Transcriptional activation of the minimal human Pro α 1(I) collagen promoter: obligatory requirement for Sp1

Helen M. POPPLETON* and Rajendra RAGHOW*†‡

*Department of Pharmacology, College of Medicine, The University of Tennessee, Memphis, TN, and †Veterans Affairs Medical Center, 1030 Jefferson Avenue, Memphis, TN 38104, U.S.A.

A construct containing human Pro α 1(I) collagen gene promoter/enhancer-driven chloramphenicol acetyltransferase (CAT), pCOL-KT, failed to be expressed significantly in Sp1-deficient Schneider *Drosophila* line 2 (SL2) cells. However, CAT expression was induced 200-fold in SL2 cells co-transfected with pCOL-KT and pPACSp1, an Sp1-expression vector driven by the *Drosophila* actin 5C promoter. Elimination of the four potential Sp1-binding sites from pCOL-KT (pCOL-KT Δ I), by removal of the first intron, did not abrogate Sp1-mediated induction of CAT. Even more significantly, a minimal Pro α 1(I) collagen promoter (–100 to +117 bp), containing a TATA box (–28 to –25 bp) and one putative Sp1-binding site (–87 to –82 bp), elicited strong Sp1-induced transactivation. Furthermore, mutation of the Sp1 motif

in the minimal Pro α 1(I) collagen promoter–CAT construct abolished Sp1-induced expression of the reporter gene. Purified Sp1 protein bound specifically to DNA fragments of the Pro α 1(I) minimal promoter encompassing the putative Sp1-binding site; Sp1 binding could be competed out by a double-stranded oligonucleotide containing the wild-type Sp1 sequence, while an oligonucleotide containing a mutated Sp1 site failed to compete. Based on these results, we postulate that Sp1 plays an obligatory role in the transcriptional activation of the human Pro α 1(I) collagen gene. Additionally, we propose that a *bona fide* Sp1 motif, located most proximal to the TATA box, is necessary and sufficient for Sp1-mediated activation of the minimal Pro α 1(I) collagen promoter.

INTRODUCTION

Type I collagen is a heterotrimer composed of two Pro α 1(I) and one Pro α 2(I) chains; individual α chains are encoded by two unique genes located on different chromosomes [1,2]. Collagen genes are regulated in response to developmental cues, in a tissue-specific manner, as well as by a variety of biological and pharmacological stimuli [3,4]. Although expression of collagen genes is controlled by transcriptional, post-transcriptional and post-translational mechanisms, the predominant mode of regulation appears to be transcriptional [2–5].

Regulation of collagen gene expression has been analysed in a variety of cells cultivated *in vitro*, as well as in transgenic mice. Transcriptional control of the human Pro α 1(I) collagen gene appears to involve complex interactions among the positive and negative regulatory sequences, located on either side of the start site of transcription; recent studies have unravelled a number of apparent inconsistencies between the tissue-specific expression of the collagen gene *in vitro* and in transgenic mice [5]. Based on experiments done in cells in culture, it has been suggested that the *cis*-acting elements located in the first intron of the human and mouse Pro α 1(I) and Pro α 2(I) collagen genes are important in transcriptional regulation [5]. A 274 bp *Ava*I fragment of the human Pro α 1(I) gene (between +820 and +1093 bp) has been shown to inhibit expression of the reporter gene in an orientation-dependent manner [6,7], whereas a 782 bp intronic fragment encompassing the same region of the human Pro α 1(I) collagen gene (between +820 and +1602 bp) has been shown to significantly increase transcription [8]. Recent studies *in vitro* in our laboratory have corroborated the notion that *cis*-acting elements

in the promoter and the first intron interact to regulate transcription and these interactions maintain tissue-specific expression of the human Pro α 1(I) collagen gene *in vivo* [9,10]. Deletion of the *cis*-regulatory element located between –804 and –331 bp in the human Pro α 1(I) collagen promoter enabled non-mesenchymal cell lines (K562 and PC-12) to express the reporter gene chloramphenicol acetyltransferase (CAT). However, this effect was abrogated by deletion of an intronic fragment between +670 and +1440 bp [9]. An enhancer element, between +418 and +1524 bp, has also been described in the first intron of the mouse Pro α 2(I) collagen gene [5].

