

Les mécanismes de contrôle de l'expression des gènes α et β -globine ne sont pas identiques

Les travaux effectués ces dernières années ont fourni un modèle cohérent de régulation du *locus* β -globine. Un ensemble de séquences régulatrices situées en *cis*, appelées LCR (*locus control region*), activent successivement, par l'intermédiaire de facteurs transactivateurs, les différents gènes du *locus*, délimitant une unité d'expression distincte dans un domaine chromatinien. L'activation spécifique de tissu de ce LCR « ouvre » la structure chromatinienne, modifie le moment de la réplication, qui se produit plus précocement au cours du cycle cellulaire, et est marquée par l'apparition de sites hypersensibles à la DNase I (HSS). Les limites du domaine ainsi activé (environ 100 kb pour le *locus* β -globine) sont marquées par des séquences spécialisées de liaison à la matrice nucléaire, séquences importantes dans les transitions d'expression en fonction de la nature du tissu.

Ce modèle, très conservé au cours de l'évolution, est sans doute extrapolable

à d'autres familles géniques. La question se pose, en particulier, pour les gènes du *locus* α -globine, puisque l'expression de ces deux familles géniques, α et β , est strictement coordonnée. Des différences avaient cependant été notées dès l'origine. Ces différences concernent la structure des deux *loci* : le *locus* α est riche en GC, on y trouve de nombreux îlots CpG non méthylés, mais, en revanche, aucune séquence de liaison aux structures nucléaires n'a pu être mise en évidence sur plus de 200 kb. Il existe également des différences fonctionnelles : la réplication du *locus* α est précoce dans tous les tissus ; dans les systèmes expérimentaux, les gènes α sont exprimés de façon constitutive dans les lignées permanentes telles que les cellules ELM (érythroleucémie murine), sans que l'induction de la différenciation modifie cette expression ; dans les souris transgéniques, en revanche, on n'a pas réussi à faire exprimer les gènes α entourés seulement de leur environne-

ment immédiat. Une série de travaux, dont la plupart ont été effectués à Oxford par l'équipe de D. R. Higgs, montre les ressemblances et les différences dans la régulation des deux *loci* de globine.

Une première étape a été la démonstration de l'existence, loin en amont des gènes α -globine, d'un élément de régulation positive comparable au LCR β . Cet élément, situé à 40 kb en 5' du site Cap du gène ζ -globine, est marqué par des HSS spécifiques des cellules érythroïdes ; lié au gène α , il permet son expression dans les souris transgéniques (*m/s* n° 9, vol. 6, p. 918) [1]. Le rôle de cette région (HS-40) est confirmé par l'étude de mutants naturels. Quatre mutants ont actuellement été décrits, dans lesquels une délétion située en amont du *locus* α -globine, mais respectant les gènes eux-mêmes, se traduit par une non-expression de ces gènes. Ces délétions, d'étendue et de topographie variables, ont en commun d'impliquer le frag-

