The transcriptional tissue specificity of the human pro α 1(I) collagen gene is determined by a negative *cis*-regulatory element in the promoter

Carl P. SIMKEVICH, James P. THOMPSON, Helen POPPLETON and Rajendra RAGHOW*

Departments of Pharmacology and Medicine, University of Tennessee and the Veterans Affairs Medical Center, Memphis, TN 38104, U.S.A.

The transcriptional activity of plasmid pCOL-KT, in which human pro α 1(I) collagen gene upstream sequences up to –804 and most of the first intron (+474 to +1440) drive expression of the chloramphenicol acetyltransferase (CAT) gene [Thompson, Simkevich, Holness, Kang & Raghow (1991) *J. Biol. Chem.* **266**, 2549–2556], was tested in a number of mesenchymal and non-mesenchymal cells. We observed that pCOL-KT was readily expressed in fibroblasts of human (IMR-90 and HFL-1), murine (NIH 3T3) and avian (SL-29) origin and in a human rhabdomyosarcoma cell line (A204), but failed to be expressed in human erythroleukaemia (K562) and rat pheochromocytoma (PC12) cells, indicating that the regulatory elements required for appropriate tissue-specific expression of the human pro α 1(I) collagen gene were present in pCOL-KT. To delineate the nature of *cis*-acting sequences which determine the tissue specificity of pro α 1(I) collagen gene expression, functional consequences of deletions in the promoter and first intron of pCOL-KT were tested in various cell types by transient expression assays. *Cis* elements in the promoter-proximal and intronic sequences displayed either a positive or a negative influence depending on the cell type. Thus deletion of fragments using *EcoRV* (nt –625 to –442 deleted), *XbaI* (–804 to –331) or *SstII* (+670 to +1440) resulted in 2–10-fold decreased expression in A204 and HFL-1 cells. The negative influences of deletions in the promoter-proximal sequences was apparently considerably relieved by deleting sequences in the first intron, and the constructs containing the *EcoRV/SstII* or *XbaI/SstII* double deletions were expressed to a much greater extent than either of the single deletion constructs. In contrast, the *XbaI** deletion (nt –804 to –609), either alone or in combination with the intronic deletion, resulted in very high expression in all cells regardless of their collagen phenotype; the *XbaI**/(–*SstII*) construct, which contained the intronic *SstII* fragment (+670 to +1440) in the reverse orientation, was not expressed in either mesenchymal or non-mesenchymal cells. Based on these results, we conclude that orientation-dependent interactions between negatively acting 5'-upstream sequences and the first intron determine the mesenchymal cell specificity of human pro α 1(I) collagen gene transcription.

INTRODUCTION

An important structural constituent of connective tissues, type I collagen is a major secretory gene product of mesenchymal cells (Miller & Gay, 1987; Kuhn, 1987; Vuorio & de Crombrughe, 1990). Type I collagen is assembled as a heterotrimer consisting of two pro α 1(I) and one pro α 2(I) chains; the two α chains are products of unique genes located on two separate chromosomes (Miller & Gay, 1987; Kuhn, 1987; Vuorio & de Crombrughe, 1990). Collagen genes are controlled in both tissue- and development-specific modes (Miller & Gay, 1987; Kuhn, 1987; Vuorio & de Crombrughe, 1990), as well as in response to a number of biological and pharmacological stimuli (Raghow *et al.*, 1984, 1986, 1987a; Roberts *et al.*, 1986; Choe *et al.*, 1987; Rossi *et al.*, 1988; Penttinen *et al.*, 1988; Postlethwaite *et al.*, 1988; Solis-Herruzo *et al.*, 1988; Cockayne & Cutroneo, 1988; Goldstein *et al.*, 1990). Regulation of collagen genes in response to transforming growth factor β (Roberts *et al.*, 1986; Raghow *et al.*, 1987b; Rossi *et al.*, 1988; Penttinen *et al.*, 1988), fibrogenic factor (Raghow *et al.*, 1984; Choe *et al.*, 1987), interleukin 1 (Postlethwaite *et al.*, 1988), tumour necrosis factor α (Solis-Herruzo *et al.*, 1988), phorbol esters (Goldstein *et al.*, 1990) and glucocorticoids (Raghow *et al.*, 1986; Cockayne & Cutroneo, 1988) has been demonstrated previously. Although both transcriptional and post-transcriptional mechanisms are involved

in the regulation of collagen gene expression in response to such diverse stimuli, the predominant mode of control under most conditions appears to be transcriptional (Raghow & Thompson, 1989; Bornstein & Sage, 1989; Ramirez & DiLiberto, 1990).

A number of studies have shown that, in addition to the promoter-proximal sequences, DNA in the first intron modulates transcription of the human pro α 1(I) collagen gene (Rossouw *et al.*, 1987; Bornstein *et al.*, 1987, 1988; Bornstein & McKay, 1988). However, the precise organization of sequence motifs in the promoter and the regulatory elements of the first intron of pro α 1(I) remains incompletely defined. Existence of both positive and negative regulatory sequences in the intron has been proposed. For instance, Bornstein and co-workers reported that a 274 bp (nucleotides +820 to +1093) *AvaI* fragment (A274) from the first intron of the human pro α 1(I) collagen gene linked to a 430 bp fragment of the pro α 1(I) promoter exhibited orientation-specific inhibition of the reporter gene (Bornstein *et al.*, 1987). These authors also showed that nuclear extracts from HeLa, lymphoid and L cells recognized three sequence motifs, which correspond to bases 919–944, 951–978 and 986–1009 within the A274 fragment, suggesting that the *trans*-acting factors binding to A274 were present in both collagen-producing and non-producing cells. The complex array of additional regulatory sequences in the first intron was further suggested by the observation that the inhibitory effect of the A274 fragment could

Abbreviations used: DMEM, Dulbecco's modified Eagle's medium; FBS, fetal bovine serum; CAT, chloramphenicol acetyltransferase.

