

## Régulation de l'expression des gènes du collagène de type I : implications pour l'étude de la sclérodermie

Le collagène de type I, la protéine la plus abondante, est synthétisée, chez les mammifères, principalement par les fibroblastes et les ostéoblastes. Cette synthèse est augmentée dans tous les types de fibroses et dans la sclérodermie. Il s'agit d'un hétérotrimère, formé de deux chaînes  $\alpha 1$  et d'une chaîne  $\alpha 2$ . La stœchiométrie 2:1, observée aussi bien pour la synthèse polypeptidique que pour celle des ARNm, a fait rechercher, chez la souris, des éléments de régulation communs aux deux gènes *colla1* et *colla2*, et a permis la découverte d'un nouveau facteur de transcription C-Krox. Son expression, sous la dépendance de facteurs de croissance, a lieu dans les fibroblastes et non dans les ostéoblastes. L'approfondissement des connaissances concernant les régulateurs des gènes de collagène devrait permettre de mieux comprendre leur rôle dans l'augmentation de synthèse du collagène I dans les fibroses et la sclérodermie et, peut-être, d'ouvrir de nouvelles voies thérapeutiques.

Philippe Galéra  
Christelle Desbois  
Patricia Ducy  
Gérard Karsenty

### ADRESSE

P. Galéra : *chercheur post-doctoral*. C. Desbois : *chercheur post-doctoral*. P. Ducy : *chercheur post-doctoral*. G. Karsenty : *assistant professor*. MD Anderson Cancer Center, the university of Texas, department of molecular genetics, Box 45, 1515 Holcombe boulevard, Houston, Texas 77030, États-Unis.

### TIRÉS À PART

G. Karsenty.

Les collagènes constituent une famille de glycoprotéines qui sont les composantes majeures de la matrice extracellulaire. A ce jour, seize isotopes différents ont été décrits chez les vertébrés, certains d'entre eux ayant une expression tissulaire restreinte [1]. Les collagènes sont constitués de trois chaînes polypeptidiques associées en une triple hélice; les chaînes varient selon les types mais pour une même isoforme, elles peuvent être identiques (collagène de type II) ou

différentes (collagène de type I: deux chaînes  $\alpha 1$  (I) et une chaîne  $\alpha 2$  (I)). Selon la structure de leur triple hélice, les collagènes ont été répartis en deux catégories. D'une part, les collagènes fibrillaires ou interstitiels, possédant des triples hélices continues [2] et regroupant les isotopes I, II, III, V et XI. D'autre part, les collagènes non fibrillaires, caractérisés par de fréquentes interruptions de leur triple hélice et représentés par les collagènes IV, VI, VII, VIII, IX, X et XII, entre autres.

## Le collagène de type I

Le collagène de type I, l'archétype des collagènes interstitiels, est la protéine la plus abondante du corps. Il est synthétisé principalement dans les os, les tendons et la peau. La synthèse du collagène de type I est restreinte essentiellement à deux types cellulaires : ostéoblastes et fibroblastes. Les deux chaînes polypeptidiques  $\alpha 1(I)$  et  $\alpha 2(I)$  sont produites dans un rapport stœchiométrique 2:1, de même que leurs ARNm respectifs. Les mécanismes de la régulation coordonnée de l'expression des deux gènes du collagène de type I (*colla1* et *colla2*) ne sont pas connus. Il est capital de comprendre comment cette corégulation s'effectue si l'on veut comprendre comment la synthèse du collagène de type I peut être altérée dans certaines maladies. Une hypothèse qui permettrait d'expliquer la régulation coordonnée de la synthèse des chaînes  $\alpha 1(I)$  et  $\alpha 2(I)$  au niveau transcriptionnel est que des éléments régulateurs ou éléments *cis* communs seraient présents dans les deux gènes, situés à la même distance du site d'initiation de transcription et lieraient des facteurs de transcription identiques exerçant la même fonction. L'expression du collagène de type I est modifiée dans un grand nombre de situations physiologiques, comme le développement embryonnaire et le processus de cicatrisation, où l'on observe une augmentation transitoire de la synthèse du collagène de type I. Au cours du développement embryonnaire de la souris, la synthèse du collagène de type I apparaît dès le huitième jour du développement [3, 4]. La synthèse de la chaîne  $\alpha 1(I)$  est essentielle puisque, en son absence, l'embryon meurt *in utero* après treize jours de gestation [5]. La synthèse du collagène de type I est également accrue dans diverses situations pathologiques sévères, toutes marquées par une réaction fibrotique. Une réaction fibrotique est une caractéristique commune à de nombreuses maladies telles que la sclérodémie, l'arthrose, l'athérosclérose, la cirrhose hépatique, la fibrose pulmonaire, etc. D'un point de vue histologique, la fibrose se caractérise par une accumulation anormale de collagènes de types I et III, ainsi que d'autres protéines de la matrice extracellulaire.

Cette accumulation exagérée de certaines protéines de la matrice extracellulaire est à l'origine de la symptomatologie clinique biologique observée dans ces maladies. De nombreuses études ont montré que, indépendamment de la cause d'une maladie fibrotique donnée, il existe dans la majorité de ces maladies une élévation de la transcription des deux gènes du collagène de type I et probablement d'autres gènes codant pour des protéines de la matrice extracellulaire. Parmi les nombreuses maladies fibrotiques, la sclérodémie constitue un modèle privilégié pour étudier la régulation de l'expression des gènes du collagène de type I. La sclérodémie est une maladie acquise, probablement d'origine auto-immune, caractérisée par un dépôt excessif de tissu conjonctif dans le derme et dans certains organes internes. Dans les fibroblastes dermiques de patients atteints de sclérodémie, une augmentation de l'activité transcriptionnelle des gènes *colla1* et *colla2* a été démontrée. L'étude de la régulation de l'expression de ces gènes dans les fibroblastes est donc particulièrement importante si l'on veut comprendre la physiopathologie moléculaire de la sclérodémie. L'approche que nous avons utilisée pour comprendre les mécanismes qui conduisent à cette augmentation de l'activité transcriptionnelle des gènes du collagène de type I dans la sclérodémie et d'autres maladies consiste tout d'abord à élucider les mécanismes qui contrôlent la régulation de l'expression de ces gènes dans des conditions physiologiques, puis à étudier, dans un second temps, leurs perturbations éventuelles dans le processus fibrotique. Cet article se propose de présenter les résultats récents concernant la régulation de l'expression des gènes codant pour les chaînes du collagène de type I afin de répondre à quelques questions précises : quels sont les principaux éléments *cis* qui gouvernent l'expression de ces gènes ? Quels sont les facteurs de transcription qui interagissent avec ces séquences *cis* ? Quels sont les éléments régulateurs communs aux deux gènes codant pour les deux chaînes du collagène de type I ? Existe-t-il des facteurs de transcription spécifiques des fibroblastes qui puissent être impliqués dès les stades initiaux d'une réaction fibrotique ?

