

Transcriptional control mechanisms for the expression of type I collagen genes

Benoit de Crombrughe, Tuula Vuorio, Gerard Karsenty, Sankar Maity, E Cristy Rutheshouser, Howard Goldberg

Fibrosis is a hallmark symptom in a number of human diseases, including scleroderma, lung fibrosis, liver cirrhosis, atherosclerosis, osteoarthritis, and keloids. The most prominent biochemical manifestation of fibrotic lesions is an abnormal accumulation of extracellular matrix components, including type I and III collagen in mesenchymal tissues. This abnormal accumulation often results in severe malfunction of the affected tissues.

It is important to note that the formation of fibrous tissue can also be a normal physiological response as it occurs, for instance, in wound healing. This suggests that in fibrotic diseases regulation of this normal physiological process is altered. A study of the mechanisms regulating the normal response should help in understanding the abnormal control of this process.

A model for the fibrotic response

Several possible mechanisms might have a role in the exaggerated accumulation of type I collagen and other extracellular matrix components in fibrotic lesions. Firstly, recruitment and proliferation of fibroblasts would be expected to lead to an increased accumulation of extracellular matrix components. Secondly, decreased secretion of collagenases and other extracellular proteases or increased secretion of inhibitors for these proteases would have an analogous effect. Finally, stimulation of the synthesis of type I collagen and other extracellular matrix components would also produce the same phenotype. All three mechanisms presumably have a role. This article considers the third alternative, stimulation of the synthesis of extracellular matrix molecules, which is likely to be an important aspect of the fibrotic response.

In many fibrotic diseases the nature of the initial aetiological agent or injury is probably an autoimmune process. The nature of this immune reaction and the nature of the triggering mechanism itself that induces the fibrotic response by the affected cells are still poorly understood. It is, nonetheless, reasonable to postulate that cytokines (such as transforming growth factor β , interleukin 1, platelet derived growth factor, fibroblast growth factor) and other factors might play an important part in stimulating cells to overproduce extracellular matrix components. By binding to specific receptors on the surface of fibroblasts these cytokines would stimulate intracellular signalling pathways, which in turn would lead to the activation of a pleiotropic genetic programme characteristic of the fibrotic response. Activation of this genetic programme would result in

changes in the expression of specific genes causing the fibrotic phenotype displayed by these cells. The changes in gene expression include stimulation of expression of the genes for type I and III collagen and genes for a number of other extracellular matrix proteins, but could eventually also include decreased synthesis of collagenases and other extracellular proteases as well as increased synthesis of inhibitors for these enzymes. It is likely that in many instances of fibrosis a genetic programme is activated that is also activated in fibroblasts under normal physiological conditions but that the control of this programme becomes unchecked. Figure 1 presents a schematic illustration of some of these proposed regulatory steps which control the synthesis of type I collagen.

The various components of this hypothetical model are still poorly defined. The exact nature of the cytokines that would induce the fibrotic response in various diseases is largely unknown. So is the nature of the probable immunological injury itself that would cause the increased concentrations of cytokines. Moreover, essentially nothing is known about the intracellular signalling pathways which would be triggered by the cytokines and would activate the genetic programme characteristic of this fibrotic response in the cell.

We consider that regardless of the primary cause of fibrosis a better understanding of the regulation of the synthesis of extracellular matrix components, such as type I and III collagen, in normal fibroblastic cells should lead to a better comprehension of the nature of the pleiotropic genetic programme that is activated in fibrosis. Because so little is known about the causes of these diseases we studied one of their symptoms (synthesis of collagen) with the expectation that a careful analysis of the factors that regulate the expression of the genes for type I and III collagen would lead to a better understanding of the intracellular signalling pathways controlling the expression of these genes and eventually provide useful information about the mechanisms that trigger the activation of these pathways. Scleroderma fibroblasts show an increased rate of type I collagen RNA synthesis. Hence our studies focused on the transcriptional control of the type I and III collagen genes in normal fibroblasts. It should be clear, however, that other control mechanisms that are active post-transcriptionally may also be important.