* To whom correspondence should be addressed: Veterans Affairs Medical Center, Research Service (151), 1030 Jefferson Avenue, Memphis, TN 38104, U.S.A.

be overridden by sequences surrounding it [e.g. 864 bp *Bgl*II/*Bam*HI fragment (+544 to +1407)].

Another important *cis*-regulatory sequence was identified recently by Liska *et al.* (1990), who showed that an Ap-1-like element was critically involved in the orientation-dependent enhancing ability of the intronic sequence. These investigators also provided evidence that the factor(s) which bind to the intronic Ap-1 site of the *pro α 1(I)* gene were present only in the cells of mesenchymal origin (Liska *et al.*, 1990). However, analyses of the Ap-1-like element of the human *pro α 1(I)* gene in our laboratory revealed results that differed from the data of Liska *et al.* (1990) in two important ways. First, the deletion of the Ap-1-like element resulted in much greater expression of the plasmid pCOL-KT, suggesting that Ap-1 acted as a negative regulatory element. Second, the nuclear extracts from both collagen-producing and non-producing cells contained *trans*-acting factor(s) which were recognized by the Ap-1-like element (H. Katai, J. D. Stephenson, D. P. Simkevich, J. P. Thompson & R. Raghov, unpublished work). In another recent study, a minigene version of the human *pro α 1(I)* gene containing 2.3 kb of the 5'-flanking sequence, with or without most of the first intron, was found to be expressed abundantly in stably transfected NIH 3T3 cells (Olsen *et al.*, 1991). Although that study raised important questions about the contribution of intronic sequences in the regulation of the human *pro α 1(I)* gene, it did not address the question of tissue specificity. Therefore, based on the published findings, we surmise that the questions regarding (1) the precise identity of determinants that confer tissue-specific expression to the human *pro α 1(I)* gene, and (2) the contribution of the first intron to such transcriptional specificity, remain unanswered. A cursory comparison between the human and murine *pro α 1(I)* collagen genes further complicates this situation, and suggests that there may be unique species-specific differences in the regulatory mechanisms of *pro α 1(I)* gene expression (Rippe *et al.*, 1989; Brenner *et al.*, 1989). It was reported that, in contrast to what has been demonstrated for the human *pro α 1(I)* gene, only 220 bp upstream of the transcription start site of the murine collagen *pro α 1(I)* gene, containing two identical 12 bp repeats, were sufficient for maximal tissue-specific expression of the reporter gene. It is noteworthy, however, that even though the 12 bp sequences resembled the recognition sites of Sp-1 and Ap-2 nuclear factors, the binding specificities of these sites were different from those of the authentic Sp-1 and Ap-2 proteins (Rippe *et al.*, 1989; Brenner *et al.*, 1989).

With the primary aim of defining the *cis*-acting sequences involved in tissue-specific transcriptional regulation, we created constructs in which expression of a reporter gene was driven by a sequence 2200 bp upstream of the transcription start site plus most of the first intron of the human *pro α 1(I)* gene. Extensive deletion analysis of the promoter-proximal and the intronic sequences shows that: (i) 804 bp of upstream sequence was sufficient for optimal tissue-specific expression, and (ii) the determinants of mesenchymal cell specificity of *pro α 1(I)* gene transcription consist of a negative regulatory element(s), removal of which leads to promiscuous expression of the *pro α 1(I)* gene in all cell types, regardless of their lineage.

MATERIALS AND METHODS

Materials

Restriction endonucleases and other nucleic acid-modifying enzymes (purchased from Bethesda Research Laboratories, Gaithersburg, MD, U.S.A., and International Biotechnologies, New Haven, CT, U.S.A.) were used according to the manufacturers' specifications. [¹⁴C]-labelled chloramphenicol and acetyl-CoA were purchased from New England Nuclear,

Cambridge, MA, U.S.A., and Sigma, St. Louis, MO, U.S.A., respectively. Plasmids containing various DNA fragments were transfected into *Escherichia coli* strain JM 83, and bacteria were grown in LB broth under appropriate antibiotic selection conditions. Silica gel t.l.c. plates were obtained from J. T. Baker Inc., and were spray-coated with ENhance (New England Nuclear, Boston, MA, U.S.A.) and autoradiographed.

Cells

Continuous lines of human fibroblasts (HFL-1 and IMR-90), and a human erythroleukaemia cell line (K562) were obtained from American Tissue Culture Collection (ATCC). A rhabdomyosarcoma cell line (A204) was obtained from Dr. Helene Sage, University of Washington, Seattle, WA, U.S.A. Embryonic chicken fibroblasts (SL29) were purchased from ATCC, and a rat pheochromocytoma cell line (PC12) was obtained from Helen Fillmore, U.T. Memphis. HFL-1 and A204 cells were grown in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum (FBS) and antibiotics (100 units of penicillin/ml, 100 µg of streptomycin/ml and 25 µg of fungizone/ml; Gibco). K562 cells were grown in 50% DMEM/38% RPMI/10% FBS antibiotics/1% glutamine. SL29 cells were cultivated in DMEM containing 10% FCS and antibiotics, and PC12 cells were grown in RPMI 1640 supplemented with 10% FCS, 5% horse serum and antibiotics.