## RÉFÉRENCES

1. Van Der Rest M, Garrone R. Collagen family of proteins. *FASEB J* 1991 ; 5 : 2814-23.
2. Myers JC, Emanuel BS. Chromosomal localization of human collagen genes. *Coll Rel Res* 1987 ; 7 : 149-59.
3. Adamson ED, Ayers SE. The localization and synthesis of some collagen types in developing mouse embryos. *Cell* 1979 ; 16 : 953-65.
4. Leivo I, Vaheri A, Timpl R, Wartiovaara J. Appearance and distribution of collagens and laminin in the early mouse embryo. *Dev Biol* 1980 ; 76 : 100-14.
5. Harbers K, Kuehn M, Delius H, Jaenisch R. Insertion of retrovirus into the first intron of  $\alpha 1(I)$  collagen gene leads to embryonic lethal mutation in mice. *Proc Natl Acad Sci USA* 1984 ; 81 : 1504-8.
6. Rippe RA, Lorensen SI, Brenner DA, Breindl M. Regulatory elements in the 5' flanking region and the first intron contribute to transcriptional control of the mouse  $\alpha 1$  type I collagen gene. *Mol Cell Biol* 1989 ; 9 : 2224-7.
7. Karsenty G, De Crombrughe B. Two different negative and one positive regulatory factors interact with a short promoter segment of the  $\alpha 1(I)$  collagen gene. *J Biol Chem* 1990 ; 265 : 9934-42.
8. Hatamochi A, Golumbek PT, Van Schaftingen E, De Crombrughe B. A CCAAT DNA binding factor consisting of two different components that are both required for DNA binding. *J Biol Chem* 1988 ; 263 : 5940-7.

Quelles sont les principales cytokines qui règlent l'expression des gènes du collagène de type I et comment agissent-elles?

### Les éléments régulateurs ou éléments cis du gène *col1a1*, codant pour la chaîne $\alpha 1(I)$

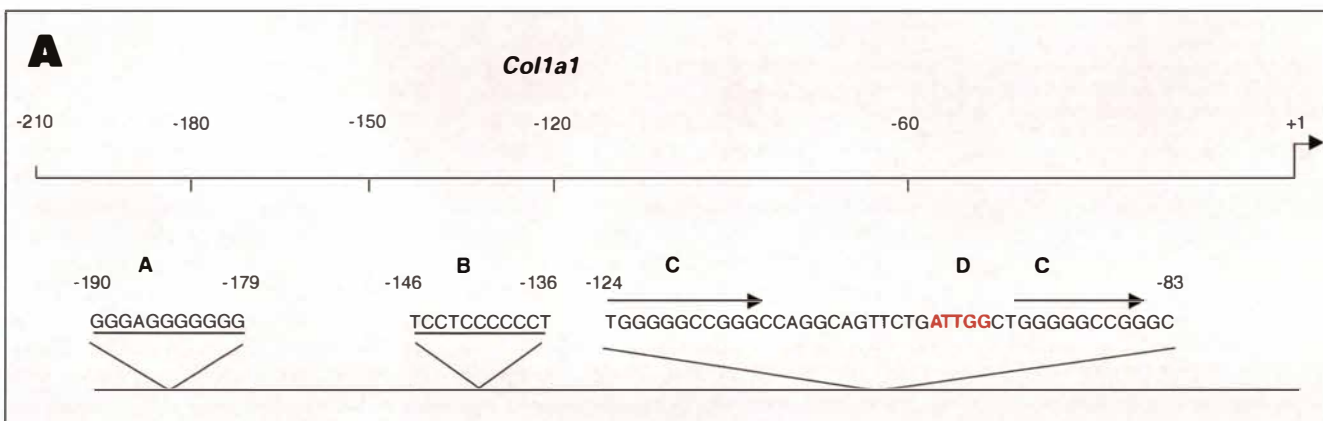
Au cours de ces dernières années, les travaux de nombreux laboratoires ont permis d'identifier des éléments *cis* présents de part et d'autre du site d'initiation de transcription du gène *col1a1*. Les premiers éléments *cis* identifiés le furent dans le premier intron du gène *col1a1* humain qui comporte à la fois des éléments *cis* activateurs et inhibiteurs. Du fait de la complexité de l'organisation de cet intron, il est cependant encore difficile de déterminer le rôle exact de ces divers éléments

dans la régulation et la spécificité tissulaire d'expression de ce gène. Un plus grand nombre d'études ont porté sur les séquences régulatrices situées en amont du site d'initiation de transcription. Deux études initiales ont montré qu'un fragment court du promoteur du gène *col1a1* de souris, contenant les séquences comprises entre -220 et +116, présente une forte activité transcriptionnelle dans les fibroblastes, mais non dans des cellules ne produisant pas de collagène de type I. Ce promoteur minimal représente, en fait, l'un des promoteurs les plus actifs parmi les promoteurs eucaryotes répertoriés, et l'addition de séquences plus en amont tend à en décroître l'activité [6, 7]. Ces résultats suggèrent que les 220 paires de bases du promoteur minimal contiennent une partie des éléments *cis* indispensables pour l'expression spécifique du gène dans les fibroblastes, et que ces

éléments sont nécessaires pour l'obtention d'un niveau d'expression maximal. Compte tenu de la taille du gène *col1a1* (plus de 30 kb) et de sa double spécificité cellulaire d'expression (ostéoblaste et fibroblaste), il apparaît évident que d'autres éléments *cis* doivent être présents, ailleurs, dans ce gène. L'analyse aussi précise que possible de la structure de ce promoteur minimal constitue cependant une première étape nécessaire pour comprendre les mécanismes plus complexes gouvernant l'expression de ce gène.

### Quatre éléments *cis* liant trois facteurs de transcription différents sont présents dans le promoteur minimal *col1a1*

L'approche utilisée pour identifier les éléments *cis* et définir leurs fonctions a été la suivante. (1) Initialement, l'existence de régions liant des pro-



**B** oligonucléotides  $\alpha 1(I)$

		* * *	
<b>C+D</b>	5'	TTCCAAATGGGGGCCGGGCCAGGCAGTTCTGATTGGCTGGGGGCCGGGCTG	3'
	3'	AAGTTTAACCCCGGCCCGTCCGTCAAGACTAACCAGACCCCGGCCGAC	5'
		** *	
<b>B</b>	5'	CCTTCCTTTCCTCCTCCCCCTCTCG	3'
	3'	GGAAGGAAAGGGAGGAGGGGGAGAAGC	5'
		***	
<b>C</b>	5'	TTGCGGGAGGGGGCGCGCTGGGTGGAC	3'
	3'	AACGCCCTCCCCCGCGACCCACCTG	5'

Figure 1. **Séquence et localisation des sites de liaison à l'ADN de protéines présentes dans les extraits nucléaires de fibroblastes NIH-3T3 au niveau du promoteur *col1a1*.** A. Les séquences promotrices mentionnées ont été identifiées dans des expériences de protection à la DNase (figure 2). Le motif CCAAT proximal qui correspond à l'élément D est présenté en rouge. L'élément C, qui entoure le pentanucléotide CCAAT et correspond au site de liaison Sp1 est caractérisé par une flèche. Les éléments A et B qui lient un nouveau facteur de

transcription sont soulignés. B. Séquence des oligonucléotides double brin utilisés dans les expériences de retardement sur gel. Les mutations sont indiquées par des astérisques. Les mutations au niveau de l'élément A remplacent trois nucléotides G par trois A. Les trois mutations au niveau de l'élément B échan- gent les nucléotides C contre des résidus G. La mutation de trois paires de bases au niveau de l'élément C distal change G--G-G en C--C-T.

