

Du nouveau dans les séquences activatrices des gènes de globine (LCR)

La synthèse de la globine humaine est contrôlée par deux familles de gènes situés, pour les gènes α -globine sur le bras court du chromosome 16, pour le gène β -globine sur le bras court du chromosome 11. Dans les deux cas, leur position de 5' en 3' est parallèle à l'ordre de leur expression au cours du développement ontogénique (voir *m/s* n° 7, vol. 4, p. 427). De nombreux travaux, depuis une dizaine d'années, ont exploré les mécanismes transcriptionnels contrôlant l'expression de ces gènes strictement spécifiques de tissu, dont l'expression est coordonnée au cours du développement, donnant ainsi lieu à la production d'hétérodimères successifs. L'intérêt de ces travaux est double. Ils sont susceptibles de devenir un modèle pour d'autres familles multigéniques. Ils tendent aussi à préciser quelles seraient les séquences minimales nécessaires dans la perspective d'une thérapie génique [1]. Les résultats, bien qu'encore incomplets, sont cependant tout à fait remarquables. Ils ont été rendus possibles par des avancées méthodologiques : études de conformation chromatinienne, études de structure de l'ADN et détermination des séquences liant des facteurs protéiques transactivateurs, expression en lignées cellulaires, puis en souris transgéniques.

Nous ne reviendrons pas sur les grandes avancées qu'ont représentées, au milieu des années 1980, le transfert de gènes de globine à la souris, la mise en évidence de sites hypersensibles à la DNase I encadrant le *locus* β -globine et le rôle activateur des séquences entourant ces sites sur l'expression d'un gène transféré (*m/s* n° 4, vol. 4, p. 252). Une étape ultérieure était la démonstration que cette même région, située largement en amont du gène embryonnaire, et appelée depuis LCR (*locus control region*), pouvait également activer des gènes hétérologues, ainsi que le

gène α -globine, entraînant chez la souris la production de globine humaine normale ou pathologique (*m/s* n° 3, vol. 6, p. 314). Des constructions variées incluant ce LCR et différents fragments d'ADN, puis leur transfert chez la souris ou leur expression en lignées permanentes, ont également, depuis deux ou trois ans, permis de formuler des hypothèses sur les mécanismes impliqués dans la commutation physiologique de l'hémoglobine fœtale à l'hémoglobine adulte (*m/s* n° 6, vol. 6, p. 572) ; la mise en évidence d'une zone, également hypersensible à la DNase I et présentant des propriétés activatrices, en amont du *locus* α -globine, a soulevé la question de la régulation coordonnée et équilibrée des deux familles de gènes (*m/s* n° 9, vol. 6, p. 918). D'autres travaux mettaient en évidence l'extrême conservation, à la fois de structure et d'organisation, de ce LCR dans l'évolution des espèces, et donc l'importance de ce type d'élément pour le *locus* α aussi bien que pour le *locus* β . Des mutants naturels, par ailleurs, confirmaient l'importance du LCR : sa délétion entraîne la non-expression de gènes présents et que l'on peut cependant exprimer *in vitro*. Enfin, il faut insister sur l'adoption en 1990 d'une nomenclature unifiée corrigeant certaines nomenclatures antérieures : les sites en amont du gène ϵ sont appelés, de 3' en 5', 5'HS1, 5'HS2, 5'HS3 et 5'HS4 ; le site loin en aval du gène β est le site 3'HS. De très nombreux travaux ont continué à être publiés depuis cette date. On peut schématiquement les regrouper sous quatre têtes de chapitre : (1) la « dissection » des régions du LCR, pour identifier les séquences d'interaction avec des facteurs actifs en *trans* ; (2) l'étude du mécanisme autonome et/ou compétitif de l'expression des différents gènes ; (3) l'extension des données concernant le *locus* α , par

comparaison au *locus* β ; (4) enfin, les modalités de la manipulation pharmacologique.

• Dissection des régions du LCR β -globine

L'analyse détaillée des différents sites, isolés ou associés, a montré que la présence des quatre sites est nécessaire à la totalité de l'effet activateur [2]. Pour, cependant, procéder à une analyse, il faut d'abord récapituler l'ensemble des propriétés de ce LCR : *in vivo*, il est responsable de l'ouverture d'un domaine chromatinienn spécifique des tissus érythroïdes ; il a une fonction *enhancer* classique ; et, de plus, il sensibilise les gènes aux inducteurs endogènes (par exemple, l'érythropoïétine) ou pharmacologiques. Par ailleurs, dans les études *in vitro*, tous les événements d'intégration sont efficaces, d'où une expression qui est fonction du nombre de copies intégrées.