Olsen et al. [11] analysed the expression of a human Pro α 1(I) collagen mini-gene containing 2.3 kb of the 5'-upstream sequences and lacking the first intron and also containing the last six exons and introns, and 2 kb of the 3'-flanking sequence in NIH 3T3 cells. These authors argued that the first intron contributes little to the expression of the Pro α 1(I) gene. Similarly, deletion of the first intron from a mouse Pro α 2(I) collagen–CAT chimeric gene neither diminished its transcription in transgenic mice nor affected its tissue-specific expression [12,13]. In contrast, tissue-specific expression of the human Pro α 1(I) gene in transgenic mice was apparently mediated by *cis*-acting elements in the promoter and first intron; furthermore, when all but 440 bp of the human promoter sequences were deleted, expression of the transgene was significantly reduced in such collagen-producing tissues as lung, bone, tail and skin [5,14].

To date, a number of *trans*-acting factors which bind to the Pro α 1(I) collagen promoter/enhancer have been described. These include a factor(s) binding to the transforming growth factor- β 1-activating element (TAE) in the rat Pro α 1(I) collagen promoter

Abbreviations used: CAT, chloramphenicol acetyltransferase; TAE, transforming growth factor- β 1-activating element; IF2, inhibitory factor 2; SL2, Schneider *Drosophila* line 2; pCOL-KT, plasmid vector in which expression of CAT is driven by 5'-upstream sequences and most of the first intron of the human Pro α 1(I) collagen gene; pCOL-KT Δ I, intronless pCOL-KT; EMSA, electrophoretic gel mobility shift assay; pPACSp1, plasmid vector containing the *Drosophila* actin 5C promoter to drive expression of human Sp1; pPACSp1(–), expression plasmid in which Sp1 cDNA sequences are cloned in the antisense orientation; hGH, human growth hormone.

‡ To whom correspondence should be addressed at the Veterans Affairs Medical Center.

[5]; apparently, the mouse Pro α 2(I) TAE is located between -308 and -322 bp [5]. The proteins binding to the TAE have not been identified, but appear to have a molecular mass of 34 kDa and are apparently distinct from NF1 and AP2, which also bind to this sequence [16]. NF1 and Sp1 have both been shown to bind to two regions of the mouse Pro α 1(I) collagen gene between -129 and -110 bp and between -105 and -78 bp; over-expression of NF1 in NIH 3T3 cells increased expression of a Pro α 1(I) promoter-luciferase construct [5]. Furthermore, 2 bp mutations in the NF1-binding sites led to reduced expression of the reporter gene [5]. The functional significance of Sp1 binding to the mouse Pro α 1(I) promoter is less well defined. Over-expression of Sp1 in NIH 3T3 cells inhibited both basal and NF1-stimulated expression of a mouse Pro α 1(I) collagen promoter-CAT construct [5], in contrast with the stimulatory activity more usually ascribed to Sp1; it was speculated that over-expression of Sp1 may interfere with binding of NF1 and consequently attenuate the ability of NF1 to activate transcription [17]. Another *trans*-acting factor, inhibitory factor 2 (IF2), which binds to the same GC-rich region of the mouse Pro α 1(I) promoter, appears to be distinct from Sp1 [18]. In the human Pro α 1(I) gene two potential Sp1-binding sites are located 806 bp upstream of the start of transcription site, and four putative Sp1 sites are found in the first intron [5]. We report here that expression of the human Pro α 1(I)-CAT constructs in Schneider *Drosophila* line 2 (SL2), which lack Sp1 [19], was undetectable and expression could be strongly induced by exogenously provided Sp1. Additionally, the most proximal authentic Sp1-binding motif, located at -88 bp in the minimal human Pro α 1(I) promoter, was necessary and sufficient for Sp1-mediated transcriptional activation.