Plasmid construction

All recombinant DNA manipulations were carried out by standard protocols (Sambrook *et al.*, 1989). The plasmid pCOL-KT was constructed by isolating a 2.3 kb *Hind*III fragment of a genomic clone (Barsh *et al.*, 1984) of human *pro α 1(I)* and inserting it into the *Hind*III recognition sequence of the multiple cloning site of plasmid pKT (Reue *et al.*, 1988). The translational initiation site and first exon (spanning nucleotides +68 to +478) of *pro α 1(I)* were deleted as described (Thompson *et al.*, 1991). Digestion of the 5 kb *Bam*HI fragment of the genomic *pro α 1(I)* DNA (Barsh *et al.*, 1984) with *Sph*I liberated a 1500 bp fragment representing additional 1197 bp of the 5'-upstream sequences. The *Sph*I fragment was ligated into *Sph*I-digested pCOL-KT; recombinant clones with the correct orientation of the additional sequences, designated [pCOL-KT(L)], were grown and used for transfection studies.

Creation of deletion mutations

Deletions of pCOL-KT were made using several different restriction endonucleases. pCOL-KT was digested with *Eco*RV to remove the fragment spanning nucleotides -625 to -442, and blunt-end-ligated with T4 DNA ligase to recircularize the plasmid. Similarly, digestion of pCOL-KT with *Xba*I deleted bases -804 to -331. Double digestions of pCOL-KT with *Xma*III and *Sph*I were used to remove bases -477 to -83; the ends were filled with Klenow enzyme and blunt-end-ligated. The putative regulatory sequences in the first intron were removed by digestion with *Sst*II, which removed a fragment encompassing nucleotides +670 to +1440, and recircularized. A number of double deletions were also constructed. pCOL-KT containing the *Eco*RV deletion was further digested with *Sst*II to remove bases +670 to +1440, to create *Eco*RV/*Sst*II (-625 to -442 and +670 to +1440 deleted). Similarly, pCOL-KT containing the *Xba*I deletion was digested with *Sst*II and recircularized to create *Xba*I/*Sst*II (-804 to -331 and +670 to +1440 deleted). Partial digestion of pCOL-KT containing the *Sst*II deletion with *Xba*I followed by relegation with T4 DNA ligase was done to create *Xba*I*/*Sst*II (-804 to -609 and +670 to +1440 deleted). Precise deletion points were confirmed by diagnostic enzyme fragmentation of the deleted constructs. We created *Xba*I* and

*Xba*I*/(–*Sst*II) by religating the *Sst*II fragment in both orientations in the *Xba*I* construct after *Sst*II digestion.

Cell transfections and transient expression assays

HFL-1, A204 and SL29 cells were grown as monolayers to about 70% confluency and transfected with pCOL-KT and its derivatives by the calcium phosphate precipitation method (Gorman *et al.*, 1982); cells were glycerol-shocked for 2 min with 15% (v/v) glycerol in Hanks balanced salt solution 4 h later (Corsaro & Pearson, 1981), washed three times with phosphate-buffered saline (138 mM-NaCl, 8.1 mM-NaH₂PO₄, 2 mM-KH₂PO₄, 2.7 mM-KCl, 0.9 mM-CaCl₂ and 0.5 mM-MgCl₂) and incubated in fresh media at 37°C for an additional 24–48 h. Before harvesting, cells were sequentially washed with cold phosphate-buffered saline and STE (0.04 M-Tris/Cl, pH 7.4, 1 mM-EDTA, 0.15 M-NaCl) and scraped; cell pellets were subjected to freeze–thaw cycles in 0.25 M-Tris/HCl, pH 7.8, the cell debris was pelleted and the cell extract was used for the chloramphenicol acetyltransferase (CAT) assay. Acetylation reactions were run for various lengths of time for different cell types; reaction products were spotted on t.l.c. plates and developed in chloroform/methanol (19:1, v/v) (Gorman *et al.*, 1982). The t.l.c. plates were sprayed with ENhance and exposed to XAR-2 Kodak film. The conversion of chloramphenicol to acetylated forms was quantified by two methods: autoradiographs were quantified by laser-assisted densitometry or by scraping the t.l.c. plates and measuring radioactivity by liquid scintillation spectroscopy. To correct for differences in transfection efficiencies, in most experiments the test constructs were co-transfected with a plasmid containing β -galactosidase driven by the Rous sarcoma virus long-terminal repeat; β -galactosidase activity was determined according to previously published procedures (Chang & Brenner, 1988). Additionally, to circumvent the intrinsic variation among different cell types, CAT assays were performed using constant amounts of total cellular protein, as determined by the Bradford assay (Bradford, 1976).

RESULTS

A collagen promoter/enhancer drives expression of a heterologous gene in a tissue-specific manner

Plasmid pCOL-KT was generated by placing the 5'-non-coding and contiguous DNA spanning the first exon and intron of the pro α 1(I) gene in front of the CAT gene, followed by removal of 410 bp (+68 to +478) containing the first exon and the translation initiation site (Thompson *et al.*, 1991). Conceivably, altered spatial relationships between the promoter and the first intron (due to a 410 bp deletion) in pCOL-KT, compared with the endogenous gene, could result in altered regulatory properties of such a construct. To circumvent the potential problem of artifactual expression of pCOL-KT, we had established earlier that the site of initiation of transcripts of endogenous pro α 1(I) and of pCOL-KT was identical (Thompson *et al.*, 1991). Next, to test the transcriptional tissue specificity of the endogenous pro α 1(I) gene promoter/enhancer, we transfected a variety of mesenchymal and non-mesenchymal cells and determined CAT activity by transient expression assays. A number of mesenchymal cell lines, including human fibroblasts (HFL-1 and IMR-90) and a rhabdomyosarcoma cell line (A204), were transfected with pCOL-KT; parallel cultures were also transfected with pSV₂CAT and pSV₀CAT DNA to monitor expression of appropriate positive and negative controls. Expression of pCOL-KT occurred readily in HFL-1 and A204 cells transfected with either pSV₂CAT or pCOL-KT (Fig. 1); similar abundant expression of the reporter CAT gene was also seen in IMR-90 and NIH 3T3 cells and a number of human dermal and