Nous récapitulons ce qui est su à l'heure actuelle des différents sites dans l'ordre de leur importance croissante :

- Le site HS1, le plus proche du gène ϵ , semble n'avoir qu'un effet additif par rapport aux autres. Des modèles naturels, comportant des délétions ne respectant que lui ou au contraire n'atteignant que lui, ont montré qu'il n'est ni nécessaire ni suffisant pour le bon fonctionnement des gènes du *locus* β -globine.

- Le site HS4 semblait, lui aussi n'avoir qu'une action additive. Une étude récente, cependant, a mis en évidence une séquence 5'GACTGG3'. Cette séquence ressemble au site CCAAT du promoteur de β -globine, et lie un facteur transactivateur non spécifique de tissu, ressemblant, mais non identique aux facteurs activateurs CP-1 et 2 (*m/s* n° 4, vol. 5, p. 252). La quasi-totalité de l'activité fonctionnelle du 5'HS4 est localisée dans un sous-

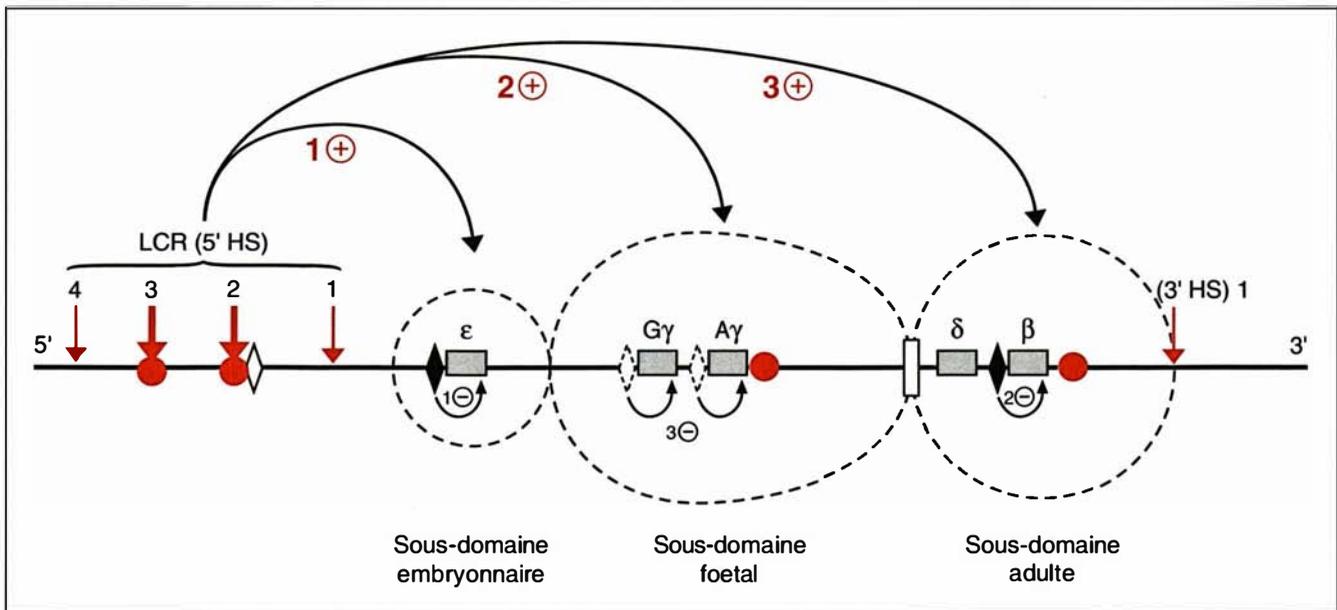


Figure 1. La figure représente les cinq gènes de structure du locus β -globine et les éléments régulateurs impliqués dans leur expression séquentielle au cours du développement. Les flèches rouges figurent les sites hypersensibles, 5'HS1, 2, 3 et 4 qui forment le LCR, et le site 3'HS1. Les flèches épaisses figurent les sites de régulation majeurs. Les enhancers sont représentés par un cercle rouge plein, aux sites 5'HS2 et 3, ainsi qu'en aval des gènes γ et β -globine. Les silenciers sont figurés sous forme de losanges noirs : en plein, les séquences identifiées pour cette fonction silencer, en amont des gènes ϵ - et β -globine ; en losange vide, une séquence identifiée, mais dont la fonction est incertaine, au niveau du 5'HS2 ; en pointillé, les zones auxquelles on attribuerait une fonction silencer, mais qui n'ont pas été identifiées, en amont des gènes foetaux. Les trois sous-domaines — embryonnaire — foetal et adulte — seraient activés successivement. Leur « extinction » est indiquée par un signe de régulation négative. Un rectangle vertical en amont du gène δ figure une séquence polypyrimidique qui pourrait être impliquée dans la transition du stade foetal au stade adulte.

fragment incluant cette séquence, qui cependant n'est pas active par elle-même, mais fortement potentialisatrice de l'action des sites 5'HS2 et 3. L'intérêt de cette constatation est de montrer qu'une régulation globale strictement érythroïde-spécifique requiert en fait une interaction complexe où interviennent des activateurs transcriptionnels spécifiques et ubiquitaires [3].

- Les sites HS2 et HS3 concentrent la majeure partie de l'activité du LCR. Dans des régions réduites à quelques centaines de paires de bases, certaines séquences sont retrouvées de façon répétitive, liant des facteurs spécifiques et non spécifiques : facteur GATA-1 le plus fréquent, facteur NFE-2, séquences GGTGG [4, 5].