EXPERIMENTAL

Materials

All media and supplements used for tissue culture, restriction enzymes and other nucleic acid modifying enzymes were purchased from Gibco-BRL (Gaithersburg, MD, U.S.A.). [14 C]Chloramphenicol and acetyl-CoA were from New England Nuclear (Boston, MA, U.S.A.) and Sigma (St. Louis, MO, U.S.A.) respectively. Silica-coated TLC plates were purchased from Baker (Phillipsburg, NJ, U.S.A.). The mutagenesis kit used to mutate the binding site of Sp1 was purchased from Promega (Madison, WI, U.S.A.). Nucleotide sequence of the mutants was verified by dideoxy sequencing using a kit purchased from United States Biologicals (Cleveland, OH, U.S.A.). Restriction fragment and double-stranded oligonucleotide probes were labelled with [α - 32 P]deoxyadenosine and [α - 32 P]deoxycytidine triphosphates purchased from New England Nuclear. A double-stranded oligonucleotide containing a consensus Sp1 site was purchased from Stratagene (San Diego, CA, U.S.A.). *Taq* polymerase and deoxynucleotide triphosphates (dNTPs) used in PCR were purchased from Perkin Elmer Cetus (Norwalk, CT, U.S.A.) and Pharmacia (Uppsala, Sweden) respectively. An Ericomp Twinblock Thermocycler System (San Diego, CA, U.S.A.) was used to carry out PCR reactions. The *Drosophila* Sp1 expression vector in sense and antisense orientations was kindly given by Dr. Robert Tjian, University of California, Berkeley, CA, U.S.A.

Oligonucleotide primers and probes

Oligonucleotide primers used for PCR were synthesized in the Molecular Biology Core Laboratory of the Specialized Center of Research on Rheumatoid Arthritis, U.T. Memphis, on an

Figure 1 A schematic summary of the deletion mutants of pCOL-KT

Linearized structures of the basic human Pro α 1(I) collagen promoter-reporter construct, pCOL-KT [21], its intronless counterpart, pCOL-KT Δ I and various congeners of pCOL-KT Δ I containing serially truncated promoter are shown. Primers containing restriction sites for *Kpn*I and *Pst*I were used to amplify the promoter sequences of the human Pro α 1(I) gene by PCR and cloned into pKT [22]. The 5' promoter-proximal, intronic and CAT DNA sequences are denoted as unfilled, shaded and solid bars respectively. Locations of the potential Sp1-binding sites in the promoter and in the first intron are identified by asterisks (*). The numbers in parentheses indicate either the Pro α 1(I) collagen-specific terminal nucleotides or the 5' location of the hexameric GC boxes (*).