synovial fibroblast lines (results not shown). Two cell lines of non-mesenchymal origin, a rat pheochromocytoma cell line (PC12) and a human erythroleukaemia cell line (K562), which do not produce type I collagen, were also transfected with pCOL-KT, pSV₀CAT and pSV₂CAT. While K562 and PC12 cells transfected with pSV₂CAT, showed substantial expression of CAT driven by the strong simian virus-40 promoter/enhancer, neither of these cells transfected with pCOL-KT expressed significant levels of CAT activity (Fig. 1). These data reveal that (i) the organization of the promoter-proximal and intronic regulatory sequences in pCOL-KT is compatible with the expression of appropriately initiated transcripts, and (ii) pCOL-KT expression occurs in a tissue-specific manner akin to the endogenous pro α 1(I) gene.

Although it seemed that sequence up to only 800 bp upstream of the transcription start site in pCOL-KT was sufficient for its expression in a tissue-specific manner, we were curious to investigate if other sequences further upstream were involved in the regulation of the pro α 1(I) gene. Thus we ligated an additional 1200 bp to pCOL-KT to create 'extended-promoter' pCOL-KT, [pCOL-KT(L)]. We found that although the patterns of tissue-specific expression of pCOL-KT(L) and pCOL-KT were identical, pCOL-KT(L) was expressed at significantly lower levels in a number of cell lines of mesenchymal origin, compared with pCOL-KT (H. Poppleton, J. Thompson & R. Raghov, unpublished work). Thus we elected to identify the *cis*-regulatory determinants of tissue specificity by site-specific alterations of pCOL-KT.

A sequence located upstream of the pro α 1(I) collagen gene promoter negatively regulates collagen gene expression

The precise interactions among the transcriptional regulatory elements located in the 5'-upstream and intronic regions of the pro α 1(I) gene remain somewhat unresolved and controversial. We reasoned that the two important sources of the discrepancies reported from different laboratories could be because either (i)

Fig. 1. The plasmid pCOL-KT is preferentially expressed in cells of mesenchymal lineage

Two mesenchymal [HFL-1 (a) and A204 (b)] and two non-mesenchymal [K562 (c) and PC12 (d)] cell lines were transfected with pCOL-KT. Parallel cultures were not transfected with DNA (no DNA) or were transfected with a promoterless CAT plasmid (pSV₀CAT) or with CAT driven by the simian virus 40 promoter/enhancer (pSV₂CAT). Cells were harvested 48 h after transfection and the cell extracts were assayed for conversion of ¹⁴C-labelled chloramphenicol into acetylated chloramphenicol by t.l.c. coupled with fluorography. pCOL-KT is readily expressed in HFL-1 and A204 cells and not expressed in K562 and PC12 cells. There is also abundant expression of pSV₂CAT in all cell types regardless of their phenotype, while there is no expression of pSV₀CAT in any of the cells.

Fig. 2. Negative and positive *cis*-acting elements in the promoter-proximal region and the first intron regulate transcription of the human *pro α 1(I)* gene in human lung fibroblasts (HFL-1) and a rhabdomyosarcoma cell line (A204)

HFL-1 (a) or A204 (b) cells were transfected with pCOL-KT and restriction-endonuclease-derived deletion constructs of pCOL-KT. Parallel cultures were also transfected with pSV₀CAT or pSV₂CAT. Cell extracts were prepared and assayed for CAT activity to determine the fraction of ¹⁴C-labelled chloramphenicol converted into its acetylated derivatives. Single deletions of nucleotides –625 to –442 (*EcoRV*), –804 to –331 (*XbaI*) or +670 to +1440 (*SstII*) resulted in decreased CAT activity of varying degrees in the two cell lines. Double deletions in the promoter (*XbaI*, nt –804 to –331; *XbaI**, nt –804 to –609) and the first intron (nt +670 to +1440) increased CAT activity and appeared to relieve the negative impact of single deletions; the most pronounced increase in CAT expression was shown by the *XbaI*/SstII* double deletion construct (nt –804 to –609 and nt +670 to +1440) in HFL-1 and A204 cells. The *SphI/XmaIII* deletion (nt –447 to –83) which removed the canonical CAAT box sequences (Fig. 3) resulted in complete lack of expression in both cell lines.

the previously tested constructs had artifactually changed the spatial organization of the promoter and intronic sequences, leading to an altered mode of expression compared with the endogenous gene, or (ii) the intrinsic variations among different test systems (e.g. *Xenopus* oocytes versus cell lines) used to support transcription of the *pro α 1(I)* gene yielded disparate observations. It should be emphasized, however, that these potential problems are not mutually exclusive, and some combination of the two could yield the variations reported from different laboratories. To circumvent the first problem, we created pCOL-KT, in which the relationship between the 5'-upstream sequences and the first intron was similar to that of the endogenous *pro α 1(I)* gene. In the following experiments, we wished to test the functional consequence of series of single and double deletions in the *pro α 1(I)* promoter/enhancer as a first step in our analysis of the *cis*-regulatory elements of this gene. A representative analysis of the relative expression of various single or double deletion constructs is shown in Fig. 2, and the results from several independent determinations are summarized in Fig. 3. In both HFL-1 and A204 cells, the *EcoRV* (lacking nucleotides –625 to –442) and *XbaI* (nt –804 to –331) deletion constructs

of pCOL-KT showed significant decreases in CAT activity (2–10-fold). Deletion of the *SstII* fragment in the intron (nt +670 to +1440) also led to a substantial decrease in expression compared with pCOL-KT. As anticipated, the *XmaIII/SphI* deletion mutant (nt –477 to –83) of pCOL-KT, which removed the canonical CAATT box located at –127, completely abolished expression of CAT (Fig. 2).