## RÉFÉRENCES

9. Maity SN, Vuorio T, De Crombrughe B. The B subunit of a rat heteromeric CCAAT-binding transcription factor shows a striking sequence identity with the yeast Hap 2 transcription factor. *Proc Natl Acad Sci USA* 1990; 87: 5378-82.
10. Vuorio T, Maity SN, De Crombrughe B. Purification and molecular cloning of the «A» chain of a rat heteromeric CCAAT-binding protein. Sequence identity with the yeast Hap 4 transcription factor. *J Biol Chem* 1990; 265: 22480-6.
11. Maity SN, Golumbek PT, Karsenty G, De Crombrughe B. Selective activation of transcription by a novel CCAAT binding factor. *Science* 1988; 241: 582-5.
12. Karsenty G, Ravazzolo R, De Crombrughe B. Purification and functional characterization of a DNA-binding protein that interacts with a negative element in the mouse  $\alpha 1(I)$  collagen promoter. *J Biol Chem* 1991; 266: 24842-8.
13. Rossi P, Karsenty G, Roche NS, Roberts AB, Sporn MB, De Crombrughe B. A nuclear factor 1 binding site mediates the transcriptional activation of a type I collagen promoter by transforming growth factor- $\beta$ . *Cell* 1988; 52: 405-14.
14. Niederreither K, D'Souza RN, De Crombrughe B. Minimal DNA sequences that control the cell lineage-specific expression of the pro $\alpha 2(I)$  collagen promoter in transgenic mice. *J Cell Biol* 1992; 119: 1361-70.
15. Galéra P, Musso M, Ducy P, Karsenty G. C-Krox, a novel transcriptional regulator of type I collagen gene expression is preferentially expressed in skin. *Proc Natl Acad Sci USA* 1994; 91: 9372-6.
16. Sham MH, Vesque C, Nonchev S, Marshall H, Frain M, Das Gupta R, Whiting J, Wilkinson D, Charnay P, Krumlauf R. The zinc finger gene *krox-20* regulates *hox B2* (*hox 2,8*) during hindbrain segmentation. *Cell* 1993; 72: 183-96.
- téines nucléaires a été déterminée par des expériences de protection à la DNase I. Ces expériences ont démontré l'existence de quatre éléments, fixant des protéines présentes dans les extraits nucléaires de fibroblastes NIH-3T3. Ces régions ont été appelées A, B, C et D, respectivement de l'extrémité 5' vers l'extrémité 3' du promoteur (figure 1). (2) Une fois la séquence de ces éléments régulateurs connue, des oligonucléotides double brin, contenant la séquence sauvage ou mutée de chacun de ces éléments, ont été utilisés dans des expériences de liaison directe et de compétition en retardement de gel. Ces expériences ont permis d'établir la spécificité de la liaison observée, et d'effectuer une caractérisation biochimique sommaire des facteurs de transcription se fixant sur ces éléments *cis*. (3) Après avoir défini quatre éléments *cis* dans ce promoteur minimal, des mutations abolissant la liaison de chacun des facteurs de transcription à leurs sites de liaison respectifs ont été introduites, et leurs effets testés par des expériences de transfections transitoires de fibroblastes NIH-3T3 de souris. Cette stratégie a permis, entre autres, de définir un élément *cis* liant un nouveau facteur de transcription préférentiellement exprimé dans les fibroblastes de la peau.

### L'élément D lie un activateur de transcription appelé CCAAT binding factor (CBF)

En amont de la séquence TATAAA ou boîte TATA (située à -26 en amont du site d'initiation de transcription), un élément centré par une séquence CCAAT est situé entre -96 et -100. Le facteur de transcription CBF se liant à ce motif a été purifié jusqu'à homogénéité. CBF est différent des autres facteurs déjà connus qui se lient à la boîte CCAAT tels que CTF/NF1 et CEBP [8]. Il s'agit d'un facteur hétéromérique composé d'au moins trois sous-unités (CBF-A, CBF-B, CBF-C) qui sont nécessaires pour sa liaison à l'ADN. Les deux sous-unités CBF-A et CBF-B ont été clonées [9, 10]. L'élément *cis* liant CBF est un activateur de la transcription du promoteur de gène *colla1* [7, 11]; une substitution d'un A par un T dans le pentanucléotide CCAAT décroît fortement la liaison de CBF à la boîte CCAAT et réduit l'activité du promoteur *colla1* de cinq

fois dans des expériences de transfections transitoires de fibroblastes NIH-3T3 [7]. La démonstration que CBF lui-même est un activateur de transcription a été obtenue par des expériences de transcription *in vitro*: l'addition de CBF purifié à des extraits nucléaires de fibroblastes NIH-3T3 préalablement déplétés en CBF par chromatographie d'affinité, stimule et initie correctement la synthèse de l'ARNm du gène *colla1* [11]. CBF est une protéine ubiquitaire et, si elle joue un rôle dans l'expression spécifique des gènes du collagène de type I, cela ne peut être qu'en conjonction avec d'autres facteurs ou en induisant une modification de la chromatine uniquement dans les fibroblastes.

### L'élément C lie une protéine nucléaire apparentée à Sp1

Situé en 5' et 3' du pentanucléotide CCAAT (entre -124 et -113 et entre -94 et -82), il existe deux séquences de douze paires de bases riches en G et C (5'-TGGGGGCCGGGC-3'). Des expériences de retardement de gel ont montré qu'une protéine différente de CBF, nécessitant la présence d'ions zinc dans le milieu de liaison, se fixait spécifiquement à cette séquence. L'introduction d'une mutation de trois paires de bases, abolissant la liaison de cette protéine à l'élément C dans le promoteur *colla1*, a pour effet d'augmenter l'activité transcriptionnelle du promoteur d'un facteur dix dans des expériences de transfections transitoires de fibroblastes de souris. Ce résultat expérimental, suggérant l'existence d'un élément répresseur dans le promoteur minimal, est en apparence contradiction avec l'activité transcriptionnelle extrêmement forte du promoteur sauvage. Plus surprenant encore, la purification et le micro-séquençage de la protéine se fixant à l'élément C a révélé que ce facteur avait une forte analogie avec Sp1, l'activateur de transcription humain le plus étudié (Desbois et Karsenty, résultats non publiés).