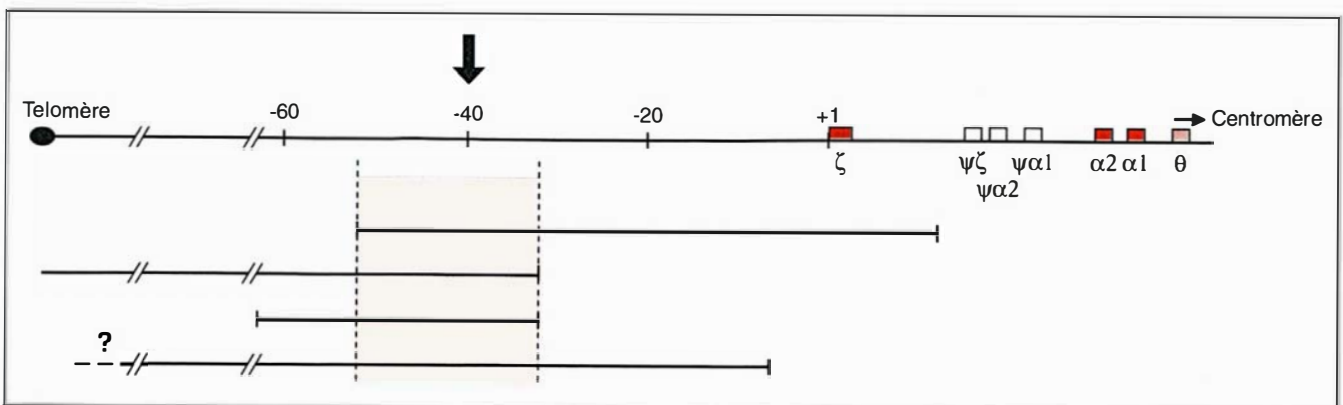


Figure 1. Représentation schématique du locus α -globine, de la région télomérique du bras court du chromosome 16 et de quatre délétions n'impliquant pas les gènes α eux-mêmes, mais empêchant leur expression. On voit que ces quatre délétions n'ont en commun qu'une zone s'étendant approximativement de -50 à -30 kb en amont du site de capping du gène ζ -globine, zone qui inclut le site hypersensible majeur HS-40. (D'après [2].)

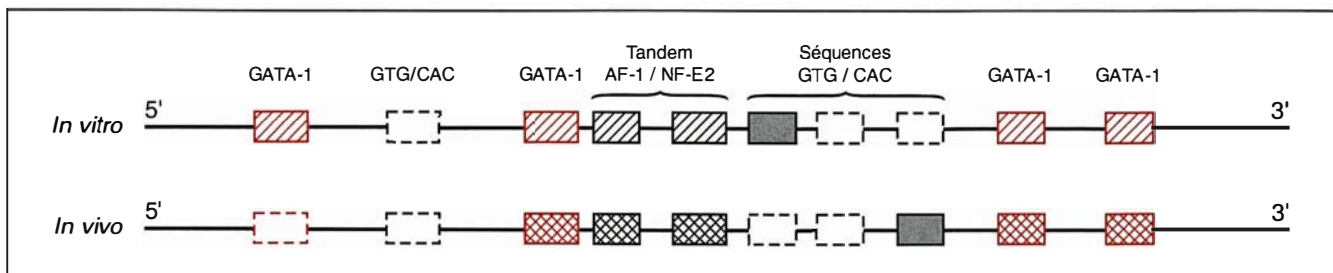


Figure 2. **Représentation schématique des facteurs transactivateurs qui se lient au fragment de 350 pb du site HS-40 : in vitro, c'est-à-dire sur ADN nu ; in vivo, c'est-à-dire en cellules entières.** En rouge, les sites GATA-1 spécifiques, dont trois seulement sont utilisés in vivo. En noir, les sites de fixation de facteurs ubiquitaires, de deux types : d'une part, le doublet de sites AP-1/NF-E2 (les expériences de compétition in vivo montrent en fait que le transactivateur lié est probablement NF-E2, c'est-à-dire un facteur de la famille de AP-1, mais spécifique des cellules érythroïdes) ; d'autre part, les séquences GGGTGG, fixant le facteur de la CAC box, inégalement occupées, et qui sont nécessaires à l'expression par une combinatoire d'interaction. (D'après[4].)

ment qui s'étend de - 50 à - 30 kb, le point + 1 étant le site Cap du gène ζ (figure 1) [2].

Les techniques classiques de déplacement sur gel, de compétition et d'empreintes ont ensuite été utilisées pour rechercher les facteurs transactivateurs par l'intermédiaire desquels s'exerçait la régulation du site HS-40. Toutes les séquences susceptibles de fixer ces facteurs ont été trouvées groupées sur un fragment de seulement 350 pb. Ces séquences sont de trois types : réparties sur les 350 pb, on trouve quatre séquences GATA-1, comme dans tous les gènes s'exprimant dans la lignée érythroïde ; deux séquences de type AP-1/NF-E2 existent, mais ici en position inversée, séparées par 26 pb ; enfin, une séquence GGGTGG, séquence CAC, voisine, fonctionne sans doute comme coactivateur [3]. L'ensemble de ces données n'est pas sans rappeler ce qu'on observe au niveau du site HS2 du LCR β (*m/s* n° 3, vol. 8, p. 255).