Double deletion mutants of pCOL-KT were constructed by removing fragments of DNA after simultaneous digestion of the parent construct by two restriction enzymes (i.e. mixtures of *EcoRV*, *XbaI* and *SstII*), and the transcriptional activity of the deleted constructs was compared with that of pCOL-KT. Based on the representative data shown in Fig. 2, it appears that the negative impact of the single deletions in the promoter-proximal sequences was consistently relieved by deletion of the intronic sequences. For example, all three double deletion mutants (i.e. *EcoV/SstII*, *XbaI/SstII* and *XbaI*/SstII*) were expressed at significantly greater levels compared with constructs containing only single deletions. Of all our constructs, the construct containing the *XbaI*/SstII* double deletion (lacking nucleotides –804 to –609 and +670 to +1440) of pCOL-KT showed the most abundant expression: a 5–10-fold increase in CAT activity was observed in different experiments (results not shown). This remarkable effect of the *XbaI*/SstII* double deletions prompted us to test the effect of *XbaI** deletion more extensively. We found that deletion with *XbaI**, either alone or with the intronic *SstII* deletion, consistently boosted expression in all cell lines regardless of whether they were mesenchymal or non-mesenchymal in origin (see below).

Determinants of tissue specificity of the transcriptional control of the human *pro α 1(I)* gene are distributed in the promoter-proximal and intronic sequences

The phenotypes of the various deletion constructs of pCOL-KT strongly support the previously published observations (Rossouw *et al.*, 1987; Bornstein *et al.*, 1987, 1988; Bornstein & McKay, 1988; Liska *et al.*, 1990) showing that sequences located upstream of the start of transcription and in the first intron determine optimal expression of human *pro α 1(I)* collagen. Our observations extend earlier results by demonstrating the existence of an additional upstream negative regulatory element in the *pro α 1(I)* promoter (e.g. '*XbaI** domain'; nucleotides –804 to –609). We tested whether the *XbaI** domain was involved in determining the tissue-specific expression of the human *pro α 1(I)* gene. A number of different cell lines of non-mesenchymal origin were transfected with pCOL-KT and its various deletion derivatives, and CAT activity was determined after transient expression. The representative data for two of these cell lines, K562 and PC12, are shown in Fig. 4. While pCOL-KT was not detectably expressed in either cell line, deletion with *XbaI** (nt –804 to –609) resulted in excellent expression of CAT activity in both K562 and PC12 cells. Although introduction of the *SstII* deletion in *XbaI** constructs to create *XbaI*/SstII* (nt –804 to –609 and +670 to +1440) curtailed expression (5–10-fold), expression of the latter was significantly greater than that of pCOL-KT (Figs. 3 and 4). Interestingly, the other two constructs containing double deletions (e.g. *EcoRV/SstII* and *XbaI/SstII*) also showed slight expression in K562 cells (Figs. 3 and 4) and in PC12 cells (results not shown). To determine the contribution of the intronic sequences to the apparent breakdown of tissue specificity of transcription as a result of the *XbaI** deletion, we reintroduced the intronic *SstII* fragment in both positive and negative orientations in the construct containing the *XbaI*/SstII* deletion. The *XbaI** construct, which contains the +670 to +1440 *SstII* intronic fragment in its natural orientation, expressed CAT to much greater extent than did any of the single

Fig. 3. Summary of the site-specific deletions/alterations of pCOL-KT and their functional consequences in mesenchymal and non-mesenchymal cells

A schematic linear map of pCOL-KT (Thompson *et al.*, 1991) and boundaries of the deleted sequences are shown. The restriction enzymes used to create a particular deletion and the base pairs deleted are indicated. The filled arrow shows inversion of the *SstII* fragment in the *XbaI**/(-*SstII*) construct. Note that the lines representing the remaining upstream sequences in the deletion constructs are not drawn to scale in this diagram. Conversion of ¹⁴C-labelled chloramphenicol into acetylated chloramphenicol in extracts from cells 48 h post-transfection was determined by t.l.c. and autoradiography. Quantification was done either by laser-assisted densitometry or by scintillation spectrometry after scraping the area representing various spots. Non-specific background radioactivity was subtracted in all calculations. All values represent the averages derived from 3–5 independent determinations and have been corrected for the variation in transfection efficiency. The numbers indicate relative change in the expression of CAT activity of a given deletion/alteration compared with the expression of pCOL-KT. The value for pCOL-KT expression, even in the non expressing cells (e.g. K562 and PC12), was arbitrarily assigned as 1; for PC12 cells, data for the expression of only two constructs, pCOL-KT and *XbaI**, have been presented.