Applied Biosystems 381A DNA synthesizer. The forward (5'GATCGGTACCCCTAGTCATGTTTCTCACCCCTG3'; sequence from -781 to -760 bp) and reverse (5'GATCCTGCAGCTAGACCCTAGACATGTAGA3'; sequence from +98 to +117 bp) primers used to create intronless pCOL-KT (pCOL-KT Δ I; Figure 1) were designed to contain unique *Kpn*I and *Pst*I sites respectively (pCOL-KT is a plasmid vector in which expression of CAT is driven by 5'-upstream sequences and most of the first intron of the human Pro α 1(I) collagen gene). The upstream sequences of pCOL-KT Δ I were serially shortened by PCR. DNA fragments amplified with forward primers, 5'GATCGGTACCTTGGCTTGGGGACGGGCGGCC3' (-100 to -80 bp), 5'GATCGGTACCTAGATTGGAGGTCCCAGGAAG3' (-330 to -310 bp) or 5'GATCGGTACCCCTAGCCACACCAGAAGACAC3' (-560 to -540 bp), and a common reverse primer (sequences from +98 to +117 bp), were ligated in front of CAT. For mutagenesis of the Pro α 1(I) collagen promoter sequence containing a potential Sp1 site the following primers were used: 5'GGGCACGGGAGGCCGCTC3' (-75 to -94 bp; C₈₄ changed to A₈₄) and 5'GCTGGGGCACGGAAAGCCGGCTCCCC3' (-79 to -98 bp; GCG₈₃₋₈₅ changed to AAA₈₃₋₈₅). Mutation of the Sp1 site at -87 to -80 bp (CGGGCGGCC) was achieved using a commercially available mutagenesis kit (Altered sites; Promega). Oligonucleotides used to prepare double-stranded DNA for gel electrophoretic mobility shift assays (EMSAs) were: wild-type Sp1, 5'GGGGCACGGGCGGCCGCTCC3' (-94 to -74 bp) and 5'AGGGGGAGCCGGCCGCGCCCTGTG3' (-70 to -90 bp); C to A mutant, 5'GGGCACGGGAGGCCGCTC3' (-93 to -76 bp) and 5'GACGAGCCGGCCTCCCGTG3' (-72 to -90 bp); GCG to AAA mutant, 5'GCTGGGGCACGGAAAGCCGGCTCCCC3' (-97 to -71 bp) and 5'GAGAGGGGGAGCCGGCTTTCCTGTGCC3' (-67 to -93 bp).

Cells

SL2 was obtained from Dr. Carl Wu, National Institutes of Health, Bethesda, MD, U.S.A. and grown at 25 °C without CO₂ in Schneider's *Drosophila* medium supplemented with 10 %

heat-inactivated fetal bovine serum and antibiotics (100 units/ml penicillin, 100 μ g/ml streptomycin and 25 μ g/ml fungizone). Cultures were seeded at a density of $(1-5) \times 10^6$ cells/ml and were subcultured by 4-5-fold dilution in fresh medium every 3 days [17].

DNA vectors

Construction of the plasmid pCOL-KT, in which 5'-upstream sequences (to -804 bp) of the human Pro α 1(I) collagen gene and most of the first intron (+474 to +1440 bp) drive expression of the CAT gene (Figure 1), has been described previously [9,20]. Intronless pCOL-KT (Figure 1) was created by PCR to amplify promoter sequences of the Pro α 1(I) gene from -781 to +117 bp. The conditions used for PCR amplification were as follows: 30 cycles of 2 min denaturation at 94 °C, 1 min annealing at 59 °C and 2 min extension at 72 °C. Reaction mixture (100 μ l) contained 10 $\times$ PCR buffer [100 mM Tris/HCl (pH 8.3)/500 mM KCl/0.1 % (w/v) gelatin], 200 μ M dNTPs, 465 ng of each primer; 1 μ g of template DNA {pUC19 clone containing a 5 kb *Bam*HI fragment of the Pro α 1(I) gene as reported earlier [21]; and 2.5 units of *Taq* polymerase. Concentration of MgCl₂ was adjusted to be 0, 1.5, 3.0 or 4.5 mM. Subsequent amplifications of DNA fragments for creation of other deletion mutants were done by using identical conditions. A 921 bp amplified DNA fragment was ligated into the multiple cloning site of plasmid pKT [22]; the resulting plasmid was designated pCOL-KT Δ I (Figure 1). Deletions in the upstream sequences of pCOL-KT Δ I were made by PCR to generate fragments from (i) -100 to +117 bp, (ii) -330 to +117 bp and (iii) -560 to +117 bp which could be ligated into the multiple cloning site of pKT, as described above.

Transfections and CAT assays

SL2 cells [$(5-8) \times 10^6$ /ml] were transfected with 1-5 μ g of pCOL-KT DNA or its derivatives, using the calcium phosphate precipitation method [20,23]. SL2 cells were grown with DNA-calcium phosphate complexes for 48 h; cells were then harvested, washed once with PBS, and suspended in 0.25 M Tris/HCl, pH 7.8. Cell extracts were prepared as previously described [9,20] and used in CAT assays [22]. In co-transfection experiments, SL2 cells were transfected with 1-5 μ g of pCOL-KT plasmids along with 10 ng to 5 μ g of Sp1-expression plasmid, pPACSp1. Total plasmid DNA in each transfection reaction was adjusted to 10 μ g with pUC19 DNA. All CAT expression levels were normalized against total cell protein as determined by the Bradford assay (Bio-Rad).