We discuss first the evidence showing increased concentrations of type I collagen RNA in fibroblasts from patients with scleroderma, as well as the evidence showing an increased rate of synthesis of these RNAs. We

Department of
Molecular Genetics,
University of Texas,
M D Anderson
Cancer Center,
1515 Holcombe
Boulevard,
Houston,
Texas 77030,
USA
B de Crombrughe
T Vuorio
G Karsenty
S Maity
C Rutheshouser
H Goldberg

Correspondence to:
Dr de Crombrughe.

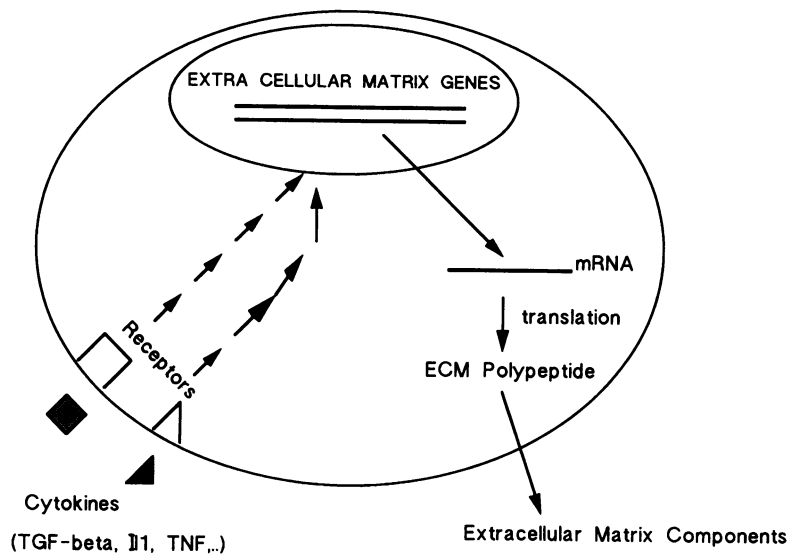


Figure 1 Schematic representation of a fibroblastic cell. The arrows pointing upwards represent mostly unknown intracellular signalling pathways, which presumably control the synthesis of extracellular matrix components. ECM=extracellular matrix; TGF=transforming growth factor; I1=interleukin 1; TNF=tumour necrosis factor.

then discuss how transcriptional control of the genes for type I collagen in normal fibroblastic cells is being studied and how such information should be useful for the understanding of the molecular pathology of fibrosis and scleroderma.

Increased levels of type I and III collagen RNA in scleroderma

Most of the studies carried out with fibroblasts cultured from scleroderma skin have indicated increased type I collagen production by these fibroblasts with no alteration in the ratio of $\alpha 1(I)$ to $\alpha 2(I)$ collagen.^{2,3} Type III collagen synthesis is also coordinately increased in scleroderma fibroblasts.³ This indicates an abnormal but coordinate regulation in the expression of at least three different collagen genes in this disorder. Scleroderma cells appear normal in their overall protein synthesis and growth rate, and they do not seem to produce increased amounts of collagenase. Not surprisingly, scleroderma cells lose their activated phenotype in culture after a few passages.^{2,3} Graves *et al* first showed that cultured scleroderma fibroblasts contained increased amounts of translatable type I collagen mRNAs.⁴ Further studies using nucleic acid hybridisation techniques have confirmed that scleroderma cells contain increased amounts of $\alpha 2(I)$, $\alpha 1(I)$, and $\alpha 1(III)$ collagen RNA.⁵⁻⁷ The increased mRNA levels for these collagens correlate closely with the amounts of collagen synthesised by these cells.

