

Mais que sont les gènes SOX ?

Le clonage, en 1990, du gène SRY humain a mis fin à la longue recherche du facteur TDF (acronyme anglais pour *testis determining factor*) porté par le chromosome Y des mammifères et déterminant le sexe mâle [1, 2]. Ce gène monoexonique code pour une petite protéine contenant un motif de 79 acides aminés, encore appelé boîte HMG (pour *high mobility group proteins*). En effet, dès sa séquence protéique établie, SRY apparut l'homologue de diverses protéines dont une, par exemple, participe à la reproduction sexuée de la levure *Schizosaccharomyces pombe* (ou protéine Mc), alors qu'une autre joue le rôle de facteur de transcription nucléolaire pour l'ARN polymérase I, ou facteur hUBF. Comme l'indique le nom de ce domaine peptidique, l'homologie est également partagée avec la classe des protéines non histones qui participent à l'organisation de la chromatine, ou protéines HMG. Cette faible homologie, restreinte aux 79 acides aminés de la boîte HMG, permit dès son clonage de postuler que la protéine SRY se liait à l'ADN et, au-delà, constituait un nouveau facteur de transcription. Après obtention de la protéine par génie génétique, des expériences de retard de migration de cibles oligonucléotidiques sur gels de polyacrylamide permirent de confirmer la capacité de la protéine SRY de se lier à l'ADN double-brin, aussi bien au niveau de séquences spécifiques riches en A-T [3], que de séquences aléatoires dans le cas d'un ADN de type cruciforme [4]. Il est à noter que cette capacité de liaison est bien

dévolue aux seuls 79 acides aminés [5] ; des altérations mineures de la séquence de ce motif, telles les mutations ponctuelles apparues spontanément chez des femmes révertantes sexuelles à caryotype 46, XY, suffisent à abolir la capacité de liaison à l'ADN de SRY [3, 6].

La découverte de SRY fournit aujourd'hui le point de départ à de nombreuses études visant à établir le mode d'action de la protéine et la nature des cibles génétiques d'un tel facteur de transcription. On espère ainsi reconstituer une voie complète de développement concourant à la mise en place d'une catégorie particulière de cellules, la cellule de Sertoli, et au processus de détermination du sexe mâle.

Parallèlement à ces travaux, l'équipe de R. Lovell-Badge, en cherchant à cloner un ADN complémentaire produit du gène *sry* murin, isola quatre séquences très homologues au niveau du motif HMG [7]. Cela permit à cette équipe d'introduire la notion de gènes SOX (pour *SRY box containing genes*), un sous-groupe à l'intérieur de la super-famille des protéines à boîte HMG [8]. Depuis cette constatation initiale, le nombre des gènes SOX n'a cessé de croître, cela dans un grand nombre d'espèces, aussi bien chez les vertébrés que chez les invertébrés [9-13]. Dans la grande majorité des cas, les séquences publiées à ce jour ne sont que partielles et reposent sur la possibilité de mettre en place la technique dite de PCR dégénérée pour amplifier sélectivement les boîtes HMG des gènes SOX. La très grande conservation du motif entre les différentes protéines SOX (*figure 1*), mais

également entre les protéines SOX de diverses espèces, suggère un rôle important pour ces gènes au cours du développement. De plus, à l'inverse de ce qui est observé pour SRY, les rares séquences complètes récemment établies pour les protéines SOX humaines et murines semblent démontrer une forte homologie entre elles, même en dehors du motif HMG [14]. Le fort degré de conservation entre ces protéines n'a pas permis de les différencier quant à leur spécificité de liaison à l'ADN [15], ce qui laisse entier le problème de leur spécificité de fonction, peut-être liée à la fenêtre d'expression de leur gène. L'implication de ces protéines dans les processus de développement devrait se trouver confirmée rapidement par la mise en évidence, chez la souris, de l'expression embryonnaire de certains gènes *Sox* au niveau de diverses structures nerveuses (J. Gollignon et R. Lovell-Badge, manuscrit soumis). Récemment, les profils d'expression de trois gènes SOX ont été plus particulièrement étudiés :

1. un ADN complémentaire humain, cartographié en 3q21-qter, correspondant à un SOX exprimé au niveau de la rétine (Sadler *et al.*, résultats non publiés) ;
2. le gène murin *sox-5* dont l'expression liée à la spermatogenèse est restreinte aux stades post-méiotiques [15] ;
3. enfin le gène SOX-4 monoexonique qui participerait au processus de différenciation lymphocytaire [16].

Il est possible que ces diverses protéines SOX révèlent des propriétés similaires à celles entrevues pour la

SRY hum :	RVKRPMNAFI	VWSRDQRRKM	ALENPRMRNS	EISKQLGYQW	KMLTEAEKWP	FFQEAQKLQA	MHREKYPNYK	YRPRRKAK
SOX4 hum :	-----	---PIE...I	MEQS.D.H.A	...R..KR.	.L.KDSD.I.	.IR..ER.RL	K.MAD.----	-----
SOX11 hum :	-----M	...KIE...I	MEQS.D.H.A	...R..KR.	...KDS..I.	.IR..ER.RL	K.MAD..D--	-----
SOX12 hum :	-----M	...SA...Q.	.H...K.H.	...R..A.	.L.D.D..R.	.VE..KR.R.	.LRD..D--	-----
Sox1 sour :M	...G....	.Q...K.H.	...R..AE.	.VMS...R.	.ID..KR.R.	I.MKEH.D.T.
Sox2 sour :M	...G....	.Q...K.H.	...R..AE.	.L.S.T..R.	.ID..KR.R.	L.MKEH.D.T.
Sox3 Lf :	-----M	...G....	.Q...K.H.	...R..AD.	.L.SD...R.	.ID..KR.R.	V.MKE-----	-----
Sox1 pleu :	-----M	...G....	.Q...K.H.	R..AD.	.L.SD...R.	.ID..KR.R.	V.MKD-----	-----
Sox12 xen :	-----	--AK.E...I	LQAF.D.H.	S...I..SR.	.SMSNG..Q.	YYE.QAR.SR	Q.L.R.----	-----
Sox14 dro :	-----	--SQME...I	CERT.DLH.A	...E..RR.	QL.SKDD.Q.	YII..E..RK	L.MIE.----	-----