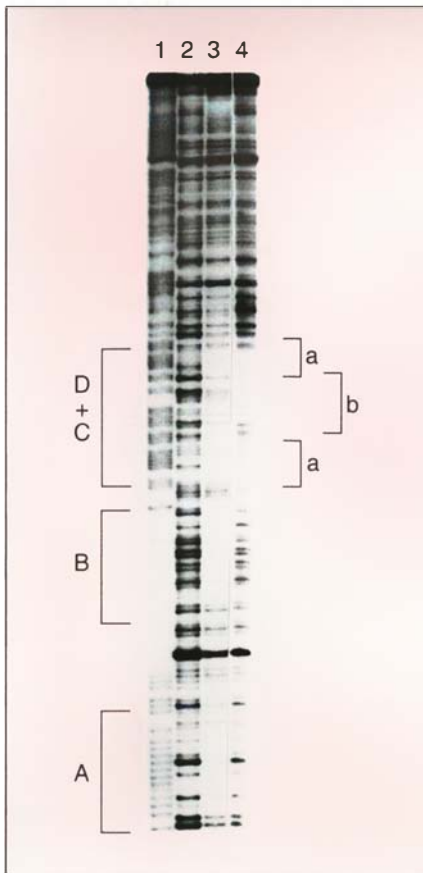


Figure 2. **Analyse du promoteur *col1a1* dans des expériences de protection à la DNase I.** Un fragment *EcoRI-HindIII* du promoteur *col1a1* de type sauvage ou mutant, marqué à son extrémité 5' *EcoRI*, a été incubé en présence ou en absence d'extraits nucléaires de fibroblastes NIH-3T3 et traité par l'ADNase I. **Ligne 1:** réaction de séquençage de Maxam-Gilbert (G+A) du fragment de type sauvage marqué à son extrémité 5'. **Ligne 2:** profil de digestion par la DNase I du fragment d'ADN de type sauvage incubé en absence d'extraits nucléaires de fibroblastes NIH-3T3. **Ligne 3:** fragment de promoteur de type sauvage incubé en présence d'extraits nucléaires de fibroblastes NIH-3T3. **Ligne 4:** fragment de promoteur *col1a1* mutant avec une substitution de trois paires de bases au niveau de l'élément C distal. Les aires de protection sont indiquées sur la gauche. Les indications sur la droite montrent les aires de protection pour les deux séquences répétées de 12 paires de bases qui correspondent au site de liaison *Sp1* (a) et le pentanucléotide CCAAT (b).

### Les éléments A et B ont une séquence similaire et lient un facteur de transcription présent dans les extraits nucléaires de fibroblastes

L'élément A est situé entre -190 et -170 par rapport au site d'initiation de transcription et le site B entre -150 et -130. Ces deux éléments ont une séquence riche en G (figure 1). Des expériences de compétition en retardement de gel ont montré qu'une protéine identique se fixait à ces deux sites. La liaison à l'ADN de la protéine présente dans les extraits nucléaires est inhibée par des concentrations croissantes d'EDTA, suggérant que ce facteur pourrait être une protéine à doigts de zinc. Des mutations qui abolissent la liaison de ce facteur à l'élément A ou à l'élément B ont été introduites dans le promoteur du gène *col1a1*, et l'activité de ce promoteur muté a été étudiée dans des expériences de transfections transitoires de fibroblastes NIH-3T3. Ces mutations introduites dans l'élément B ou l'élément A augmentent l'activité transcriptionnelle du promoteur  $\alpha 1(I)$ .

### *Sp1* et le facteur, se liant aux éléments A et B, inhibent partiellement la liaison de CBF ou à l'ADN

Comment expliquer que ce promoteur minimal soit très actif, alors que trois éléments *cis* sur quatre identifiés semblent agir comme des inhibiteurs de la transcription ? Des expériences de liaison de protéines à l'ADN *in vitro* suggèrent qu'il existe une inhibition de liaison réciproque entre CBF, d'une part, et *Sp1* et le facteur se liant aux éléments A et B, d'autre part. Comme nous l'avons mentionné, sur le promoteur de type sauvage quatre régions sont protégées : A, B, C et D (figure 2, ligne 3). Cependant, la protection autour de l'élément D centré par le pentanucléotide CCAAT est toujours partielle. Si, dans ces expériences, nous utilisons comme matrice un promoteur contenant une mutation au niveau de l'élément C qui abolit la liaison de facteurs nucléaires à cette région, trois résultats importants sont observés. Tout d'abord, les régions correspondant à l'élément C (figure 2, ligne 4, a et b) ne sont plus protégées, ce qui était attendu. De plus, les séquences correspondant aux éléments A et B ne sont plus pro-

tégées. Enfin, la protection autour du pentanucléotide CCAAT est maintenant quasi totale. Ces expériences, effectuées avec des extraits nucléaires non purifiés, suggèrent que les protéines se fixant aux éléments A, B et C forment un complexe entre elles et préviennent partiellement la liaison de CBF à la boîte CCAAT. Du fait de la forte activité transcriptionnelle de ce promoteur, on peut suggérer que ces protéines sont des activateurs de transcription. De fait, la protéine se liant à l'élément C est un homologue murin de *Sp1*, un puissant activateur de transcription.

### Les mêmes éléments *cis* régulent l'activité du promoteur du gène *col1a2*, codant pour la chaîne $\alpha 2(I)$

Si les éléments que nous avons décrits auparavant sont impliqués dans le contrôle coordonné de l'expression des gènes du collagène de type I, alors ils doivent être présents dans le promoteur du gène *col1a1*. Plusieurs éléments régulateurs *cis* ont été identifiés dans un fragment du promoteur du gène *col1a1* entre -500 et +1.

CBF, le facteur de transcription hétéromérique se liant à la séquence CCAAT entre -96 et -100 dans le promoteur du gène *col1a1* se fixe également à la même séquence CCAAT située entre -84 et -80 dans le promoteur du gène *col1a2*. CBF accroît aussi l'activité transcriptionnelle du promoteur *col1a2 in vitro*. Des séquences similaires aux éléments A, B et C sont présents approximativement à la même distance du site d'initiation de transcription dans le promoteur *col1a2*. Des études de liaison de facteurs nucléaires à l'ADN et des expériences de compétition en retardement de gel, utilisant des oligonucléotides sauvages ou mutants, ont montré qu'il existe un élément B dans ce promoteur ayant une séquence identique à l'élément B présent dans le promoteur du gène *col1a1*. Cet élément B du gène *col1a2* lie une protéine présente dans les extraits nucléaires de fibroblastes NIH-3T3 identique à celle se liant aux éléments A et B du promoteur du gène *col1a1*. Cet élément est donc le deuxième élément *cis*-après la boîte CCAAT-commun aux deux gènes.