Experimental evidence strongly suggested that this increase in mRNA levels was mainly due to an increased transcription rate of the type I collagen genes. The rate of mRNA synthesis was measured by so called nuclear run-on experiments.⁸ This assay measures the number of RNA polymerase molecules (the enzyme which synthesises RNA from a DNA template) that have initiated transcription of a given gene and are in the process of transcriptional elonga-

tion at the time of isolation of the nuclei. The method relies on the fact that, in isolated nuclei, RNA in the process of being elongated can be labelled to high specific activity by incubating the isolated nuclei with radioactively labelled nucleoside triphosphates. The radioactive mRNAs synthesised by these isolated nuclei are then hybridised to filters, to which unlabelled recombinant plasmid DNAs containing sequences specific to only one species of mRNA have been fixed. The results indicate that nuclei from scleroderma fibroblasts synthesise more type I collagen RNA than nuclei from control fibroblasts.

In theory increased production of type I and III collagen in scleroderma fibroblasts might be due to activation of all fibroblasts or, alternatively, to a selective expansion or activation (or both) of a fibroblast subpopulation producing high concentrations of type I collagen. In situ hybridisation was used to try to distinguish between these possibilities. This method, which uses radioactive DNA probes to hybridise to RNAs directly in tissue sections, is helpful in determining which cells synthesise a given RNA. The method is useful to examine the possible heterogeneity of fibroblasts in fibrotic lesions and to obtain some estimate about the concentrations of a particular RNA present in individual cells. Radioactive DNA probes for type I collagen genes were used to localise corresponding mRNAs in fibroblasts from scleroderma skin and normal control skin. In scleroderma skin there are fibroblasts that produce larger quantities of type I collagen RNA and fibroblasts that show much lower concentrations.⁹ The amount of type I collagen mRNAs in the high producing cells are above the levels seen in any fibroblasts in normal skin. In cultured cells, on the other hand, the fibroblasts appeared to be homogeneously activated, each cell containing more type I collagen mRNAs than normal control cells.⁹ Hence the in situ hybridisations indicate that a certain proportion of fibroblasts are responsible for the higher rate of synthesis of type I collagen.

Control of type I collagen gene transcription in normal fibroblasts

Based on the previous results and regardless of the primary cause of scleroderma, it can be assumed that an increase in the rate of transcription of type I collagen genes, and probably of other genes for other components of the extracellular matrix, plays a major part in the appearance of the fibrotic phenotype. To understand better the precise mechanisms of this increase it is necessary to understand the regulation of these genes in normal cells. The second part of this article considers some of these control mechanisms in normal fibroblasts. The questions we are asking are as follows: Which are the sequences in the type I collagen genes which are cis acting elements and determine the transcriptional activity of these genes? What are the transacting protein factors that interact with these cis elements? Are any of these factors tissue specific? Are there sequences in type I collagen

genes that mediate tissue specific expression of the genes? Which sequences mediate increased or decreased expression in response to cytokines? A more precise knowledge of the location of the regulatory sequences in these genes, of their role, of the protein factors which interact with these sequences, and of the cellular mechanisms whereby these regulatory sequences might respond to various growth factors and cytokines should provide important new information that is essential for a better understanding of the mechanisms of the fibrotic process.

Type I collagen genes are large and the purified genes are fairly difficult to manipulate. We assumed that in these genes, as in most other genes, at least some of the major regulatory sequences that determine the rate of transcription of these genes are localised at their 5' end; our laboratory—as well as others—therefore focused its attention on the regulatory elements present at the 5' end of these genes. It is, of course, possible that other regulatory sequences may be present elsewhere in these genes.