Figure 1. **Comparaison des séquences protéiques des domaines HMG de quelques gènes SOX.** La forte conservation des 79 acides aminés de ces boîtes apparaît clairement. Les points rouges représentent des identités et les tirets sont des indéterminations. hum : humain; sour : souris; Lf : *Larus fuscus* (oiseau); pleu : pleurodèle; xen : xénope; dro : drosophile. Code à une lettre des acides aminés : A : Ala; C : Cys; D : Asp; E : Glu; F : Phe; G : Gly; H : His; I : Ile; K : Lys; L : Leu; M : Met; N : Asn; P : Pro; Q : Gln; R : Arg; S : Ser; T : Thr; V : Val; W : Trp; Y : Tyr.

protéine SRY. Si tel est le cas, leur liaison à l'ADN, réalisée au moins partiellement au niveau du petit sillon modifierait la structure locale de l'ADN en imprimant à ce dernier une forte courbure [17, 18]. On peut alors proposer que ces protéines pourraient modifier la structure locale de la chromatine et rendre ainsi des gènes cibles accessibles à des complexes transcriptionnels spécifiques de certaines voies de développement. Un effet opposé amenant à la réalisation de blocs transcriptionnels ne peut être également écarté.

D'ores et déjà, il est possible de se demander si ces gènes SOX nous réserveront autant de surprises que leurs compères les gènes HOX? Nul doute que de nombreux résultats viendront vite répondre à cette question ■

Philippe Berta
Catherine Gozé
Francis Poulat

Centre de recherche de biochimie macromoléculaire, CNRS UPR 9008-INSERM U. 249, route de Mende, BP 5051, 34033 Montpellier Cedex, France.

TIRÉS A PART

P. Berta.

RÉFÉRENCES

- Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 1990; 346: 240-4.
- Weissenbach J, Petit C. Chromosome Y et détermination du sexe. *médecine/sciences* 1990; 6: 785-90.
- Harley VR, Jackson DI, Hextall PJ, Hawkins JR, Bergovitz GD, Socknathan S, Lovell-Badge R, Goodfellow PN. DNA binding activity of recombinant SRY from normal males and XY females. *Science* 1992; 255: 453-6.
- Ferrari S, Harley VR, Pontiggia A, Goodfellow PN, Lovell-Badge R, Bianchi ME. SRY like HMG1, recognizes sharp angles in DNA. *EMBO J* 1992; 11: 4497-506.
- Poulat F, Guichard G, Goze C, Heitz F, Calas B, Berta P. Synthesis of a large peptide mimicking the DNA binding properties of the sex determining protein, SRY. *FEBS Let* 1992; 309: 385-8.
- Poulat F, Soulier S, Goze C, Heitz F, Calas B. Description and functional implications of a novel mutation in the sex-determining gene SRY. *Hum Mutation* 1993 (sous presse).
- Gubbay J, Collignon J, Koopman P, Capel B, Economou A, Musterberg A, Vivian N, Goodfellow P, Lovell-Badge R. 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* 1990; 346: 245-50.
- Laudet V, Stéhelin D, Clevers H. Ancestry and diversity of the HMG box superfamily. *Nucleic Acids Res* 1993; 21: 2493-501.
- Griffiths R. The isolation of conserved DNA sequences related to the human sex-determining region Y gene from the lesser black-backed gull (*Larus fuscus*). *Proc R Soc Lond (Biol)* 1991; 244: 123-8.
- Denny P, Swift S, Brand N, Dahbade N, Borton P, Ashworth A. A conserved family of genes related to the testis determining gene, SRY. *Nucleic Acids Res* 1992; 20: 2887.
- Coriat AM, Muller U, Harry JL, Uwanogho D, Sharpe PT. PCR amplification of SRY-related gene sequences reveals evolutionary conservation of the SRY-box motif. *PCR Methods Appl* 1993; 2: 218-22.
- Wright EM, Snopek B, Koopman P. Seven new members of the SOX gene family expressed during mouse development. *Nucleic Acids Res* 1993; 21: 744.
- Goze C, Poulat F, Berta P. Partial cloning of SOX-11 and SOX-12, two new human SOX genes. *Nucleic Acids Res* 1993; 21: 2943.
- Whitfield LS, Lovell-Badge R, Goodfellow PN. Rapid sequence evolution of the mammalian sex-determining gene SRY. *Nature* 1993; 364: 713-5.
- Denny P, Swift S, Connor F, Ashworth A. An SRY-related gene expressed during spermatogenesis in the mouse encodes a sequence-specific DNA-binding protein. *EMBO J* 1992; 11: 3705-12.
- Schilham MW, Van Eijk M, Van de Wetering MV, Clevers HC. The murine Sox 4 protein is encoded on a single exon. *Nucleic Acids Res* 1993; 21: 2009.
- Van de Wetering M, Clevers H. Sequence-specific interaction of the HMG box proteins TCF-1 and SRY occurs within the minor groove of a Watson-Crick double helix. *EMBO J* 1992; 11: 3039-44.
- Giese K, Cox J, Grosschedl R. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell* 1992; 69: 185-95.