximales des ARN *c-myb* décrite dans les cellules d'origine murine (lignée pré-B 70Z/3B, cellules de leucémie myéloïde WEHI-3B) et aviaire (thymus de poule), nous avait conduit à postuler que l'expression du locus *c-myb* serait contrôlée par plusieurs promoteurs alternatifs [65]. Le marquage de la coiffe des transcrits exprimés dans les cellules myéloïdes du sac vitellin de poule a permis de localiser un site potentiel d'initiation de transcription entre les exons E1 et E2 (B. Perbal, non publié), et un promoteur interne, initialement reconnu sur la base de séquences consensus, a été identifié à l'intérieur de l'exon E5 (qui est le deuxième exon de *c-myb* représenté dans l'oncogène *v-myb* d'AMV).

L'analyse de clones provenant d'une

banque d'ADNc préparée à partir d'ARN messager de moelle humaine normale et le séquençage partiel de l'intron situé entre les exons humains E1 et E2 nous ont permis d'établir que la transcription de certains ARN messagers *c-myb* est initiée à partir d'un promoteur localisé dans l'intron I1 situé en amont de l'exon E2. Nous avons montré que ce promoteur est fonctionnel, qu'il contient un élément régulateur négatif et qu'il

est stimulé *in vivo* par les produits du gène *c-myb*. Ces observations suggèrent l'existence d'un mécanisme d'autorégulation positive de l'expression de *c-myb*. Une observation identique a été publiée récemment en ce qui concerne la régulation du promoteur de transcription localisé en amont de l'exon humain E1 [66]. Les séquences responsables de l'initiation de transcription en amont de E1 ont été partiellement identifiées



## RÉFÉRENCES

51. Clarke MF, Kukowska-Latallo JF, Westin E, Smith M, Prochownik EV. Constitutive expression of a *c-myb* cDNA blocks Friend murine erythroleukemia cell differentiation. *Mol Cell Biol* 1988 ; 8 : 884-92.
52. Weber BL, Westin EH, Clarke MF. Differentiation of mouse erythroleukemia cells enhanced by alternatively spliced *c-myb* mRNA. *Science* 1990 ; 249 : 1291-3.
53. Mucenski ML, McLain K, Kier AB, et al. A functional *c-myb* gene is required for normal murine fetal hepatic hematopoiesis. *Cell* 1991 ; 65 : 677-89.
54. Gewirtz AM, Calabretta B. A *c-myb* antisense oligodeoxynucleotide inhibits normal human hematopoiesis *in vitro*. *Science* 1988 ; 242 : 1303-6.
55. Anfossi G, Gewirtz AM, Calabretta B. An oligomer complementary to *c-myb* - encoded mRNA inhibits proliferation of human myeloid leukemia cell lines. *Proc Natl Acad Sci USA* 1989 ; 86 : 3379-83.
56. Venturelli D, Travali S, Calabretta B. Inhibition of T-cell proliferation by a MYB antisense oligomer is accompanied by selective down-regulation of DNA polymerase  $\alpha$  expression. *Proc Natl Acad Sci USA* 1990 ; 87 : 5963-7.
57. Merzak A, Soret J, Martinerie C, Sureau A, Crochet J, Perbal B. Transforming potential of truncated *v-myb* and stimulation of replication by gag-*myb* fusion products. *Oncogene* 1992 (sous presse).
58. Howe KM, Watson RJ. Nucleotide preferences in sequence-specific recognition of DNA by *c-myb* protein. *Nucleic Acids Res* 1991 ; 19 : 3913-9.
59. Sakura H, Kancsi-Ishii C, Nagase T, Nakagoshi H, Gonda TJ, Ishii S. Delineation of three functional domains of the transcriptional activator encoded by the *c-myb* proto-oncogene. *Proc Natl Acad Sci USA* 1989 ; 86 : 5758-62.
60. Howe KM, Reakes CFL, Watson RJ. Characterization of the sequence-specific interaction of mouse *c-myb* protein with DNA. *EMBO J* 1990 ; 9 : 161-9.
61. Bortner DM, Ostrowski MC. Analysis of the *v-myb* structural components important for transactivation of gene expression. *Nucleic Acids Res* 1991 ; 19 : 1533-9.
62. Benezra R, Davis RL, Lockshon D, Turner DL, Weintraub D. The protein Id : a negative regulator of helix-loop-helix DNA binding proteins. *Cell* 1990 ; 61 : 49-59.
63. Lamb P, McKnight SL. Diversity and specificity in transcriptional regulation : the benefits of heterotypic dimerization. *Trends Biol Sci* 1991 ; 16 : 417-22.
64. Westin EH, Gorse KM, Clarke MF. Alternative splicing of the human *c-myb* gene. *Oncogene* 1990 ; 5 : 1117-24.
65. Soret J, Vellard M, Martinerie C, Crochet J, Perbal B. Organization of 5'-proximal *c-myb* exons in chicken DNA. Implications for *c-myb* tissue-specific transcription. *FEBS Lett* 1988 ; 232 : 227-34.
66. Nicolaidis NC, Gualdi R, Casadevall C, Manzella L, Calabretta B. Positive auto-regulation of a *c-myb* expression *via* myb binding sites in the 5' flanking region of the human *c-myb* gene. *Mol Cell Biol* 1991 ; 11 : 6166-76.

chez la poule et la souris. Cette région contient une fréquence élevée de dinucléotides CpG, un nombre important de sites potentiels de fixation du facteur Sp1 et un motif de type TATA [49]. Une séquence de 124 paires de bases [67], conservée dans les deux espèces et située juste en amont d'un site majeur de transcription chez la poule, pourrait jouer un rôle important dans la régulation transcriptionnelle.