To facilitate the study of the regulatory elements of the type I collagen genes we substituted most of the structural portion of these genes with a much smaller surrogate or marker gene. In these studies, as in the studies of other genes, one commonly used marker gene

is the bacterial chloramphenicol acetyltransferase (CAT) gene,¹¹ which is not endogenous to eukaryotic cells. Figure 2 shows a schematic representation of a prototypic chimeric gene, in which the mouse $\alpha 2(I)$ collagen 5' flanking sequences (2000 base pairs) are fused to the CAT gene. In a position 3' to the CAT gene is a segment of the first intron of the mouse $\alpha 2(I)$ collagen gene that is thought to contain a transcriptional enhancer element.¹¹

To study the function of the regulatory DNA segments, chimeric genes such as the one illustrated in fig 2 are introduced in tissue culture cells by a gene transfer method called DNA transfection. In transient expression experiments the cells are harvested 48 hours after DNA transfection, lysed, and assayed to determine the activity of the enzyme chloramphenicol acetyltransferase. This activity is directly proportional to the activity of the transcriptional regulatory elements that drive the expression of the CAT gene. Such experiments provide information about the activities of these regulatory elements (in our case those of the type I collagen genes) in specific tissue culture cells.

Regulatory elements of the $\alpha 2(I)$ collagen gene

We first wished to obtain evidence that the regulatory sequences of the $\alpha 2(I)$ collagen gene present in the chimeric construction of fig 2 were active in intact animals. This DNA construction was, therefore, microinjected in fertilised mouse oocytes and these microinjected oocytes reimplanted in the oviduct of pseudo-pregnant female mice to generate transgenic mice. The chimeric gene is integrated in the genome of a certain proportion of the pups born. This new gene is transmitted to the next generations in a characteristic Mendelian fashion. The pattern of expression of the CAT gene in various tissues of these mice was compared with the pattern of expression of the endogenous $\alpha 2(I)$ collagen gene. We found that these patterns were extremely similar—that is, high expression in tail, somewhat less in bone and in skin, and much less expression in other tissues (Goldberg H *et al*, unpublished data). In fact, with a construction containing only the 2000 base pairs upstream of the start of transcription of the $\alpha 2(I)$ collagen gene, a similar tissue specific pattern was obtained.¹² This experiment indicates that these 2000 base pairs are sufficient to confer tissue specific expression of the gene in intact animals. The results imply that the transacting factors which confer tissue specificity must interact with these 2000 base pairs upstream of the transcription start. These transgenic mice might also be a useful model system to examine whether the expression of the transgene is increased in experimental fibrosis.

We examined in more detail the 500 base pairs which lie immediately upstream of the start of transcription in the $\alpha 2(I)$ collagen gene (fig 3). Within these 500 base pairs there are several specific sites to which DNA binding factors present in nuclear extracts of NIH-3T3 fibroblasts have been shown to bind. A hetero-

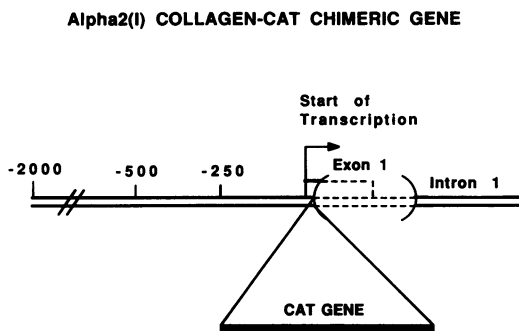


Figure 2 Schematic representation of a chimeric gene used in DNA transfection experiments. The chimeric gene consists of 2000 base pairs of the 5' flanking sequences of the $\alpha 2(I)$ collagen genes and a very small portion of the first exon of this gene fused to the gene for chloramphenicol acetyltransferase (CAT). A portion of the first intron of the $\alpha 2(I)$ collagen gene, which contains a transcriptional enhancer element, has been added to the construction. Most of the first exon and a portion of the first intron as well as all other segments of the $\alpha 2(I)$ collagen gene of 35 000 base pairs have been deleted.