Cette exploration, faite *in vitro* sur un ADN « nu », aurait pu cependant être entachée d'erreur, par méconnaissance d'un facteur protéique présent en faible concentration ou être insensible à la conformation chromatinienne, et détecter alors des sites de fixation qui sont inutilisables *in vivo*. La première étude a donc été complétée par une autre similaire faite dans des cellules intactes après protection des sites susceptibles d'interagir avec les facteurs

protéiques. L'exploration de ces sites donne alors une image de l'interaction ADN-protéine *in situ*. On a retrouvé les trois mêmes types de séquence que dans l'étude précédente, ce qui démontre que les trois facteurs qui se lient à ces séquences représentent le minimum requis pour l'activation de l'élément régulateur. Cependant, seules trois séquences GATA-1 sont utilisées. Par ailleurs, les séquences de type AP-1 ont été explorées par des expériences de compétition : elles fixent le facteur NF-E2, de la famille AP-1, mais spécifique de la lignée érythroïde (figure 3). De cette étude ressortent deux faits essentiels. C'est d'abord le caractère étonnamment compact de ce LCR α . C'est ensuite que ces activations ne mettent en jeu qu'un petit nombre de facteurs, toujours les mêmes, mais dont la combinatoire semble hautement spécifique [4].

On a ensuite cherché à savoir si le gène α -globine, comme le gène β , était présent dans un domaine chromatinienn distinct, spécifique des cellules érythroïdes, dont l'ouverture serait un trait caractéristique de la mise en activité. Deux approches ont été utilisées pour l'étude comparative de la chromatine des loci α - et β -globine : l'exploration de la sensibilité à la DNase I, et la caractérisation d'autres gènes éventuellement présents dans cette région d'ADN. On constate que les HSS spécifiques des cellules érythroïdes sont répartis sur environ 65 kb à partir du

HS-40, ce qui définirait la taille minimale d'un domaine érythroïde. Les HSS ubiquitaires, en revanche, sont disséminés sur l'ensemble des 200 kb explorés.

L'existence d'îlots CpG non méthylés marque habituellement l'extrémité 5' d'un gène, elle est constante à l'extrémité de tous les gènes d'expression ubiquitaire (gènes *housekeeping*) ; nous avons vu que l'abondance de ces îlots CpG était une caractéristique de l'environnement du locus α -globine. Ils ont été recherchés systématiquement à l'aide d'une enzyme spécifique. Outre ceux qui correspondent aux différents gènes de la famille α , quatre îlots CpG plus éloignés en 5', échelonnés entre - 99 et - 14 kb, ont été identifiés. Des ARN messagers ont été alors recherchés avec des sondes correspondant à chacun de ces îlots, et retrouvés dans toutes les lignées cellulaires, érythroïdes et non érythroïdes. Aucun transcrit, en revanche, n'a été mis en évidence par une sonde comprenant la séquence HS-40. Ce fait, ainsi que l'étude détaillée des ADN complémentaires correspondants obtenus par le criblage de différentes banques cellulaires, a permis une caractérisation partielle des gènes en cause, qui sont tous de nature ubiquitaire, et orientés de façon variable, l'un vers le centromère, comme l'ensemble du locus α , les autres de façon inverse, vers le télomère. Leur comparaison avec différentes banques d'ADNc d'origine animale