- L'étude la plus détaillée du site HS3 a été faite en le couplant à un gène adulte β -globine, vis-à-vis duquel il présente une fonction *enhancer*. Curieusement, en système foetal (cellules K562), il ne semble lier aucun facteur diffusible, ce qui serait compatible avec

un rôle plus spécifique du stade adulte [6].

- Le site HS2 est sans doute celui qui présente la gamme la plus complète de propriétés régulatrices [7]. Il a été étudié par plusieurs équipes, et dans des systèmes d'expression variés ; cette étude a mis en évidence la totalité des fonctions signalées plus haut, même si certaines ne sont pas efficaces à 100 %. Outre des séquences, spécifiques de tissus ou non, analogues à celles que l'on trouve dans les autres sites, deux éléments sont caractéristiques du HS2. Deux séquences AP1, séparées par un court intervalle, conservent à elles seules une grande partie de l'activité *enhancer* [8]. Plus en 3', une séquence riche en purines/pyrimidines alternées pourrait avoir une fonction modulatrice négative [9]. Elle est polymorphe dans les populations humaines [10].

• Expression autonome et/ou compétitive des différents gènes

Les premières expériences semblaient

favoriser l'hypothèse d'une régulation compétitive (*m/s* n° 7, vol. 7, p. 740). Plus récemment, on a été amené à nuancer cette proposition :

- Le gène ϵ -globine a incontestablement une régulation autonome [11]. Ce fait permet d'attribuer un rôle au cours de l'ontogénie à une séquence extinctrice (*silencer*) dont l'existence a été démontrée en 5' du gène.

- Les gènes γ -globine, également, peuvent « éteindre » leur activité s'ils sont intégrés en même temps que leurs séquences flanquantes [12]. Le mécanisme de cet arrêt d'activité est encore mal élucidé.

- Différentes constructions recombinantes ont démontré que l'ordre physique des gènes vis-à-vis du LCR était important pour l'ordre de leur expression séquentielle, le gène proximal étant exprimé avant le gène distal. Cet ordre intervient donc dans l'expression ontogénique. L'organisation des gènes du locus β -globine se traduit de façon à la fois compétitive et polaire. Il s'agit d'une compétition non réciproque [12].

Une séquence riche en pyrimidines, située en amont des gènes δ , pourrait également avoir un rôle dans ce phénomène de polarité. Elle est, en effet, susceptible de modifier sa conformation secondaire en formant une triple hélice, et elle fixe un facteur transactivateur spécifiquement présent dans des cellules adultes ; cela évoque un rôle possible dans le développement des cellules hématopoïétiques et leur commutation vers l'état adulte (*m/s* n° 2, vol. 8, p. 188) [14].