Preparation of nuclear extracts and EMSAs

Nuclear extracts were prepared from SL2 and NIH 3T3 cells as previously described [24]. Known protein concentrations of nuclear extracts, as determined by the Bradford assay, and affinity-purified human recombinant Sp1 (0.5 footprinting units; 15 ng of protein) were used in EMSAs. EMSAs were performed as described previously [25]. A 217 bp *Eco*RI-*Pst*I fragment containing sequences from -100 to +117 bp of the Pro α 1(I) gene was incubated at room temperature for 20 min with 100 μ Ci of [α -³²P]dATP (800 Ci/mM), 200 μ M dNTPs and 5 units of the Klenow enzyme and the 217 bp end-labelled fragment excised from a 6 % (w/v) non-denaturing gel in 1 $\times$ TBE [100 mM Tris (pH 8.3)/123 mM boric acid/2 mM EDTA]. Radiolabelled probe was eluted from the gel overnight at room temperature in 10 mM Tris/HCl (pH 7.5)/1 mM EDTA (pH 8.0)/1 M NaCl and precipitated with ethanol. Restriction fragments containing the mutated Sp1 sites were labelled in the same way. Unlabelled

*Eco*RI-*Pst*I restriction fragments prepared as above were used in competition experiments. Unlabelled oligonucleotide fragments containing a consensus Sp1 element, and wild-type or mutant Pro α 1(I) collagen Sp1 elements, were also used in competition experiments.

RESULTS

Sp1 is an obligatory *trans*-acting factor for the expression of the human Pro α 1(I) collagen gene

The *Drosophila* cell line, SL2, was chosen to investigate the role of the transcription factor Sp1 in the regulation of human Pro α 1(I) collagen gene expression, because *Drosophila* cells provide an Sp1-deficient background [26]. Since the major mechanistic features of the process of transcription are thought to be highly conserved between *Drosophila* and mammalian systems [27,28], these cells have been used previously to test the precise contribution of Sp1 to the process of transcriptional activation of Sp1-dependent promoters [19].

The plasmid vector containing the *Drosophila* actin 5C (pPACSp1) promoter to drive expression of the human Sp1 has been shown to be expressed efficiently in SL2 cells [19]. In the initial series of experiments, SL2 cells were co-transfected with pPACSp1 and our basic Pro α 1(I) collagen promoter/enhancer-reporter plasmid, pCOL-KT [5,9,20], and CAT activity was measured 48 h after transfection. As illustrated in Figure 2, CAT expression was undetectable in SL2 cells transfected with the reporter plasmid pCOL-KT alone. In contrast, when pCOL-KT was co-transfected with pPACSp1, CAT expression was dramatically enhanced (Figure 2). The expression of pCOL-KT did not occur in SL2 cells co-transfected with pPACSp1(-), an expression plasmid in which the Sp1 cDNA sequences have been cloned in the antisense orientation, thus clearly demonstrating the specificity of the Sp1-mediated activation of the human Pro α 1(I) gene. This observation also satisfies one of the criteria

Figure 2 Sp1 is an obligatory *trans*-acting factor for expression of pCOL-KT in SL2 cells

SL2 cells were transfected with pCOL-KT and an Sp1 expression plasmid, pPACSp1, or control plasmids as detailed below. CAT activity in SL2 cell extracts was measured 48 h after transfection by TLC coupled with fluorography. A representative autoradiogram is shown. SL2 cells either transfected with pCOL-KT alone (lane 1) or co-transfected with pCOL-KT and pUC19 (lane 2) did not express CAT. However, co-transfection of SL2 cells with pCOL-KT and an Sp1 expression plasmid pPACSp1 (lane 3) increased CAT expression 200-fold. In contrast, expression of CAT remained unchanged in SL2 cells co-transfected with pCOL-KT and pPACSp1(-), an expression plasmid in which the Sp1 cDNA is in the antisense orientation (lane 4).