## RÉFÉRENCES

17. Schneider-Maunoury S, Topilko P, Seitanidou T, Levi G, Cohen-Tannoudji M, Pournin S, Babinet C, Charnay P. Disruption of *krox-20* results in alteration of rhombomeres 3 and 5 in the developing hindbrain. *Cell* 1993; 75: 1199-214.
18. Roberts AB, Flanders KC, Kondaiah P, Thompson NL, Van Obberghen-Schilling EV, Wakefield L, Rossi P, De Crombrughe B, Heine U, Sporn M B. Transforming growth factor- $\beta$ : biochemistry and roles in embryogenesis, tissue repair and remodeling, and carcinogenesis. *Rec Prog Horm Res* 1988; 44: 157-97.
19. Roberts AB, Sporn MB, Assoian RK, Smith JM, Roche NS, Wakefield LM, Heine UI, Liotta LA, Falanga V, Kehrl JH, Fauci AS. Transforming growth factor type-beta: rapid induction of fibrosis and angiogenesis *in vivo* and stimulation of collagen formation *in vitro*. *Proc Natl Acad Sci USA* 1986; 83: 4167-71.
20. Penttinen RP, Kobayashi S, Bornstein P. Transforming growth factor  $\beta$  increases mRNA for matrix proteins in the presence and in the absence of changes in mRNA stability. *Proc Natl Acad Sci USA* 1988; 85: 1105-8.
21. Inagaki Y, Truter S, Ramirez F. TGF- $\beta$  stimulates  $\alpha 2(I)$  collagen gene expression through a *cis*-acting element that contains an Sp1 binding site. *J Biol Chem* 1994; 269: 14828-34.
22. Solis-Herruzzo JA, Brenner DA, Chojrier M. Tumor necrosis factor- $\alpha$  inhibits collagen gene transcription and collagen synthesis in cultured human fibroblasts. *J Biol Chem* 1988; 263: 5841-5.
23. Mauviel A, Heino J, Kähäri V-M, Hartmann DJ, Loyau G, Pujol JP, Vuorio E. Comparative effects of interleukin-1 and tumor necrosis factor- $\alpha$  on collagen production and corresponding procollagen messenger-RNA levels in human dermal fibroblasts. *J Invest Dermatol* 1991; 96: 243-9.
24. Schmidt A, Setoyama C, De Crombrughe B. Regulation of a collagen gene promoter by the product of viral *mos* oncogene. *Nature* 1985; 314: 286-9.

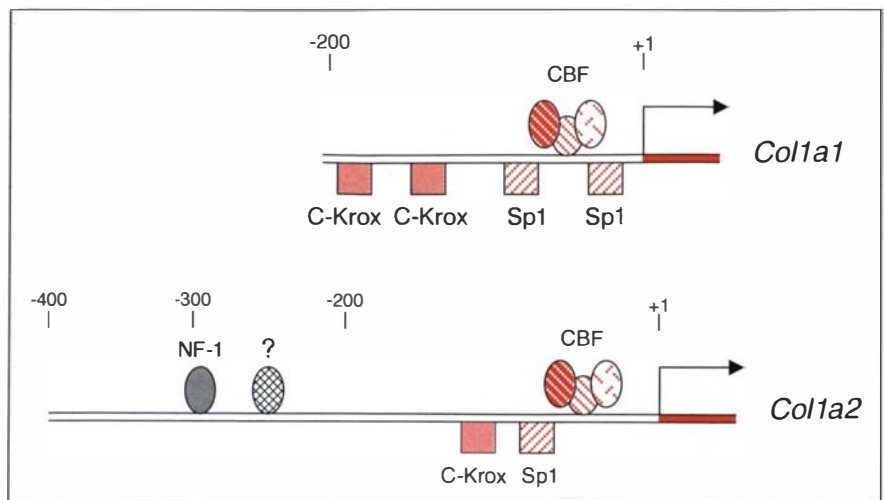


Figure 3. Représentation schématique des éléments cis liant des facteurs de transcription identifiés dans les promoteurs des gènes *col1a1* et *col1a2*. La figure du haut représente le promoteur du gène *col1a1* avec les sites de liaison pour CBF (CCAAT binding factor), Sp1 et C-Krox. La figure du bas représente le promoteur du gène *col1a2* qui inclut les sites de liaison pour CBF, Sp1, C-Krox et NF-1.

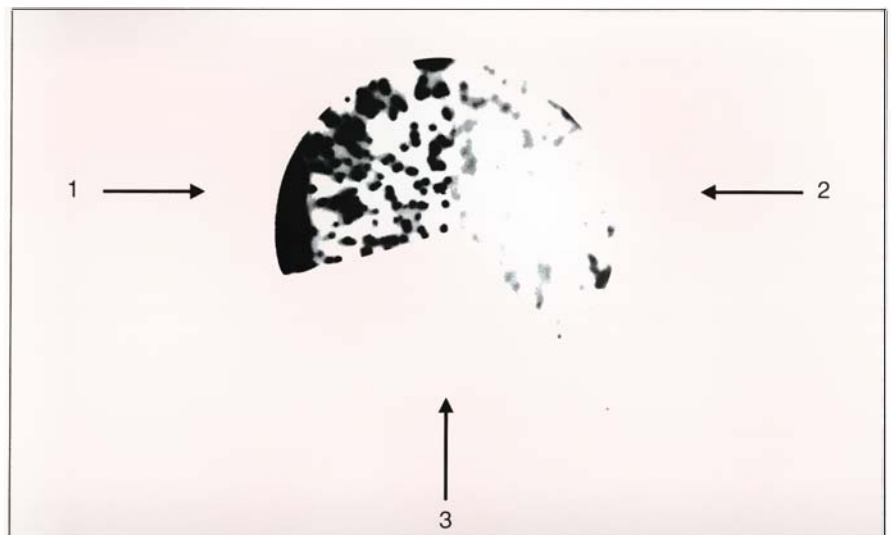


Figure 4. Détection *in situ* du clone d'ADNc C-Krox. Un phage contenant un ADNc codant pour une protéine qui se lie à l'élément  $\alpha 1(I)$  a été isolé dans une banque d'expression. Le filtre obtenu après le troisième criblage et contenant la protéine codée par ce clone d'ADNc a été coupé en trois morceaux qui ont été hybridés avec des sondes représentant un tétramère de l'oligonucléotide  $\alpha 1(I)$  [1], ou un monomère de l'oligonucléotide  $\alpha 1(I)$  [2], ou un tétramère de l'oligonucléotide  $\alpha 1(I)$  mutant [3].

	Identité %	Similitude %
CPVCHKIIHGAGKLP RHMRTHTGKPFACEVCGVRFTRNDK LKIHM RKHTGERPYS CPHCPARFLHSYDLKNHMLH		
--E-D-R-TRDHH-KT---W-----H-SH--RQSVQVAN--R-W-V-----T-EI-D---SD-N---S--L--	45	56
--AEG-D-R-SRS-E-T---I---H---Q-R--MRN-S-S-H-TT---T-----A-DY-GR--A--D-R-R-TK--	42	56
---ES-D-R-SRS-E-T---I-----Q-R--MRN-S-S-H-TT---T-----A-DI-GR--A--D-R-R-TK--	44	59
-DE-G-T--QS-S-L--Q-I-----T---K--IERSS-T--Q-T-----K-HE-GKA-S--M--TV--RT-	41	59

Figure 5. **Comparaison de la séquence en acides aminés de la région des doigts de zinc de C-Krox et de certains membres de la famille Krox.** Les pourcentages d'identité et de similitude au niveau de la séquence en acides aminés entre les divers membres de la famille Krox sont indiqués à droite. Les traits représentent les acides aminés qui sont identiques à C-Krox. Les résidus histidines et cystéines sont indiqués en rouge.