La comparaison des extrémités 5'-proximales des transcrits de *c-myb* a révélé que les points où ces séquences divergent ne correspondent pas à des sites consensus accepteurs d'épissage, suggérant que l'hétérogénéité des extrémités 5'-terminales des ARN *c-myb* résulte de l'utilisation de différents promoteurs et ne provienne pas d'épissage d'exons 5'-proximaux alternatifs [49]. En faveur de cette hypothèse, des expériences d'extension d'amorce et de protection con-

tre la digestion par la nucléase S1 ont permis l'identification de trois sites majeurs d'initiation de transcription en aval des séquences représentées à l'extrémité 5'-proximale d'ARN *c-myb* exprimés dans des cellules tumorales de type B [68]. De plus, plusieurs ADNc *c-myb* d'origines aviaire et humaine contiennent à leur extrémité 5'-proximale des séquences qui ne sont pas contiguës à l'exon E1 dans l'ADN génomique [49].

### L'épissage des exons ET et E1 (*c-myb*) fait intervenir une recombinaison intermoléculaire

L'isolement d'un ANDc contenant à son extrémité 5'-proximale 125 paires de bases codantes représentées uniquement dans les ARNm exprimés dans les thymocytes et leurs précurseurs médullaires [69, 70] confirmait

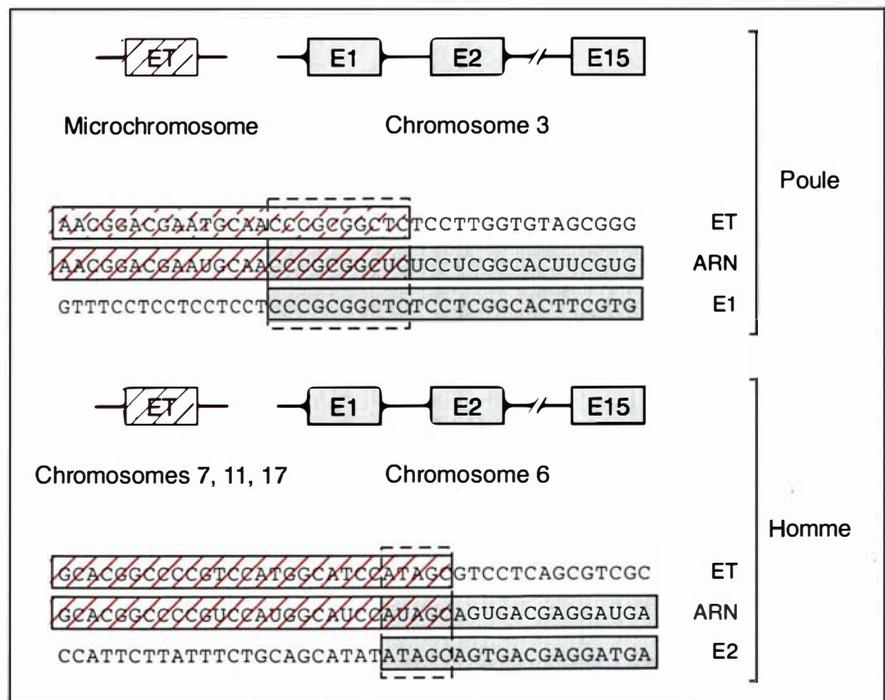


Figure 5. **Épissage en trans des ARN *c-myb*.** Chez la poule, des ARN de thymus normal contiennent une séquence décamerique à la jonction des exons ET et E1. Cette séquence est retrouvée, respectivement, aux extrémités 3' et 5' proximales des exons ET et E1. Chez l'homme, un exemple de recombinaison intermoléculaire fait intervenir les exons ET et E2. Les ARN de thymus contiennent une séquence pentamerique qui est présente aux extrémités 3' et 5' proximales des exons ET et E2.