(*zoo libraries*) montre une très grande conservation au cours de l'évolution. Enfin, le site HS-40 se situe dans un intron du plus grand de ces gènes, lui-même d'orientation télomérique (figure 3). Quelques conclusions majeures, ici encore, ressortent de cette étude. C'est d'abord la très haute densité de gènes dans cette région télomérique du bras court du chromosome 16. On sait que les gènes actifs ne sont pas répartis également dans l'ensemble du génome, et il serait intéressant d'explorer, dans cette perspective d'autres télomères. On doit remarquer aussi l'existence, jamais observée jusqu'alors, d'une séquence régulatrice majeure présente à l'intérieur même d'un gène hétérologue. Cette séquence HS-40 est, en fait, physiquement plus proche du promoteur de ce gène d'expression ubiquitaire qu'elle ne l'est des promoteurs du locus α -globine. Le fait qu'elle soit active sur ces derniers et non sur le premier pose un problème de régulation non encore résolu. Enfin, si le HS-40 se situe dans le corps d'un gène exprimé dans tous les tissus, il doit être présent dans une

structure de chromatine ouverte en permanence. Cela expliquerait la réplification précoce et attire l'attention sur une importante différence entre le HS-40 α et le LCR β : la fonction d'ouverture de la chromatine n'est sans doute pas utile au premier [5] alors qu'elle est essentielle à la fonction du second.

Un modèle de compétition se référant à la position des gènes par rapport au LCR a été proposé pour expliquer l'activation successive des différents gènes du locus β (*m/s n° 3, vol. 8, p. 255*). Des expériences ont cherché à établir s'il en était de même pour les gènes du locus α . Différentes constructions ont été introduites dans des souris transgéniques, et leur expression mesurée, par comparaison avec les gènes endogènes, aux étapes successives du développement. Aucune interdépendance n'a pu être trouvée entre l'expression des gènes ζ embryonnaires et celle des gènes α fœtaux/adultes. Ces gènes sont chacun soumis à une régulation autonome contemporaine de celle des gènes endogènes [6]. En revanche, une commutation par-

tielle a été mise en évidence entre les deux gènes α eux-mêmes : à partir d'une expression égale de α^2 et α^1 , on voit se développer progressivement la prédominance de α^2 , telle qu'elle sera observée pendant toute l'existence adulte [7].

Une dernière différence entre les zones régulatrices des deux loci de globine doit être signalée. Parmi les caractères du LCR, il est classique de mettre en évidence, dans les expériences de souris transgéniques, une expression proportionnelle au nombre de copies intégrées, si celui-ci n'est pas trop élevé. Cette propriété n'est pas retrouvée dans les expériences utilisant le HS-40. Le taux d'expression des gènes transfectés sous le contrôle du HS-40 est en général voisin de 50 % de celui des gènes endogènes. Il n'augmente pas quand le nombre de copies intégrées augmente, le taux d'expression par copie étant donc inversement proportionnel au nombre de copies, et cela quelle que soit la construction employée (D. R. Higgs, communication personnelle, et Eighth Conference on Hemoglobin Switching, Rosario