D'autres éléments régulateurs *cis*, qui ne sont pas présents dans le promoteur court du gène *colla1*, ont été identifiés en amont du site d'initiation de la transcription du promoteur *colla2*. Un site de liaison situé à -250 lie un facteur de transcription qui n'a pas encore été caractérisé. Le site localisé à -300 (-315 à -284) fixe un activateur de la transcription appelé CTF/NF1 [13]. Chacun de ces deux éléments agit comme activateur de transcription. Il est maintenant possible de représenter les éléments *cis* communs aux deux promoteurs susceptibles d'intervenir dans la régulation de l'expression des deux gènes du collagène de type I (figure 3).

### Spécificité de l'expression tissulaire des gènes du collagène de type I

Comme nous l'avons mentionné dans l'introduction, le collagène de type I est principalement synthétisé dans deux types cellulaires qui sont les ostéoblastes et les fibroblastes. L'étude systématique des séquences régulatrices requises pour établir cette spécificité d'expression a été effectuée dans divers laboratoires. Les résultats obtenus peuvent être résumés de la façon suivante. (1) Un fragment relativement court (300 paires de bases) du promoteur du gène *colla2* est nécessaire et suffisant pour obtenir l'expression d'un gène rapporteur dans les fibroblastes et les ostéoblastes de souris transgéniques [14]. Un fragment également court du gène *colla1* (220 paires de bases) semble contenir les éléments nécessaires pour diriger

l'expression d'un gène rapporteur dans des fibroblastes en culture. (2) Les éléments *cis* responsables de l'expression ostéoblastique du gène *colla1* semblent être présents plus en amont, entre -1,7 kb et -2,3 kb. Ainsi, il apparaît que les éléments *cis* responsables de l'expression spécifique des deux gènes du collagène de type I dans les fibroblastes sont présents dans un fragment court de leurs promoteurs. La détermination de la nature et de la localisation des éléments *cis* présents dans ces courts promoteurs a révélé qu'au moins deux de ces éléments lient des facteurs qui sont exprimés de façon ubiquitaire : CBF qui se lie à la séquence CCAAT et Sp1 qui se lie à l'élément C. Nous nous sommes initialement intéressés aux éléments A et B car ils pouvaient permettre d'expliquer la co-régulation de l'expression de ces gènes. De façon inattendue, cette étude nous a permis de caractériser un facteur de transcription exprimé principalement dans les fibroblastes.

### Les éléments A et B présents dans les promoteurs des deux gènes du collagène de type I lient un nouveau facteur de transcription synthétisé de façon prédominante dans la peau

Afin de caractériser le facteur de transcription se liant aux éléments A et B, nous avons criblé une banque d'expression de fibroblastes NIH-3T3 suivant la technique dite de *Southwestern*, en utilisant comme sonde un fragment d'ADN contenant quatre copies de l'élément A sauvage du gène *colla1*. Les clones positifs ont été en-

suite criblés à l'aide d'une sonde contenant quatre copies du même élément, mais cette fois-ci, muté. Cette approche nous a permis de cloner l'ADNc correspondant à une protéine se liant spécifiquement à l'élément A sauvage, mais non à l'élément A muté (figure 4) [15]. La séquence en acides aminés de cette protéine, déduite à partir de la séquence de l'ADNc, a révélé la présence de trois doigts de zinc du type C<sub>2</sub>H<sub>2</sub> près de l'extrémité C-terminale. Dans la région des doigts de zinc, cette protéine présente 55 à 60 % d'homologie avec la protéine Krüppel, un facteur de transcription de *Drosophila melanogaster*, mais aussi avec certains homologues murins de Krüppel, les protéines Krox 1, Krox 2, Krox 20 et Krox 24 (figure 5). Les protéines Krox ont toutes en commun l'existence de plusieurs doigts de zinc, ce qui suggère un possible rôle comme facteurs de transcription. Cependant, jusqu'à présent, l'existence d'un gène cible n'a pu être démontrée que pour un seul membre de cette famille, Krox 20. Krox 20, qui est exprimé de façon exclusive dans les rhombomères 3 et 5, règle l'expression du gène *Hox B2*. La destruction du gène *Krox 20* chez la souris est marquée par un développement anormal du cerveau postérieur [16, 17].

Cette nouvelle protéine Krox, que nous avons appelée C-Krox (collagène-Krox), présente plusieurs particularités intéressantes dans le cadre de l'étude des mécanismes moléculaires de la fibrose. Cette protéine se lie de façon spécifique aux éléments A et B du promoteur du gène *colla1*, mais également à l'élément B du promo-

teur *colla2*, constituant ainsi le second membre de la famille Krox pour lequel un gène cible a été identifié. C-Krox, dont l'expression est réglée par les facteurs de croissance présents dans le sérum, accroît l'activité transcriptionnelle des promoteurs *colla1* et *colla2*. Le résultat probablement le plus intéressant concernant l'étude moléculaire de la fibrose est que le gène *C-Krox* est bien plus abondamment exprimé dans la peau, et plus spécifiquement dans les fibroblastes dermiques, que dans n'importe quel autre tissu, alors qu'il n'est pas exprimé dans les os, l'autre source majeure de collagène de type I (figure 6). Le fait que *C-Krox* soit préférentiellement exprimé dans les fibroblastes est en accord avec les résultats expérimentaux montrant qu'un promoteur *colla1* court est exprimé spécifiquement dans les fibroblastes *in vitro*. L'absence d'expression de ce gène dans l'os suggère que les mécanismes réglant l'expression des gènes du collagène de type I dans l'ostéoblaste et le fibroblaste sont différents. Notre hypothèse est que C-Krox pourrait jouer un rôle important dans l'expression fibroblastique des gènes du collagène de type I dans les situations physiologiques et pathologiques. Plusieurs expériences doivent être effectuées pour confirmer cette hypothèse. En particulier, l'étude du profil d'expression du gène *C-Krox* durant le développement embryonnaire murin fournira des informations importantes sur la fonction possible de ce nouveau facteur de transcription. A plus long terme, l'étude *in vivo* de la fonction de C-Krox permettra de déterminer son rôle durant le développement et dans certaines maladies.

### Régulation de l'expression des gènes du collagène de type I par les cytokines

L'expression des gènes du collagène de type I est réglée par de nombreuses cytokines ainsi que par certains oncogènes. Il est probable qu'une synthèse anormale de certaines de ces cytokines dans les maladies fibrotiques soit à l'origine de l'augmentation de la synthèse du collagène de type I comme d'autres protéines de la matrice extracellulaire.