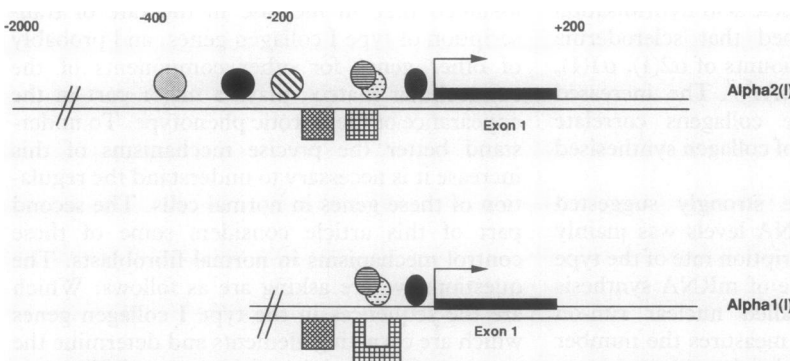


Figure 3 Schematic representation of some of the $\alpha 2(I)$ and $\alpha 1(I)$ collagen regulatory segments. Circles represent stimulatory transcription factors, whereas squares represent inhibitory factors. Both types of factor bind to specific sequences in these DNAs.

dimeric CCAAT binding factor binds to a CCAAT motif present around -80 .¹⁰ This is indicated by a symbol representing two overlapping circles in fig 3. Several single base pair mutations have been introduced in this site.¹³ They abolish or greatly decrease the binding of this CCAAT binding factor and also reduce five- to 10-fold the transcriptional activity of the $\alpha 2(I)$ collagen promoter in DNA transfection experiments. This strongly suggests that the factor that binds to this sequence is an activator of transcription for this gene. Binding to this DNA motif requires both subunits.¹⁰ This protein also binds to a similar site in the $\alpha 1(I)$ promoter. Furthermore, the purified factor stimulates the activity of both the $\alpha 2(I)$ and the $\alpha 1(I)$ collagen promoters in an in vitro reconstituted transcription system.¹⁴ The two subunits of the protein that bind to this site have been purified to homogeneity and cDNA clones have been isolated for each of these subunits.¹⁵ These reagents will help in studying changes in this transcription factor which may occur in both physiological and pathological situations.

Just upstream of the binding site for the CCAAT factor there are binding sites for two other factors first identified as binding to the $\alpha 1(I)$ promoter.¹⁶ They are indicated by squares in fig 3. Mutagenesis of these binding sites suggests that at least one of these factors acts as a transcriptional inhibitor of the $\alpha 2(I)$ promoter.

In addition to these three binding sites that are common to both the $\alpha 1(I)$ and $\alpha 2(I)$ collagen genes, three other binding sites have been identified in the mouse $\alpha 2(I)$ collagen promoter: at -250 , -300 , and -400 (indicated by circles in fig 3). The regulatory element located at -300 is a binding site for a DNA binding protein called nuclear factor 1.¹⁷ We found that activation of this promoter by transforming growth factor β is mediated by this binding site.¹⁸ Point mutations or small substitution mutations in the -250 and -300 binding sites, which greatly decrease the binding of specific factors, also generally inhibit the activity of the promoter in DNA transfection experiments between five- and 10-fold.¹³ These experiments suggest that the factors which bind to these elements act as transcriptional activators. The various transcriptional factors and their cognate cis acting binding sites in the $\alpha 2(I)$ collagen promoter segment are schematically represented in fig 3. The circles represent activators of transcription, whereas the squares represent inhibitors of transcription.

Regulatory elements of the $\alpha 1(I)$ collagen gene

Regulatory elements of the $\alpha 1(I)$ gene are distributed on either side of the start transcription. The first intron appears to contain both stimulatory and inhibitory segments but the precise delineation of these sequences has not yet been established.¹⁹⁻²¹ Several cis acting elements have been identified upstream of the transcription start. Our laboratory focused primarily on a short promoter fragment containing the 222 base pairs immediately upstream of the start of transcription, though additional

elements exist that are further upstream. When this fragment is fused to a CAT gene it constitutes a strong promoter in DNA transfection experiments of NIH-3T3 fibroblasts.¹⁶