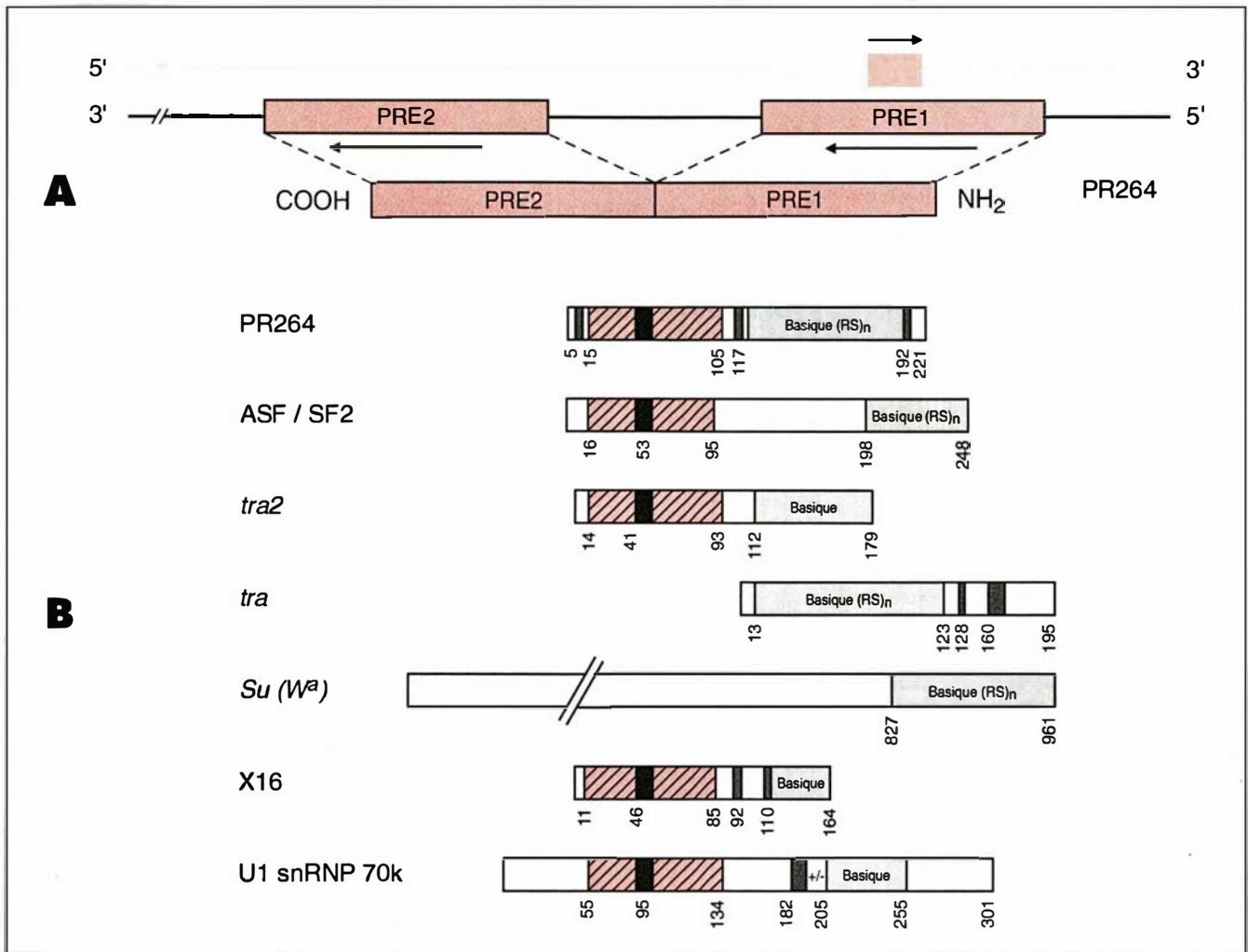


Figure 6. **Structure de la protéine PR264.** La protéine PR264 est exprimée sur le brin antisens du locus ET, à partir de deux exons, PRE1 et PRE2, dont l'épissage est apparemment gouverné par un mécanisme classique d'épissage en cis (**A**). C'est un facteur potentiel d'épissage dont les deux domaines (codés par PRE1 et PRE2) sont représentés dans d'autres facteurs régulateurs d'épissage tels que les facteurs ASF/SF2 et Tra. (**B**) La zone hachurée représente le domaine RNP80, spécifique des protéines interagissant avec les ARN. Le domaine riche en répétitions Arg-Ser est représenté par (RS)<sub>n</sub>. Su (W<sup>a</sup>): gène suppresseur de la mutation white apricot.

l'existence d'exons alternatifs susceptibles d'être présents à l'extrémité 5'-proximale des ARN *c-myb*. Ces observations suggéraient également que l'épissage alternatif de cet exon que nous avons dénommé ET (exon thymique) était gouverné de manière spécifiquement tissulaire, puisqu'il n'était pas représenté dans les ARN provenant de cellules lymphoïdes B [46].

Par hybridation *in situ*, nous avons pu établir que les séquences ET sont exprimées à partir d'une unité transcriptionnelle qui est située sur un microchromosome chez la poule, alors

que le reste des exons de *c-myb* est localisé sur le chromosome 3 [46]. De la même manière, des séquences humaines homologues à l'exon ET de poule sont localisées sur le chromosome 17 (17q25) alors que le locus *c-myb* est situé sur le chromosome 6 (6q22,23) [46].

La dualité de localisation chromosomique des séquences exoniques ET et *c-myb* (E1-E15) dans les deux espèces suggérerait que l'association des séquences codantes de l'ARN *c-myb* fasse intervenir soit une recombinaison chromosomique, soit un épissage *in trans* d'ARN pré-messagers expri-

més à partir de deux loci physiquement séparés (figure 5).