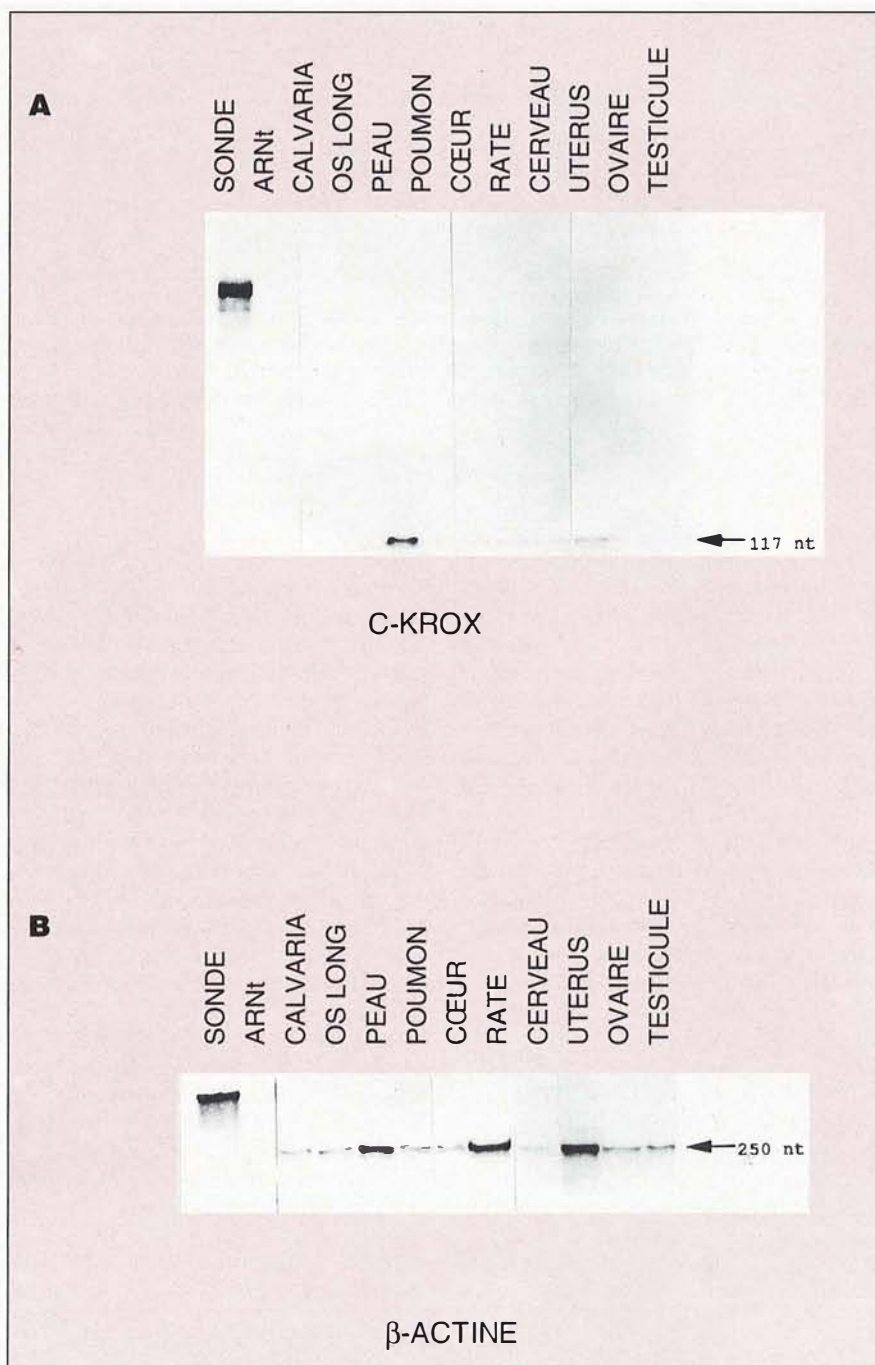


Figure 6. Analyse de l'expression du gène C-Krox chez la souris par la méthode de protection à la RNase. A. 10 µg d'ARN totaux extraits à partir de diverses sources tissulaires de souris ont été hybridés avec une sonde C-Krox localisée dans la région 5' non traduite ainsi qu'une partie de la région codante de l'ADNc de C-Krox. La sonde d'ARN antisens non digérée a une taille de 160 nucléotides alors que la taille du fragment protégé est de 117 nucléotides (indiqué par une flèche). L'ARNt de levure a été utilisé comme contrôle négatif. B. 1 µg d'ARN total collecté à partir de divers tissus de souris a été hybridé avec une sonde témoin β-actine. La sonde non digérée a une taille de 300 nucléotides alors que le fragment protégé est de 250 nucléotides (indiqué par une flèche).



Le *transforming growth factor-β* (TGFβ) est un facteur de croissance multifonctionnel qui stimule notamment la synthèse du collagène de type I et de la fibronectine dans les fibroblastes en culture [13, 18]. Cet effet peut également être reproduit chez l'animal : des injections de TGFβ à des souris nouveau-nées provoquent l'apparition d'un granulome, composé essentiellement de fibroblastes et de protéines de la matrice extracellulaire [19]. D'autre part, le TGFβ agit au niveau de la dégradation de la matrice extracellulaire ; il diminue la synthèse de deux métalloprotéinases, la collagénase et la transine/stromélysine. Le TGFβ stimule également la sécrétion des inhibiteurs de protéases et règle donc l'activité des protéases ; ainsi la sécrétion de l'inhibiteur de l'activateur du plasminogène et du *tissue inhibitor of metalloproteases* (TIMP) sont élevées sous l'effet de ce facteur [18]. La résultante de ces multiples effets du TGFβ sur les fibroblastes est de favoriser l'augmentation de la synthèse des protéines de la matrice extracellulaire. Du fait de ses multiples effets sur le fibroblaste, le TGFβ joue probablement un rôle crucial dans le processus fibrotique.