Four different cis acting sequences which bind factors present in nuclear extracts of fibroblasts were identified in this short promoter using DNA binding experiments. The same heterodimeric factor which binds to a CCAAT motif in the $\alpha 2(I)$ collagen promoter also binds to the proximal of two such motifs in the $\alpha 1(I)$ promoter.^{14 16} A mutation in this CCAAT sequence which inhibits binding of this factor decreases the strength of this promoter four- to fivefold, indicating that this promoter acts as a transcriptional activator of the $\alpha 1(I)$ collagen promoter.¹⁶ This factor is represented by a double circle in the schematic model of the $\alpha 1(I)$ promoter of fig 3. Two other factors bind to this short promoter upstream of the CCAAT sequence. Both of these factors also bind to similar sequences and at similar distances from the start of transcription in the $\alpha 2(I)$ promoter.¹⁶ As shown in fig 3 one of these factors binds to two adjacent sites, whereas the other binds to two direct repeats that bracket the proximal CCAAT box. These factors are indicated by rectangular symbols in fig 3 and are shown below the double line representing the DNA. Analysis of mutant promoters with base substitutions in these binding sites suggests that both these factors act as inhibitors of transcription of the $\alpha 1(I)$ collagen promoter.¹⁶

It is interesting to note that both collagen genes contain a small cluster of common regulatory elements to which three different transcription factors bind. One is a transcriptional activator, whereas the two others are transcriptional inhibitors. It is likely that these factors participate in the coordinate control of these two genes which are coexpressed in almost all physiological and pathological situations.

Conclusions

One important conclusion of these studies is that a multiplicity of both positive and negative transcription factors interacts with regulatory sequences in the promoters of the $\alpha 1(I)$ and $\alpha 2(I)$ collagen genes. We believe that the complexity of these functional cis acting elements and their cognate transacting factors corresponds to a multiplicity of cytokines and other hormones, which by interacting with specific receptors on the surface of fibroblasts are known to influence the expression of these genes.

In summary, a detailed study of the regulatory elements in the type I collagen genes and an analysis of the transacting transcription factors which interact with these elements will lead to a significantly enhanced understanding of the molecular mechanisms which control the activity of these genes. These studies will also provide a better framework to understand how various cytokines that interact with specific receptors on the cell surface influence the activity of type I collagen genes. It is likely that there are also multiple intracellular signalling pathways which can be used to activate or repress these genes.

Possibly, in scleroderma fibroblasts, one or several of these intracellular pathways has become more active than in normal fibroblasts as a result of a long term and persistent stimulation, presumably by one or more cytokines. Furthermore, the protracted nature of the stimulatory input in scleroderma and other fibrotic diseases might be the consequence of an autocrine mechanism whereby the production of an active cytokine is continuously stimulated by the newly secreted cytokine itself.