Fig. 4. Tissue-specific repression of the human pro α 1(I) collagen gene in non-mesenchymal cells is determined by both the promoter-proximal and intronic sequences

Cultured human erythroleukaemia K562 cells (a) or rat pheochromacytoma PC12 cells (b) were transfected with pCOL-KT and various deletion constructs of pCOL-KT. Cell extracts were assayed 48 h after transfection, and CAT activity was determined by t.l.c. coupled with fluorography. The *XbaI** construct (deletion of nucleotides -804 to -609) was expressed in both K562 and PC12 cells; combined deletions of the promoter-proximal sequences and the first intron (e.g. *XbaI/SstII*; *XbaI*/SstII*; *EcoRV/SstII*) were expressed to a much lesser extent in K562 cells. Altering the orientation of the putative enhancer in the first intron (nt +670 to +1440; *SstII* fragment) in constructs containing in *XbaI** deletion [e.g. *XbaI*/(-SstII)*] abolished transcriptional activity in both A562 and PC12 cells. Expression of pSV₂CAT was readily apparent in K562 and PC12 cells.

or double deletions in K562 cells (Fig. 4). In sharp contrast, although both K562 and PC12 cells expressed the *XbaI** deletion well, the *XbaI*/(-SstII)* plasmid, containing the *SstII* (+670 to

Fig. 5. Stereo-specific interaction between promoter-proximal and intronic sequences is required for optimal transcription of pCOL-KT in A204 cells

Cell monolayers were transfected with various constructs [e.g. pSV₀CAT, pSV₂CAT, pCOL-KT, *XbaI** or *XbaI*/(-SstII)*], glycerol-shocked, and cell extracts were prepared 48 h later for CAT assays as described in the Materials and methods section. Equal aliquots of extracts were incubated with [¹⁴C]chloramphenicol, and acetylated derivatives of chloramphenicol were visualized by t.l.c. coupled to autoradiography. While the construct containing the *XbaI** deletion was found to be expressed much more abundantly than pCOL-KT in these cells, there is no detectable expression of this construct if the *SstII* fragment is placed in the reverse orientation [e.g. *XbaI*/(-SstII)*].

+1440) fragment in the opposite orientation, was not expressed in either K562 cells (Fig. 4) or PC12 cells (results not shown). Similarly, the *XbaI*/(-SstII)* construct failed to be expressed in A204 cells (Fig. 5) or in many other cells of mesenchymal origin

(C. P. Simkevich, unpublished work). We interpret these data to mean that the *XbaI** domain contains *cis*-acting negative regulatory element(s) which determine tissue-specific expression of this gene and removal of these sequences from pCOL-KT leads to promiscuous expression of the collagen gene in non-mesenchymal cells. Furthermore, based on the failure of expression of *XbaI**/(*-SstII*), we hypothesize that orientation-dependent interactions between the *XbaI** domain and the intronic sequences determine the transcriptional specificity of the pro α (I) collagen gene.

DISCUSSION

Examination of the functional consequences of a number of deletion mutants of pCOL-KT in mesenchymal and non-mesenchymal cells revealed that the transcriptional tissue specificity of the human pro α 1(I) collagen gene is determined by a negative *cis*-regulatory element(s) located upstream of the start of transcription. Removal of this 195 bp DNA fragment (*XbaI** domain; -804 to -609) from pCOL-KT leads to abundant expression of this construct in non-mesenchymal cells. We also demonstrate that the putative negative element(s) and sequences adjacent to the fragment interact with the DNA elements located in the first intron of the pro α 1(I) gene, and this interaction occurs in an orientation-dependent manner. The first intron of the human pro α 1(I) gene contains both positive and negative *cis*-regulatory modules (Rossouw *et al.*, 1987; Bornstein *et al.*, 1987, 1988; Bornstein & McKay, 1988; Liska *et al.*, 1990). Apparently, the enhancer-like action of the intronic sequences of the human pro α 1(I) collagen gene is particularly orientation-dependent and spatially restricted (Bornstein *et al.*, 1988; Boast *et al.*, 1990). Our analyses strongly corroborate the notion of stereo-specificity of promoter-intron interactions, as postulated by Bornstein *et al.* (1988). Since the action of enhancers is generally orientation-independent and spatially unrestricted (i.e. they activate transcription when placed either upstream or downstream of the promoter), the human pro α 1(I) collagen gene enhancer differs from the classical enhancers (Atchison, 1988) in a fundamental way. This is especially intriguing, since the transcriptional enhancer identified in the first intron of the murine pro α 2(I) collagen gene apparently resembles classical enhancers and was found to be orientation- and location-independent (Rossi & de Crombrughe, 1987). At present a precise molecular explanation for the underlying stereo-specificity of the human pro α 1(I) enhancer is not available. Such orientation specificity may result from a unique combinatorial arrangement of the *cis*-regulatory modules in the intron and conformational constraints imposed by the optimal interaction of the intronic sequences with the 5' regulatory sequences (Bornstein *et al.*, 1988; Boast *et al.*, 1990). Although we favour the highly speculative but plausible model for the potential interaction between the promoter-proximal and intronic sequences of the human pro α 1(I) gene (Bornstein *et al.*, 1988), the precise molecular interactions await further experimentation.

At present, the nature of the *cis*-acting sequences which determine the tissue specificity of transcription of the human pro α 1(I) collagen gene is somewhat confusing and controversial. Consistent with our observations, it was shown that a 804 bp segment located upstream of the start of transcription in conjunction with the intronic DNA was sufficient to dictate mesenchymal-cell-specific transcription of the human pro α 1(I) collagen gene (Bornstein *et al.*, 1987). A construct containing the -331 to +98 pro α 1(I) promoter linked to CAT, but lacking the first intron, was expressed somewhat non-specifically in baby hamster kidney (BHK) cells, HeLa cells and chicken tendon fibroblasts (CTF); ligation of an intronic fragment A274 (+819

to +1093) to sequences upstream of the promoter resulted in the cell-specific inhibition of transcription (Bornstein *et al.*, 1987). Based on these data, these authors concluded that the intronic sequences were involved in the mesenchymal cell specificity of transcription of the human pro α 1(I) gene.