Les expériences d'amplification par PCR que nous avons réalisées pour vérifier l'expression d'ARN hybrides (ET/*c-myb*) dans les thymus de poule, ne nous ont pas encore permis de mettre en évidence de recombinaison des deux loci au niveau génomique. Cette approche ne permet cependant pas d'écarter la possibilité selon laquelle les exons ET et E1 (*c-myb*) soient situés à une grande distance l'un de l'autre sur une même molécule d'ADN recombinante et ligaturés ultérieurement, lors d'un épissage

## RÉFÉRENCES

67. Urbanek P, Dvorak M, Bartunek P, Pecenska V, Paces V, Travnicek M. Nucleotide sequence of chicken *myb* proto-oncogene promoter region : detection of an evolutionarily conserved element. *Nucleic Acids Res* 1988 ; 16 : 11521-30.
68. Hahn SL, Hahn M, Hayward WS. Structural organization of upstream exons and distribution of transcription start sites in the chicken *c-myb* gene. *Mol Cell Biol* 1989 ; 9 : 837-43.
69. Rosson D, Reddy EP. Nucleotide sequence of a chicken *c-myb* complementary DNA and implications for *myb* oncogene activation. *Nature* 1986 ; 319 : 604-6.
70. Kim WK, Baluda MA. Hematopoietic lineage-specific heterogeneity in the 5'-terminal region of the chicken proto-*myb* transcript. *Mol Cell Biol* 1989 ; 9 : 3771-6.
71. Sureau A, Perbal B. Intermolecular recombination of human *c-myb* proto-oncogene coding sequences. *CR Acad Sci Paris* 1991 ; 312 : 323-8.
72. Punyammalee B, Crabeel M, de Lannoy C, Perbal B, Glansdorff N. Two *c-myb* proteins differing by their aminotermini exhibit different transcriptional transactivation activities. *Oncogene* 1991 ; 6 : 11-9.
73. Perbal B, Vellard M. Intermolecular recombination of the *c-myb* proto-oncogene coding sequences in chicken : clues for a mechanism. *CR Acad Sci Paris* 1990 ; 311 : 467-72.
74. Tice-Baldwin K, Fink GR, Arndt KT. BAS1 has a myb motif and activates *HIS4* transcription only in combination with BAS2. *Science* 1989 ; 246 : 931-5.
75. Gilbert W. The RNA world. *Nature* 1986 ; 319 : 618.
76. Vellard M, Soret J, Sureau A, Perbal B. A novel type of RNA-binding protein is potentially encoded by the opposite strands of the trans-spliced *c-myb* coding exon. *CR Acad Sci Paris* 1991 ; 313 : 591-7.
77. Nagai K, Oubridge C, Jessen TH, La J, Evans PR. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* 1990 ; 348 : 515-20.
78. Hoffman DW, Query CC, Golden BL, White SW, Keene JD. RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proc Natl Acad Sci USA* 1990 ; 88 : 2495-9.
79. Vellar M, Sureau A, Soret J, Martinerie C, Perbal B. A potential splicing factor is encoded by the opposite strand of the trans-spliced *c-myb* exon. *Proc Natl Acad Sci USA* 1992 ; 89 : 2511-5.
80. Li H, Bingham PM. Arginine/serine-rich domains of the *su(w<sup>2</sup>)* and *tra* RNA processing regulators target proteins to a subnuclear compartment implicated in splicing. *Cell* 1991 ; 67 : 1-20.

conventionnel de l'ARN pré-messager. Toutefois, le fait qu'aucune des sondes utilisées n'ait permis la détection de fragments recombinants ET/*c-myb* dans l'ADN de thymus n'est pas en faveur d'une telle hypothèse. Alors que les exemples d'épissage en *trans* décrits précédemment mettent en jeu des séquences consensus également importantes pour l'épissage en *cis*, aucun site canonique d'épissage n'est présent en aval de l'extrémité 3' de l'exon ET ou en amont de l'extrémité 5' de l'exon E1 (*c-myb*). En revanche, la séquence décarnérique (CCCCGGGCTC) qui constitue la jonction des exons ET et E1 dans l'ADNc thymique de poule est représentée aux extrémités 3' et 5' des exons ET et E1 respectivement [46], ce qui suggère que l'épissage de ces deux exons fasse appel à de nouveaux facteurs (protéines, ARN, ou ribonucléoprotéines) ou requière des associations d'un type nouveau qui assure la jonction des ARN adéquats. Il est intéressant de remarquer que le nucléotide situé directement en 3' du décarnère correspond à une position identifiée comme un site potentiel d'initiation de transcription par cartographie à la nucléase S1 [68]. Dans le cadre d'un épissage en *trans* des exons ET et E1, il pourrait représenter un site accepteur.

L'analyse par PCR des ARN *c-myb* exprimés dans les cellules de thymus humain n'a pas encore permis d'identifier d'ARNm recombinants dans lesquels les séquences ET seraient fusionnées aux séquences E1. En revanche, cette étude a révélé l'existence d'espèces recombinantes possédant soit des extrémités 5'-proximales non codantes différentes (A. Sureau, résultats non publiés), soit une jonction directe ET-E2 *c-myb*, avec élimination des séquences E1 [71]. Cette observation prend une signification tout à fait intéressante à la lumière de l'identification récente du promoteur alternatif fonctionnel situé dans l'intron séparant E1 et E2 (voir plus haut).