• Données concernant le locus α -globine. Similitudes et différences

L'étude du locus α -globine est postérieure à celle du locus β -globine. Plusieurs données sont à l'origine de cette étude : a) deux gènes, situés sur deux chromosomes différents, sont soumis à une régulation telle que leurs produits d'expression sont équilibrés ; b) dans des souris transgéniques, l'expression d'un gène α -globine et de ses séquences flanquantes n'est pas détectable ; c) le couplage du même gène à un LCR β aboutit à une expression du même ordre de grandeur que celle des globines endogènes. L'hypothèse d'une zone comparable au LCR β se trouvait corroborée par l'existence de mutants naturels. Chez un malade porteur d'une hémoglobinose H, avec génotype $\alpha\text{-}/(\alpha\alpha)^0$, on a pu mettre en évidence une délétion *de novo* de 35 kb en 5' du locus, respectant intégralement l'ensemble des gènes [15].

Les premières étapes, recherche de sites hypersensibles et constructions permettant de tester le rôle de ces sites couplés au gènes α pour l'expression en souris transgéniques, ont permis de cerner une séquence de 350 pb, située à environ 40 kb en amont du gène ζ 2 (*m/s* n° 9, vol. 6, p. 918). On a, depuis, cherché à caractériser cette région de façon plus précise, par fragmentations successives et expression dans des cellules érythrocytaires murines (cellules MEL transformées par le virus de Friend), en utilisant le gène *neo* comme gène rapporteur ; les expériences d'empreinte (*footprint*) et les expériences de compétition ont confirmé la fixation sur ce segment de nombreux facteurs transactivateurs, les mêmes que ceux qui activent le locus β . On a ainsi identifié plusieurs empreintes GATA-1 et des séquences NFE-2. On trouve éga-

lement un couple de sites AP-1, ici inversés [16]. Dans ce cas comme dans celui du gène β -globine, se pose le problème de l'action activatrice de transcription, puissante, dans les cellules érythroïdes, d'un facteur AP-1 qui ne serait pas spécifique, et qui doit agir en combinaison avec des facteurs spécifiques, mais non activateurs par eux-mêmes, comme GATA-1, ou, autre hypothèse, du rôle de NFE-2, dont la séquence consensus recouvre celle de AP-1. Comme dans le cas du locus β , ce site, le plus important, ne semble néanmoins pas suffisant pour une expression maximale des gènes α -globine. D'autres régions hypersensibles, en particulier un site à - 33 kb, non actif à l'état isolé, pourraient avoir une action complémentaire.

A l'intérieur du locus α -globine, la régulation de l'expression des gènes embryonnaire ζ et fœtaux/adultes α n'est pas compétitive. Quant à la régulation coordonnée du locus α et du locus β , elle n'est encore qu'incomplètement expliquée. Outre un élément traductionnel connu depuis de nombreuses années, il existe sûrement un élément transcriptionnel dans lequel le rôle du LCR est évoqué.

• Manipulation pharmacologique

La production d'hémoglobine fœtale dans des cellules adultes a d'abord été étudiée *in vivo* chez le singe ou chez l'homme. La souris transgénique pourrait-elle être un modèle pour l'étude de ces inductions pharmacologiques ? L'expérience a été faite après transfert d'un gène γ -globine couplé à une construction ne conservant que les séquences essentielles de chacun des quatre sites hypersensibles (μ LCR- γ), l'expression du gène étant contrôlée à différents niveaux : cellulaire (production de cellules F, c'est-à-dire contenant de l'Hb F en quantité détectable par immunofluorescence), protéique (rapport $\gamma/\gamma + \beta$), production d'ARNm (PCR et hybridation). Dans ces conditions, les différents agents inducteurs connus (érythropoïétine, 5-azacytidine, hydroxyurée, butyrate) induisent tous une expression détectable des gènes γ . Afin d'éliminer une action isolée, soit sur les sites activateurs proximaux des gènes γ , soit sur les sites activateurs distaux du LCR, des contre-expériences ont été faites qui

n'utilisaient que les uns ou les autres : le transfert d'un gène γ sans LCR ou d'une construction μ LCR- β ont été sans résultat, ce qui atteste d'une synergie nécessaire entre les séquences proximales du gène γ et celles du LCR [17]. Même si ce modèle peut encore être affiné pour une meilleure définition des séquences impliquées, il est évidemment très important de disposer d'un modèle pharmacologique d'étude sur petit animal au moment où la réactivation des gènes fœtaux est envisagée comme une thérapeutique potentielle des hémoglobinopathies majeures, β -thalassémies et drépanocytose [18] ■

Dominique Labie

ICGM, centre hospitalier universitaire, Cochin, 24, rue du Faubourg, Saint-Jacques, 75014 Paris, France.