More recently, Boast *et al.* (1990) showed that reporter constructs driven by upstream sequences containing nucleotides -2500 to +123 of the human pro α 1(I) gene were expressed at low levels in both fetal fibroblasts and immortalized human lymphoblasts; an incremental removal of the upstream sequences from nucleotides -2500 to -331 increased CAT activity slightly in fibroblasts, but had little effect in the lymphoblast cell line. Boast *et al.* (1990) placed the putative intronic enhancer upstream of the -806 bp pro α 1(I) promoter and observed that it had an orientation-preferred enhancing effect in human fetal fibroblasts and immortalized human lymphoblasts. Thus it was claimed that placing the 782 bp intronic fragment upstream did not confer tissue specificity with the -806 bp pro α 1(I) promoter. Furthermore, when the 782 bp intronic enhancer was placed upstream of the truncated promoter (containing only 331 bp of the upstream sequences), there was no stimulatory effect seen in either orientation in the human foreskin fibroblasts, which were the only cells expressing CAT activity with the truncated promoter. Neither the lymphoid cells nor the two other mesenchymal cell lines (EF and MC2) expressed constructs containing the shortened promoters. Based on these data, it was concluded that there was a virtual lack of tissue-specificity of the stimulatory effect of the 782 bp intronic DNA (Boast *et al.*, 1990). These data are in sharp contrast to those obtained by Sherwood & Bornstein (1990); these authors reported that the 786 bp intronic fragment (+822 to -1607), ligated upstream of the human pro α 1(I) promoter at nt -804 in either orientation, had no effect in SV-MSC cells and a modest (2-fold) inhibitory effect in CTF cells. Our observations are more consistent with those of Bornstein *et al.* (1987) and Slack *et al.* (1991), and suggest that the 804 bp upstream sequences, in combination with the first intron, are indeed sufficient for proper transcriptional tissue specificity of the pro α 1(I) gene. We also provide evidence for the conclusion that, although there exist additional regulatory sequences upstream of position -804, these are not required to confer tissue specificity (H. Poppleton, J. Thompson & R. Raghov, unpublished work); it is conceivable that these sequences play a potential role in modulating pro α 1(I) collagen gene expression in response to other physiological or pharmacological stimuli known to affect collagen genes (Raghov & Thompson, 1989; Bornstein & Sage, 1989; Ramirez & DiLiberto, 1990).

The search for tissue-specific determinants of transcription of genes in higher eukaryotes is a very active area of investigation (Ptashne, 1988; Mitchell & Tjian, 1989; Ptashne & Gann, 1990). While the precise molecular mechanisms are poorly understood, it is thought that interaction between the *cis*-regulatory sequences located on DNA and tissue-specific *trans*-acting factors regulate tissue-specific transcription (Johnson & McKnight, 1989; Latchman, 1990). While in some genes only a few hundred base pairs upstream from the initiation of transcription are needed (Edlund *et al.*, 1985; Behringer *et al.*, 1988), in others tissue-specific enhancers located 10-50 kb away from the start of transcription have been shown to regulate transcriptional tissue specificity (Hammer *et al.*, 1987; Talbot *et al.*, 1989). In a number of genes, tissue-specific enhancers were found to contain both positively and negatively acting sequence motifs located in close proximity to each other (Atchison, 1988). Previous observations suggested that tissue-specific transcription of the immunoglobulin heavy chain (Imler *et al.*, 1987) and the pro α 1(II) collagen genes (Savagner *et al.*, 1990) is dictated by negative *cis*-regulatory elements. It was postulated that, in non-expressing

cells, binding of a putative tissue-specific *trans*-acting factor(s) to the negative *cis*-regulatory motifs prevented interaction of the ubiquitous factors that are needed for transcriptional activation of these genes. Whether this occurs by direct steric hindrance between the target sequences and the *trans*-acting factor, or through more complicated associations involving protein-protein interactions (Ptashne, 1988; Mitchell & Tjian, 1989; Johnson & McKnight, 1989; Ptashne & Gann, 1990), is currently unknown.

These studies, performed mainly by C.P.S. as part of Ph.D. dissertation, benefited greatly from the moral support of many individuals in and out of the Raghow laboratory; C.P.S. particularly thanks Dr. M. Heimberg, Dr. S. Solomon and Dr. A. Kang for their continued encouragement and support. These investigations were supported by grants from the National Institutes of Health Specialized Center of Research in Rheumatoid Arthritis (P50-AR39166), and the Department of Veterans Affairs to R.R., who is an Associate Career Scientist of the Department of Veterans Affairs. We also acknowledge the skilled, patient and cheerful secretarial assistance of Kathy Pearson.