### Épissage de l'exon ET et propriétés biologiques de *c-myb*

Nous avons montré que les séquences polypeptidiques codées par l'exon ET constituent un domaine régulateur

négatif de l'activité *trans*-activatrice de transcription de *c-myb*, mesurée chez la levure [72] ou dans les cellules hématopoïétiques d'origine humaine (résultats non publiés). Par ailleurs, la caractérisation de formes tronquées de *c-myb* exprimées dans diverses tumeurs hématopoïétiques d'origines aviaire et murine (figure 4) suggère que l'activation oncogénique de *c-myb* résulte de modifications conformationnelles qui modulent son activité biologique, soit en perturbant des interactions potentielles entre les différents domaines d'activité, soit en affectant la liaison de facteurs cellulaires susceptibles d'agir en synergie avec *c-myb*. L'existence de plusieurs produits de *c-myb* provenant d'épissages alternatifs dans les cellules hématopoïétiques normales suggère que le locus *c-myb* code pour une famille de protéines qui diffèrent par leurs propriétés biologiques et qui sont spécifiquement exprimées à certains stades de la différenciation cellulaire [73]. Les cellules tumorales, bloquées à des étapes bien déterminées du processus de différenciation, exprimeraient alors le produit de *c-myb* correspondant à chacun de ces stades dans les cellules normales.

On peut ainsi envisager un modèle selon lequel une fonction de base assurée par les exons de *c-myb* contenant le domaine de liaison à l'ADN ait pu être modulée de plus en plus finement au cours de l'évolution grâce à l'adjonction d'exons comportant des domaines régulateurs négatifs et positifs. La très grande conservation phylogénétique du domaine de liaison à l'ADN est en faveur d'un tel modèle. Il est intéressant de souligner, par exemple, que le domaine homologue chez la drosophile peut lier spécifiquement le motif nucléotidique consensus reconnu par les protéines Myb des vertébrés et que des séquences faiblement homologues de ce domaine de liaison ont été retrouvées chez la levure [74] et certaines archaebactéries (B. Perbal, non publié).

Dans le cadre de cette hypothèse, qui serait en accord avec la théorie de brassage des exons (*exon shuffling*) proposée il y a quelques années par W. Gilbert [75], on peut envisager que l'exon ET, qui est actuellement localisé sur un chromosome différent du reste du locus *c-myb* (voir plus haut), s'intègre dans le futur au locus *c-myb*.

Une autre possibilité attrayante serait que le domaine régulateur codé par l'exon ET puisse être épissé en *trans* sur des ARN pré-messagers codant pour d'autres facteurs régulateurs de transcription.

### Locus ET et épissage en trans

Les mécanismes qui gouvernent l'addition de l'exon ET de manière spécifiquement tissulaire sont susceptibles de jouer un rôle important dans le processus de la différenciation cellulaire puisqu'ils aboutissent à la modification de l'activité *trans*-activatrice de protéines impliquées dans la prolifération et la différenciation des cellules hématopoïétiques. L'examen des séquences nucléotidiques flanquant les exons ET et E1

dans l'ADN génomique et dans les ARN de thymus de poule a révélé, en plus du motif décamerique de jonction, l'existence d'une répétition directe de 16 nucléotides localisés respectivement à 63 et 201 nucléotides en amont et en aval du codon initiateur situé dans ET qui est utilisé pour l'initiation de la traduction de la protéine ET-*myb* [46]. Une séquence identique étant située à 62 et 201 nucléotides de l'ATG initiateur dans l'exon ET humain, il est probable que ce motif conservé joue un rôle dans le processus de la recombinaison intermoléculaire. L'identification des ARN précurseurs intervenant dans l'épissage en *trans* nécessite la caractérisation des transcrits exprimés à partir du locus ET. Alors qu'un locus ET unique a été caractérisé chez la poule, quatre loci

ET ont été identifiés sur des chromosomes humains distincts ([46] et C. Guyon et J. Soret, non publié). L'utilisation de sondes spécifiques de petites fractions de ET (A. Sureau, non publié) et le clonage des différents loci (C. Guyon et J. Soret, non publié) nous ont permis d'établir que les gènes ET sont structurellement très proches. Les résultats d'hybridations réalisées à l'aide de sondes ARN et d'oligonucléotides sens et antisens ont établi que les loci ET aviaire et humain étaient transcrits dans les deux orientations. La caractérisation des différents ARN a révélé que les espèces majeures d'ARN sens exprimées à partir du locus ET aviaire ne sont pas polyadénylées (M. Vellard, résultats non publiés) et représentent donc de bons candidats pour des précurseurs