Rajagopal Krishnamoorthy

Inserm U.120, hôpital Robert-Debré, 48, boulevard Sérurier, 75019 Paris, France.

RÉFÉRENCES

1. Orkin SH. Globin gene regulation and switching : circa 1990. *Cell* 1990 ; 63 : 665-72.
2. Antoniou M, Grosveld F. β -globin dominant control region interacts differently with distal and proximal promoter elements. *Genes Dev* 1990 ; 4 : 1007-13.
3. Walters M, Kim C, Gelinas R. Characterization of a DNA binding activity in DNase I hypersensitive site 4 of the human globin locus control region. *Nucl Acids Res* 1991 ; 19 : 5385-93.
4. Collis P, Antoniou M, Grosveld F. Definition of the minimal requirements within the human β -globin gene and the dominant control region for high level expression. *EMBO J* 1990 ; 9 : 233-40.
5. Mignotte V, Eleouet JF, Raich N, Romeo PH. *Cis-* and *trans-*acting elements involved in the regulation of the erythroid promoter of the human porphobilinogen deaminase gene. *Proc Natl Acad Sci USA* 1989 ; 86 : 6548-52.
6. Ikuta T, Kan YW. *In vivo* protein-DNA interactions at the β -globin gene locus. *Proc Natl Acad Sci USA* 1991 ; 88 : 10188-92.
7. Tuan DYH, Solomon WB, London IM, Lee DP. An erythroid-specific, developmental-stage-independent enhancer far upstream of the human « β -like globin » genes. *Proc Natl Acad Sci USA* 1989 ; 86 : 2554-8.

RÉFÉRENCES

8. Ney PA, Sorrentino BP, McDonagh KT, Nienhuis AW. Tandem AP-1 binding sites within the human β -globin dominant control region function as an inducible enhancer in erythroid cells. *Genes Dev* 1990 ; 4 : 993-1006.
9. Moi P, Kan YW. Synergistic enhancement of globin gene expression by activator protein-1-like proteins. *Proc Natl Acad Sci USA* 1990 ; 87 : 9000-4.
10. Perichon B, Bennani M, Lu CY, Moi P, Krishnamoorthy R. β -globin locus control region is polymorphic in human populations. Congrès international sur les maladies génétiques de l'hémoglobine. Nice-Acropolis, 6-7-8 novembre 1991 : 156.
11. Raich N, Enver T, Nakamoto B, Josephson B, Papayannopoulou T, Stamatoyannopoulos G. Autonomous developmental control of human embryonic globin gene switching in transgenic mice. *Science* 1990 ; 250 : 1147-9.
12. Dillon N, Grosveld F. Human γ -globin genes silenced independently of other genes in the β -globin locus. *Nature* 1991 ; 350 : 252-4.
13. Hanscombe O, Whyatt D, Fraser P, *et al.* Importance of globin gene order for correct developmental expression. *Genes Dev* 1991 ; 5 : 1387-94.
14. O'Neill D, Bornschlegel K, Flamm M, Castle M, Bank A. A DNA-binding factor in adult hematopoietic cells interacts with a pyrimidine-rich domain upstream from the human δ -globin gene. *Proc Natl Acad Sci USA* 1991 ; 88 : 8953-7.
15. Liebhaber SA, Griese EU, Weiss I, *et al.* Inactivation of human α -globin gene expression by a *de novo* deletion located upstream of the α -globin gene cluster. *Proc Natl Acad Sci USA* 1990 ; 87 : 9431-5.
16. Jarman AP, Wood WG, Sharpe JA, Gourdon G, Ayyub H, Higgs DR. Characterization of the major regulatory element upstream of the human α -globin gene cluster. *Mol Cell Biol* 1991 ; 11 : 4679-89.
17. Constantoulakis P, Josephson B, Mangahas L, *et al.* Locus control region- γ transgenic mice : a new model for studying the induction of fetal hemoglobin in the adult. *Blood* 1991 ; 77 : 1326-33.
18. Ley TJ. The pharmacology of hemoglobin switching : of mice and men. *Blood* 1991 ; 77 : 1146-52.