Les mécanismes moléculaires de l'action du TGFβ sur les fibroblastes ont été étudiés dans plusieurs laboratoires. Après interaction au niveau de récepteurs membranaires spécifiques ayant une activité sérine/thréonine kinase, le TGFβ entraîne une cascade d'événements encore mal connus qui conduisent à une augmentation de l'expression des gènes codant pour des protéines de la matrice extracellulaire. Le TGFβ agit sur différentes étapes de la biosynthèse collagénique et accélérerait notamment la transcription, la maturation et la sécrétion des collagènes nouvellement synthétisés. Parallèlement, il augmenterait aussi la stabilité des ARNm correspondants [20]. Le TGFβ stimule, notamment, l'activité des promoteurs des deux gènes *colla1* et *colla2*. Une séquence située entre -350 et -290 dans le promoteur *colla2* de souris et de l'homme semblerait être la séquence cible du TGFβ. Des expériences de liaison de protéines à l'ADN et de mutagenèse dirigée ont indiqué que, chez la souris, le site de liaison CTF/NF-1 était responsable de l'activation transcriptionnelle du gène

*colla2* par le TGFβ1 [13]. Une étude sur le promoteur du gène *colla2* humain a révélé la présence d'un TGFβ *responsive element* (TβRE), localisé entre -378 et -183, qui est responsable de l'activation de la transcription du gène *colla2* par le TGF-β [21]. Les conclusions de cette étude montrent que l'effet stimulant du TGFβ sur l'activité transcriptionnelle du gène *colla2* est dû à l'augmentation de la liaison d'un complexe de protéines appelé TβRE *protein complex* (TβRC) qui se lie au TβRE, et que cet effet serait relayé par Sp1. L'augmentation de l'activité transcriptionnelle du promoteur *colla1* induite par le TGFβ nécessite la présence d'un élément *cis* situé plus en amont, à -1 627 paires de bases. La nature du facteur de transcription se liant à cet élément n'a pas encore été déterminée.

La cachectine ou *tumor necrosis factor-α* (TNFα) est un facteur polypeptidique également produit par les macrophages. Il provoque une diminution de la transcription des gènes du collagène de type I : la synthèse de la protéine correspondante chez des fibroblastes humains et des cellules 3T3 en présence de sérum est réduite [22, 23]. Les mécanismes moléculaires d'action du TNFα sur l'activité transcriptionnelle des gènes du collagène de type I sont actuellement en cours d'étude. Des résultats préliminaires indiquent que le TNF-α induit une diminution de l'activité transcriptionnelle du promoteur du gène *colla2* humain, et que cet effet implique la même région située entre -300 et -290 nécessaire à l'effet activateur du TGFβ (Y. Inagaki, F. Ramirez ; communication personnelle). Finalement, le produit de certains oncogènes tels que *v-src*, *v-mos* et *v-ras* inhibe le programme de différenciation des fibroblastes et la production de nombreuses protéines de la matrice extracellulaire telles que le collagène de type I [24]. Cette diminution de l'expression des gènes du collagène de type I est un des phénomènes initiaux lors de la transformation oncogénique des fibroblastes. D'autres cytokines, telles l'interféron-γ et la leucoréguline, règlent également l'expression des gènes du collagène de type I. Cependant, à l'exception du TGFβ et probablement du TNFα, le

(ou les) éléments *cis* responsables de la régulation par ces différentes cytokines n'ont pas été caractérisés.

## Conclusion

L'étude détaillée des éléments *cis* présents dans les gènes du collagène de type I, ainsi que l'analyse des facteurs de transcription qui interagissent avec eux, ont déjà permis de mieux comprendre les mécanismes moléculaires qui contrôlent l'activité de ces gènes dans des situations pathologiques. Il apparaît clairement que cette connaissance est encore fragmentaire et que ces études doivent être poursuivies, notamment pour rechercher de nouveaux facteurs de transcription ayant une spécificité tissulaire d'expression. Compte tenu de la taille et de la complexité de l'organisation des gènes *colla1* et *colla2*, il est probable que d'autres facteurs de transcription, jouant un rôle important dans l'expression osseuse et/ou dermique de ces gènes, restent à découvrir.

L'étude de C-Krox et de son possible rôle dans la physiopathologie de la sclérodémie devra être poursuivie et cela dans plusieurs directions. D'une part, nous devons étudier son profil d'expression au cours du développement embryonnaire de la souris afin de savoir si C-Krox commence à être exprimé avant les gènes du collagène de type I et peut jouer un rôle dans la différenciation du fibroblaste. Le rôle fonctionnel de C-Krox doit être appréhendé dans des modèles de souris transgéniques contenant dans leur génome plusieurs copies du site de liaison de C-Krox, par l'étude de l'expression de C-Krox dans des modèles animaux de fibrose et de sclérodémie et, enfin, grâce à la production de souris transgéniques dépourvues de C-Krox. Enfin, l'expression du gène codant pour C-Krox étant réglée par des facteurs de croissance présents dans le sérum, il sera important de déterminer si certaines de ces cytokines ou facteurs de croissance connus pour régler l'expression des gènes du collagène de type I affectent également l'expression du gène codant pour C-Krox.

Ces diverses études fourniront également des informations qui nous permettront de mieux comprendre comment les différentes cytokines, qui interagissent avec des récepteurs spé-

cifiques membranaires, peuvent affecter l'activité transcriptionnelle des gènes du collagène de type I. Nous pensons que la diversité des facteurs de transcription qui interagissent au niveau des éléments régulateurs des gènes *coll1a1* et *coll1a2* est le reflet de la multiplicité d'action engendrée par les diverses cytokines connues pour influencer l'activité de ces gènes. Nous savons d'ores et déjà que les cytokines interagissent avec des récepteurs membranaires spécifiques et qu'après activation de ces derniers, le message peut être transmis par diverses voies de transduction qui peuvent activer ou inhiber l'expression des gènes du collagène de type I. Nous pouvons donc émettre l'hypothèse selon laquelle, dans les fibroblastes sclérodermiques, une ou plusieurs de ces voies de transduction intracellulaires peuvent devenir plus actives que dans des fibroblastes normaux, ceci ayant pour cause une stimulation chronique induite par une ou plusieurs cytokines ■

## Summary

### Regulation of the expression of type I collagen genes : implications for the study of Scleroderma

Type I collagen, the most abundant protein of the body, is preferentially synthesized in bone, dermis, and tendons by two cell types, the osteoblast and the fibroblast. The expression of type I collagen is increased in the various forms of fibrosis such as lung, liver, bone marrow fibrosis and scleroderma. Type I collagen is a heterotrimer molecule consisting of two  $\alpha 1(I)$  chains and one  $\alpha 2(I)$  chain. The two polypeptide chains are synthesized in a 2:1 stoichiometry. The same 2:1 ratio is observed for the rate of synthesis of the corresponding mRNAs. One hypothesis that would explain how this coregulation occurs at the transcriptional level is that common cis-acting elements are present on both genes. These common regulatory elements would bind identical transcription factors displaying the same function. The characterization of the various regulatory elements present in these genes would foster our understanding of the molecular mechanisms control-

ling type I collagen genes expression in normal and in pathological situations. Over the past few years, several laboratories have identified cis-acting elements in the promoters of the *coll1a1* and *coll1a2* genes. At least, two of these cis-acting elements are common to both promoters. One is centered by a pentanucleotide CCAAT and binds a ubiquitously expressed heteromeric CCAAT binding factor. A second one is centered by a G-rich region and it binds a new transcription factor called C-Krox. Interestingly, *C-Krox* gene, whose expression is regulated by growth factors, is preferentially expressed in skin fibroblasts in mice and is absent in bone suggesting that it plays a role in the fibroblast specific expression of type I collagen genes. The knowledge of how this and eventually other transcription factors act to regulate collagen expression will eventually lead to a better understanding of the increased type I collagen gene expression seen in diseases such as scleroderma.