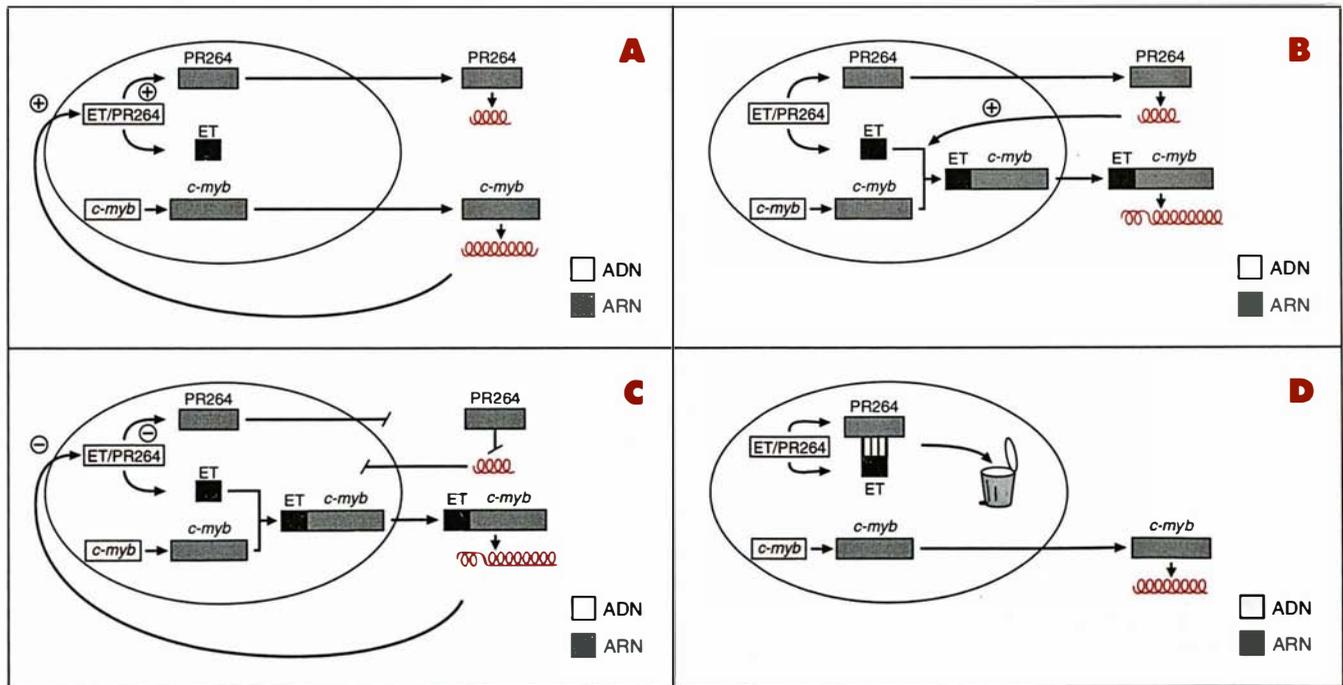


Figure 7. **Modèle d'autorégulation de l'épissage en trans de c-myb.** La transcription des loci *c-myb* et *PR264/ET* conduit à la production d'ARN ET non polyadénylé séquestré dans le noyau, et à l'expression des protéines *c-Myb* et *PR264* à partir d'ARN messagers transportés dans le cytoplasme. Le promoteur *PR264* est stimulé par le produit de *c-myb* (A). La protéine *PR264* transportée dans le noyau permet l'épissage en trans des séquences ET et *c-myb* (B), et par conséquent, la synthèse d'un produit ET/*c-myb* qui possède une activité *trans*-activatrice de transcription réduite vis-à-vis du promoteur *PR264* (C). Les quantités relatives d'ARN sens (ET) et antisens *PR264* sont alors en défaveur de l'ARN *PR264* qui est « piégé » dans les structures bicaténaires sélectivement modifiées et détruites (D). Le niveau de protéine *PR264* n'est plus suffisant pour assurer un épissage en trans efficace ; une protéine *c-myb* dépourvue de ET est exprimée. La régulation de cette boucle reposerait essentiellement sur le contrôle de l'adressage des ARN et protéines en jeu.

## RÉFÉRENCES

81. Fu XD, Maniatis T. The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. *Proc Natl Acad Sci USA* 1992 ; 89 : 1725-9.
82. Mancebo R, Lo PCH, Mount SM. Structure and expression of the *Drosophila melanogaster* gene for the U1 small nuclear ribonucleoprotein particle 70K protein. *Mol Cell Biol* 1990 ; 10 : 2492-502.
83. Miyajima N, Horiuchi R, Shibuya Y, et al. Two *erbA* homologs encoding proteins with different T<sub>3</sub> binding capacities are transcribed from opposite DNA strands of the same genetic locus. *Cell* 1989 ; 57 : 31-39.
84. Wagner RW, Smith JE, Cooperman BS, Nishikura K. A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells and *Xenopus* eggs. *Proc Natl Acad Sci USA* 1989 ; 86 : 2647-51.
85. Cooper TA, Ordahl CP. A single cardiac troponin T gene generates embryonic and adult isoforms via developmentally regulated alternative splicing. *J Biol Chem* 1985 ; 260 : 11140-8.
86. Santoni MJ, Barthels D, Vopper G, Boned A, Goridis C, Wille W. Differential exon usage involving an unusual splicing mechanism generates at least eight types of NCAM cDNA in mouse brain. *EMBO J* 1989 ; 8 : 385-92.
87. Dominski Z, Kole R. Selection of splice sites in pre-mRNAs with short internal exons. *Mol Cell Biol* 1991 ; 11 : 6075-83.
88. Kay RJ, Russnak RH, Jones D, Mathias C, Candido EPM. Expression of intron containing *C. Elegans* heat-shock genes in mouse cells demonstrates divergence of 3' splice site recognition sequences between nematodes and vertebrates, and an inhibitory effect of heat shock on the mammalian splicing apparatus. *Nucleic Acids Res* 1987 ; 15 : 3723-41.
89. Von Doren K, Hirsh D. Trans-spliced leader RNA exists as small nuclear ribonucleoprotein particles in *Caenorhabditis elegans*. *Nature* 1988 ; 335 : 556-8.
90. Bruzik JP, Van Doren K, Hirsh, Steitz JA. Trans-splicing involves a novel form of small nuclear ribonucleoprotein particles. *Nature* 1988 ; 335 : 559-62.
91. Shen-Ong GLC, Skurla RM, Owens JD, Mushinski JF. Alternative splicing of RNAs transcribed from the human *c-myc* gene. *Mol Cell Biol* 1990 ; 10 : 2715-22.