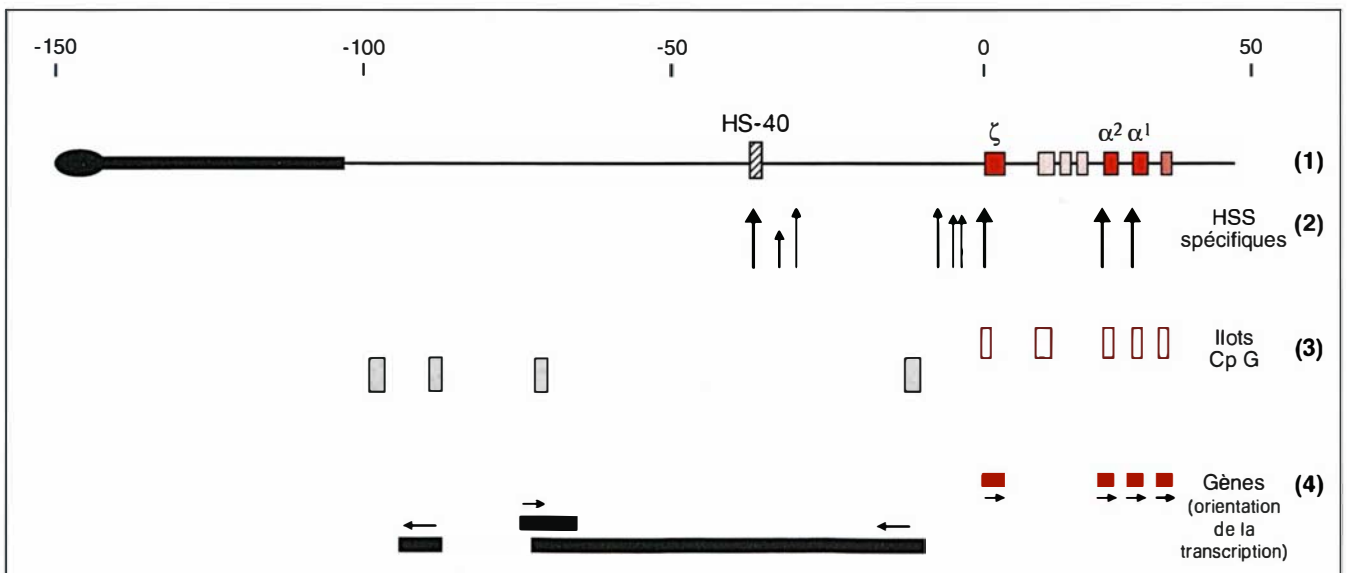


Figure 3. **Représentation schématique de toute l'extrémité télomérique du bras court du chromosome 16 (environ 200 kb).** On a représenté, de haut en bas : (1) Sur la figuration du chromosome lui-même, les gènes du locus α -globine et la zone régulatrice HS-40 ; en traits épais, la région télomérique et subtélomérique ; (2) Les sites hypersensibles des cellules érythroïdes spécifiques, groupés sur 65 kb, alors que des sites non spécifiques sont répartis sur toute la longueur des 200 kb ; (3) Les îlots CpG non méthylés — en rouge, ceux qui correspondent aux gènes de globine ; en gris, ceux qui correspondent à des gènes d'expression ubiquitaire, en cours de caractérisation ; (4) Les gènes identifiés et leur orientation : en rouge, les gènes de globine ; en noir, des gènes d'expression ubiquitaire. (D'après [5].)

Resort, Orcas Island, Washington, 29 mai-2 juin 1992).

On voit donc que cet ensemble très cohérent de résultats laisse encore des questions non résolues. Il restera aussi à comprendre l'expression parfaitement synchronisée et quantitativement équilibrée des gènes des deux *loci*, peut-être sous l'action coordonnée des mêmes facteurs transactivateurs, spécifiques ou ubiquitaires ■

Dominique Labie

ICGM, CHU Cochin, 24, rue du Faubourg-Saint-Jacques, 75014 Paris, France.

RÉFÉRENCES

1. Higgs DR, Wood WG, Jarman AP, *et al.* A major positive regulatory region located far upstream of the human α -globin gene locus. *Genes Dev* 1990 ; 1588-601.
2. Romao L, Osorio-Almeida L, Higgs DR, Lavinha J, Liebhaber SA. α -thalassemia resulting from deletion of regulatory sequences far upstream of the α -globin structural genes. *Blood* 1991 ; 78 : 1589-95.
3. 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.
4. Strauss EC, Andrews NC, Higgs DR, Orskin SH. *In vivo* footprinting of the human α -globin locus upstream regulatory element by guanine and adenine ligation-mediated polymerase chain reaction. *Mol Cell Biol* 1992 ; 12 : 2135-42.
5. Vyas P, Vickers MA, Simmons DL, Ayyub H, Craddock CF, Higgs DR. Cis-acting sequences regulating expression of the human α -globin cluster lie within constitutively open chromatin. *Cell* 1992 ; 69 : 781-93.
6. Albitar M, Katsumata M, Liebhaber SA. Human α -globin genes demonstrate autonomous developmental regulation in transgenic mice. *Mol Cell Biol* 1991 ; 11 : 3786-94.
7. Albitar M, Cash FE, Peschle C, Liebhaber SA. Developmental switch in the relative expression of the α^1 - and α^2 -globin genes in humans and in transgenic mice. *Blood* 1992 ; 79 : 2471-4.

TIRÉS A PART

D. Labie.

m/s n° 8, vol. 8, octobre 92