REFERENCES

- Atchison, M. L. (1988) *Annu. Rev. Cell Biol.* **4**, 127–153
- Barsh, G. S., Rousch, C. L. & Gelinas, R. E. (1984) *J. Biol. Chem.* **259**, 14906–14913
- Behringer, R., Matthews, L. S., Palmiter, R. D. & Brinster, R. L. (1988) *Genes Dev.* **2**, 453–461
- Boast, S., Su, M.-W., Ramirez, F., Sanchez, M. & Avvedimento, E. B. (1990) *J. Biol. Chem.* **265**, 13351–13356
- Bornstein, P. & McKay, J. (1988) *J. Biol. Chem.* **263**, 1603–1606
- Bornstein, P. & Sage, H. (1989) *Prog. Nucleic Acids Res. Mol. Biol.* **37**, 67–106
- Bornstein, P., McKay, J., Morishima, J. K., Devarayalu, S. & Gelinas, R. E. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8869–8873
- Bornstein, P., McKay, J., Liska, D. J., Apone, S. & Devarayalu, S. (1988) *Mol. Cell. Biol.* **8**, 4851–4857
- Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254
- Brenner, D. A., Rippe, R. A. & Veloz, L. (1989) *Nucleic Acids Res.* **17**, 6055–6064
- Chang, A. C. Y. & Brenner, D. G. (1988) *Focus* **4**, 66–69
- Choe, I., Aycock, R. S., Raghov, R., Myers, J. C., Seyer, J. M. & Kang, A. H. (1987) *J. Biol. Chem.* **262**, 5408–5413
- Cockyne, D. & Cutroneo, K. R. (1988) *Biochemistry* **27**, 2736–2745
- Corsaro, C. M. & Pearson, M. L. (1981) *Somatic Cell. Genet.* **7**, 603–616
- Edlund, T., Walker, M. D., Barr, P. J. & Rutter, W. J. (1985) *Science* **230**, 912–916
- Goldstein, R. H., Fine, A., Farnsworth, L. J., Poliks, C. & Polgar, P. (1990) *J. Biol. Chem.* **265**, 13623–13628
- Gorman, C. M., Moffat, F. & Howard, B. H. (1982) *Mol. Cell. Biol.* **2**, 1044–1051
- Hammer, R. E., Krumlauf, R., Camper, S. A., Brinster, R. L. & Tilghman, S. M. (1987) *Science* **235**, 53–58
- Imler, J., Lemaire, C., Waslyk, C. & Waslyk, B. (1987) *Mol. Cell. Biol.* **7**, 2558–2567
- Johnson, P. F. & McKnight, S. L. (1989) *Annu. Rev. Biochem.* **58**, 799–839
- Kuhn, K. (1987) in *The Classical Collagens: Types I, II and III in Structure and Function of Collagen Types* (Mayne, R. & Burgeson, R. E., eds.), pp. 1–42, Academic Press, New York
- Latchman, D. S. (1990) *Biochem. J.* **270**, 281–289
- Liska, D. J., Slack, J. L. & Bornstein, P. (1990) *Cell. Regul.* **1**, 487–498
- Miller, E. J. & Gay, S. (1987) *Methods Enzymol.* **144**, 3–41
- Mitchell, P. G. & Tjian, R. (1989) *Science* **245**, 371–378
- Olsen, A. S., Geddis, A. E. & Prockop, D. J. (1991) *J. Biol. Chem.* **266**, 1117–1121
- Penttinen, R. P., Kobayashi, S. & Bornstein, P. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1105–1108
- Postlethwaite, A. E., Raghov, R., Stricklin, G. P., Poppleton, H., Seyer, J. M. & Kang, A. H. (1988) *J. Cell Biol.* **106**, 311–318
- Ptashne, M. (1988) *Nature (London)* **335**, 683–689
- Ptashne, M. & Gann, A. A. F. (1990) *Nature (London)* **346**, 329–331
- Raghow, R. & Thompson, J. P. (1989) *Mol. Cell. Biochem.* **86**, 5–18
- Raghow, R., Gossage, D., Seyer, J. M. & Kang, A. H. (1984) *J. Biol. Chem.* **259**, 12718–12723
- Raghow, R., Gossage, D. & Kang, A. H. (1986) *J. Biol. Chem.* **261**, 4677–4684
- Raghow, R., Kang, A. H. & Pidikiti, D. (1987a) *J. Biol. Chem.* **262**, 8409–8415
- Raghow, R., Postlethwaite, A. E., Keski-Oja, J., Moses, H. L. & Kang, A. H. (1987b) *J. Clin. Invest.* **79**, 1285–1288
- Ramirez, F. & DiLiberto, M. (1990) *FASEB J.* **4**, 1616–1623
- Reue, K., Leff, T. & Breslow, J. L. (1988) *J. Biol. Chem.* **263**, 6857–6864
- Rippe, R. A., Lorenzen, S.-I., Brenner, D. A. & Breindl, M. (1989) *Mol. Cell. Biol.* **9**, 2224–2227
- Roberts, A. B., Sporn, M., Assoian, R. K., Smith, J. M., Roche, N. S., Wakefield, L. M., Heine, V. I., Liotta, L. A., Falanga, V., Kehrl, J. M. & Fauci, A. S. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4167–4171
- Rossi, P. & de Crombrughe, B. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5590–5594
- Rossi, P., Karsenty, G., Roberts, A. B., Roche, N. S., Sporn, M. B. & de Crombrughe, B. (1988) *Cell* **52**, 405–414
- Rossouw, C. M. S., Vergeer, W. P., du Plooy, S. J., Bernard, M. P., Ramirez, F. & de Wet, W. J. (1987) *J. Biol. Chem.* **262**, 15151–15157
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- Savagner, P., Miyashita, T. & Yamada, Y. (1990) *J. Biol. Chem.* **265**, 6669–6674
- Sherwood, A. L. & Bornstein, P. (1990) *Biochem. J.* **265**, 895–897
- Slack, J. L., Liska, D. J. & Bornstein, P. (1991) *Mol. Cell. Biol.* **11**, 2066–2074
- Solis-Herruzo, J. A., Brenner, D. A. & Chojkier, M. (1988) *J. Biol. Chem.* **263**, 5841–5845
- Talbot, D., Collis, P., Anthoniou, M., Vidal, M., Grossveld, F. & Greaves, D. R. (1989) *Nature (London)* **338**, 352–355
- Thompson, J. P., Simkevich, C. P., Holness, M. A., Kang, A. H. & Raghov, R. (1991) *J. Biol. Chem.* **266**, 2549–2556
- Vuorio, E. & de Crombrughe, B. (1990) *Annu. Rev. Biochem.* **59**, 837–872

Received 3 December 1991/21 February 1992; accepted 27 February 1992