d'épissage en *trans*. Les résultats préliminaires que nous avons obtenus suggèrent que les nombreux transcrits exprimés par l'unique *locus* ET de poule puissent être exprimés séparément par les différents *loci* humains. Une telle situation pourrait correspondre à une spécialisation fonctionnelle de chaque *locus* ET chez l'homme.

La mise en évidence d'une transcription bidirectionnelle nous avait conduit à proposer que les ARN messagers antisens ou des facteurs protéiques codés par ces transcrits soient directement impliqués dans le processus d'épissage en *trans* de ET [73]. La protéine PR264 (figure 6) qui est codée par les ARN antisens majeurs exprimés chez la poule et chez l'homme est extrêmement bien conservée [76]. Dans les deux espèces, elle contient 221 acides aminés (masse moléculaire théorique : 25 kDa) qui se répartissent en deux domaines structuraux reliés par une boucle polypeptidique riche en proline et glycine. Le domaine amino-proximal est constitué essentiellement par un motif RNP80\* dont 63 résidus sont identiques au motif consensus caractéristique des protéines se liant aux ARN et dont 13 autres résidus sont également retrouvés dans les motifs RNP80 d'autres protéines ayant de l'affinité pour les ARN [76]. La protéine PR264 se caractérise par la présence d'une lysine à la place d'un résidu aromatique conservé dans plusieurs protéines de ce type. Cette modification locale pourrait avoir des conséquences importantes quant aux propriétés biologiques de la protéine PR264 puisque le sous-domaine RNP2, qui contient le résidu aromatique conservé, serait contenu dans une structure en hélice  $\alpha$  au lieu d'être inclus dans un feuillet  $\beta$  qui semble important pour la spécificité de la liaison de la protéine U1-A\*\* avec ses cibles [77, 78].

Le second domaine de la protéine PR264 (domaine RS) est constitué

presque exclusivement par des répétitions du dipeptide Arg-Ser qui sont contenues dans une structure de type *random coil*, mais dont l'arrangement autour d'axes de symétrie potentiels suggère une organisation très particulière qui pourrait correspondre à un nouveau type de motif structural spécifique des protéines de ce type (figure 6, p. 557). Les caractéristiques structurales de la protéine PR264 correspondent à l'archétype d'une famille de facteurs impliqués dans la régulation de l'épissage (figure 6). Il est particulièrement intéressant à cet égard de remarquer que le facteur ASF/SF2 qui est impliqué dans la sélection des sites donneurs d'épissage alternatif est celui qui possède le plus d'analogie avec la protéine PR264. Dans les deux cas, les domaines RNP80 et RS sont codés par deux exons distincts dont l'épissage semble obéir aux règles consensuelles générales. La présence d'un motif RDAEDA\*\*\* dans le domaine RNP80 de ces deux polypeptides, alors qu'il n'est retrouvé dans aucune autre protéine des banques de données disponibles, pourrait résulter d'une fonction commune à ces deux facteurs.

La protéine PR264, exprimée *in vitro*, a un poids moléculaire apparent de 35 kDa en gel de polyacrylamide [79], au lieu des 25 kDa prédits à partir de sa séquence primaire. Cette différence résulte probablement de la quantité élevée de résidus arginine qu'elle contient dans son domaine RS. Des études récentes ont indiqué que le motif RS contenu dans les protéines Tra et Su (W<sup>a</sup>)\*\*\*\* de drosophile sont responsables de l'adressage de ces facteurs vers un sous-compartiment nucléaire impliqué dans l'épissage [80]. Parmi les facteurs d'épissage dont le poids moléculaire apparent est de 35 kDa [17, 20], seul SC35, qui module probablement les interactions au site accepteur [81], a été localisé de manière certaine dans ce compartiment [17]. Il serait donc intéressant de déter-

\* RNP80 : motif conservé de 80 acides aminés, impliqué dans la reconnaissance des ARN.

\*\* Protéine U1-A : protéine constitutive de la ribonucléoprotéine qui contient l'ARN U1.

\*\*\* RDAEDA : Arg-Asp-Ala-Glu-Asp-A.

\*\*\*\* Tra : abréviation de transformer. Gène responsable de la différenciation sexuelle chez la drosophile. Su (W<sup>a</sup>) : suppressor of white apricot. Gène suppresseur de la mutation white apricot.

## TIRÉS A PART

B. Perbal.

miner si les protéines SC35 et PR264 ont des propriétés biologiques communes.

Nous avons obtenu des résultats suggérant que l'expression des différents ARN PR264 résulte de processus d'épissage alternatif des séquences 3' non codantes et est contrôlée de manière spécifique au cours de l'embryogenèse [79]. Des processus d'épissage alternatif du même type ont été décrits dans le cas de la protéine 70K constituant la ribonucléoprotéine U1 de drosophile [82].