- 1 Vuorio T, Maity S N, de Crombrughe B. Purification and molecular cloning of the "A" chain of a rat heteromeric CCAAT-binding protein. Sequence identity with the yeast Hap4 transcription factor. *J Biol Chem* 1990; 265: 22480-6.
- 2 LeRoy E C. Increased collagen synthesis by scleroderma skin fibroblasts in vitro. *J Clin Invest* 1974; 54: 880-9.
- 3 Vuorio T, Mäkelä J K, Kähäri V-M, Vuorio E. Coordinated regulation of type I and type III collagen production and mRNA levels of pro alpha1(I) and pro alpha2(I) collagen in cultured morphea fibroblasts. *Arch Dermatol Res* 1987; 279: 154-60.
- 4 Graves P N, Weiss I K, Perlish J S, Fleischmajer R. Increased procollagen mRNA levels in scleroderma skin fibroblasts. *J Invest Dermatol* 1983; 80: 130-2.
- 5 Jimenez S A, Feldman G, Bashey R I, Bienkowski R, Rosenbloom J. Co-ordinate increase in the expression of type I and type III collagen genes in progressive systemic sclerosis fibroblasts. *Biochem J* 1986; 237: 837-43.
- 6 Kähäri V-M, Vuorio T, Nääntö-Salonen K, et al. Increased type I collagen in RNA levels in cultured scleroderma fibroblasts. *Biochim Biophys Acta* 1984; 781: 183-6.
- 7 Vuorio T, Mäkelä J, Vuorio E. Activation of type I collagen genes in cultured scleroderma fibroblasts. *J Cell Biochem* 1985; 28: 105-13.
- 8 Kähäri V-M, Multimäki P, Vuorio E. Elevated pro alpha2(I) collagen mRNA levels in cultured scleroderma fibroblasts result from an increased transcription rate of the corresponding gene. *FEBS Lett* 1987; 215: 331-4.
- 9 Kähäri V-M, Sandberg M, Kalimo H, Vuorio T, Vuorio E. Identification of fibroblasts responsible for increased collagen production in localized scleroderma by in situ hybridization. *J Invest Dermatol* 1988; 90: 664-70.
- 10 Hatamochi A, Golumbek P T, 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.
- 11 Rossi P, de Crombrughe B. Identification of a cell-specific transcriptional enhancer in the first intron of the mouse alpha2 (type I) collagen gene. *Proc Natl Acad Sci USA* 1987; 84: 5590-4.
- 12 Khillan J S, Schmidt A, Oerbeek P, de Crombrughe B, Westphal H. Developmental and tissue specific expression directed by the alpha2 type I collagen promoter in transgenic mice. *Proc Natl Acad Sci USA* 1986; 83: 725-9.
- 13 Karsenty G, Golumbek P, de Crombrughe B. Point mutations and small substitution mutations in three different upstream elements inhibits the activity of the mouse alpha2 (I) collagen promoter. *J Biol Chem* 1988; 263: 13909-15.
- 14 Maity S N, Golumbek P T, Karsenty G. Selective activation of transcription by a novel CCAAT binding factor. *Science* 1988; 241: 582-5.
- 15 Maity S N, Vuorio T, de Crombrughe B. The B subunit of a rat heteromeric CCAAT-binding transcription factor shows a striking sequence identity with the yeast Hap2 transcription factor. *Proc Natl Acad Sci USA* 1990; 87: 5378-82.
- 16 Karsenty G, de Crombrughe B. Two different negative and positive regulatory factors interact with a short promoter segment of the alpha1 (I) collagen gene. *J Biol Chem* 1990; 265: 9934-42.
- 17 Oikarinen J, Hatamochi A, de Crombrughe B. Separate binding sites for nuclear factor I and a CCAAT DNA binding factor in the mouse alpha (I) collagen promoter. *J Biol Chem* 1987; 262: 11064-70.
- 18 Rossi P, Karsenty G, Roberts A B, Roche N S, Sporn M B, de Crombrughe B. A nuclear factor I binding site mediates the transcriptional activation of a type I collagen promoter by transforming growth factor-beta. *Cell* 1988; 52: 405-14.
- 19 Bornstein P, McKay J, Liska D S, Apne S, Devarayalu S. Interactions between the promoter and first intron are involved in transcriptional control of alpha1 (I) collagen gene expression. *Mol Cell Biol* 1988; 3: 4851-9.
- 20 Bornstein P, McKay J, Hoishima J K, Devarayalu S, Gelinis R E. Regulatory elements in the first intron contribute to transcriptional control of the human alpha1(I) collagen gene. *Proc Natl Acad Sci USA* 1987; 84: 8869-73.
- 21 Rossouw C M S, Vergeer W P, DuPlooy S J, Bernard M P, Ramirez F, de Wet W J. DNA sequences in the first intron of the human pro-alpha (I) collagen gene enhancer transcription. *J Biol Chem* 1987; 262: 15151-7.