### **PR264 : un facteur d'épissage en trans impliqué dans la régulation de la différenciation hématopoïétique ?**

L'expression du domaine régulateur ET et d'un facteur potentiel d'épissage à partir de phases de lecture ouvertes chevauchantes ayant une polarité opposée suggère très fortement que la transcription bidirectionnelle du locus ET/PR264 permette une régulation fine des niveaux relatifs d'expression sens et antisens. Jusqu'à présent, on ne connaissait qu'un seul exemple de transcription bidirectionnelle conduisant à l'expression de produits biologiquement apparentés [83]. Il est tentant de penser que l'expression interdépendante des ARN ET et PR264 ait un fondement fonctionnel et que le facteur PR264, qui possède des caractéristiques de protéines réglant l'épissage alternatif, participe à la recombinaison intermoléculaire des ARN ET et *c-myb* et représente le premier exemple de facteur d'épissage en *trans*.

L'expression spécifique de tissu d'une classe d'ARN PR264 et la régulation de cette expression au cours de l'embryogenèse [79] pourraient alors constituer des niveaux de contrôle importants de l'épissage en *trans* et, par là-même de l'activité régulatrice de transcription (et de réplication ?) des produits de *c-myb* exprimés dans les différents tissus hématopoïétiques au cours de la différenciation.

Nous avons pu établir récemment que le promoteur du locus PR264 contient des sites de fixation spécifiques de *myb* et que son activité *in vitro* était réglée de manière différentielle par la liaison de différents produits de *c-myb* à ces

sites (manuscrit en préparation). Nous avons ainsi pu établir que la transcription initiée au promoteur PR264 est stimulée par une protéine *c-myb* dépourvue des séquences ET et proposer un modèle dans lequel l'épissage en *trans* de l'exon ET serait soumis à une boucle d'autorégulation (figure 7, p. 559). Dans ce modèle, le niveau d'épissage en *trans* de l'exon ET, qui dépendrait de la présence de la protéine PR264, serait également contrôlé par la quantité relative des ARN sens et antisens. La présence simultanée de ces ARN au sein du noyau pourrait conduire à la formation d'ARN double brin dont les adénosines seraient sélectivement modifiées en inosine [84]. De cette manière, tout excès de l'un ou l'autre de ces ARN pourrait conduire à l'élimination des ARN complémentaires. Dans le cadre de ce modèle, les processus d'adressage des ARN sens et antisens et de la protéine PR264 joueraient un rôle clé dans le déroulement de l'épissage en *trans* et l'expression de protéines *c-myb* dont les propriétés *trans*-activatrices de transcription pourraient être requises à des étapes bien déterminées de la différenciation hématopoïétique.

### **Conclusion**

Le mécanisme d'épissage en *trans* de l'exon ET, dont dépendraient les propriétés biochimiques du produit de *c-myb* exprimé dans les différentes cellules de la lignée T, pourrait donc jouer un rôle important dans la différenciation hématopoïétique, au même titre que l'épissage alternatif des transcrits *tra* et *sxl* est un élément clé de la détermination du sexe chez la drosophile.

Toute altération d'un des composants de ce système pourrait conduire à l'expression constitutive d'une protéine Myb dont les propriétés seraient incompatibles avec le déroulement normal de la différenciation des cellules qui l'expriment et induire ainsi le développement de certaines tumeurs hématopoïétiques.

Les perspectives fascinantes qu'ouvre ce nouveau concept confèrent un intérêt primordial à l'étude des mécanismes moléculaires de l'épissage en *trans* de l'exon ET et à la compréhension de sa signification biologique ■

## **Summary**

### **Trans-splicing of *c-myb* proto-oncogene and hematopoietic differentiation**

The *myb* proto-oncogene encodes a variety of products likely to play a critical role in the control of hematopoietic cell differentiation and proliferation. The biological properties of the *c-myb* products which probably rely primarily upon their ability to bind specific DNA sequences are modulated by several functional domains. We have established that the transcriptional transactivating properties of *c-myb* are down regulated by an aminoproximal domain whose corresponding exon (ET) is localized on a separate chromosome, both in chicken and human. We therefore suggested that the expression of the ET-containing mRNA species in thymic cells results from trans-splicing of precursor RNA species expressed from physically unrelated transcriptional units. We speculate that the splicing factor (PR264) which is encoded on the opposite strand of the ET locus, participates directly in the intermolecular recombination of *c-myb* coding sequences and represents the first example of trans-splicing factor. As such, the PR264 polypeptide would control the modulation of the *c-myb* transactivating properties and could therefore be involved in the regulation of hematopoietic differentiation.

### **Remerciements**

Je remercie tout particulièrement F. Galibert et E. DeMaeyer pour leurs critiques et suggestions ainsi que E. Moustacchi, J. Soret et L. Gabaron pour la relecture du manuscrit. Un grand merci à J. Soret pour avoir dessiné la figure 7.

Les travaux effectués dans notre laboratoire ont été financés par le Cnrs, l'Institut Curie, l'Association pour la recherche contre le cancer, le ministère de l'Éducation nationale, la Fondation contre la leucémie, la Ligue nationale contre le cancer et la Fondation pour la recherche médicale.