Table 1 Deletion of the intron from pCOL-KT results in moderate expression of the reporter CAT in SL2 cells

SL2 cells were transfected with pCOL-KT, intronless pCOL-KT (pCOL-KTΔI) and three intronless deletion mutants (deleted sequences are shown in parentheses). CAT activity was measured in cell extracts 48 h after transfection and quantified by scraping TLC plates and measuring radioactivity in chloramphenicol and its acetylated products by liquid-scintillation spectroscopy. Deletion of the first intron led to enhanced expression of pCOL-KT in SL2 cells. Four independent determinations were used to derive the average value of CAT in the cells transfected with the control construct, pCOL-KT.

Construct	CAT activity (μg of protein)
pCOL-KT	0.017 ± 0.009 (4)
pCOL-KTΔI	0.22
pCOL-KTΔI (−781 to −561)	0.11
pCOL-KTΔI (−781 to −331)	0.42
pCOL-KTΔI (−781 to −101)	0.32

previously suggested for assessing whether a gene is Sp1-responsive, namely that transcription of the target gene should be increased by at least 10-fold [29,30].

Sp1-mediated activation of the human Pro α 1(I) gene transcription occurs in the absence of intronic sequences

A number of recent studies have implicated the first intron in the regulation of collagen gene transcription; the exact nature of the *cis*-regulatory elements and their contribution towards transcriptional activation have not been defined precisely and are somewhat controversial [5]. The intronic region between +700 and +1300 bp of the human Pro α 1(I) gene contains four potential binding sites for Sp1 [6,7]; two of these are the high-affinity decanucleotide Sp1-binding sites (GCCCCGCCCC) and the other two have the same nucleotide sequences (GGGGCGGGRT) as the 'medium' affinity Sp1-binding sites [29,30]. Recently, Sp1 has been shown to bind to two of these intronic sites (between +919 to +944 bp and between +986 to +1009 bp). Deletion of the intron has previously been shown to decrease expression of a Pro α 1(I) collagen–human growth hormone (hGH) construct [5,6]. However, point mutations introduced into the Sp1-binding sites that no longer competed for binding in EMSA did not decrease transcription of a Pro α 1(I)–hGH construct in chick tendon fibroblasts, but rather resulted in a modest increase (1.5–2-fold) in expression [31]. We assessed the role of the potential Sp1 sites in the intron of the Pro α 1(I) collagen gene by co-transfecting the Sp1-expression vector with pCOL-KTΔI and measuring the expression of CAT in SL2 cells. As shown in Table 1, deletion of all potential Sp1-binding sites located in the intron resulted in an 8–26-fold increase in basal expression of pCOL-KT in SL2 cells even in the absence of co-transfection with pPACSp1. The increased basal transcription of the intronless constructs in SL2 cells could be observed even when 5'-promoter had been shortened to −101 bp (the basal promoter). These data suggest that deletion of the first intron, in addition to removing the four potential Sp1 sites, inadvertently removes other sequences which have a negative regulatory role in SL2 cells. The nature of these additional regulatory elements remains elusive at present. Regardless of the significant improvement of expression of pCOL-KTΔI in SL2 cells, co-transfection of the reporter constructs with pPACSp1 further enhanced their expression by 10–15-fold in various experiments (Figure 3). Consistent with the previous data, co-transfection of pCOL-KT or its variant reporter plasmids with

Figure 3 Intronic sequences are dispensable for Sp1-mediated transcriptional activation of the human Pro α 1(I) collagen gene

The extracts were prepared from SL2 cells 48 h after transfection with intronless pCOL-KT and the Sp1 expression plasmid, pPACSp1, and assayed for CAT activity. [^{14}C]Chloramphenicol and its acetylated products were separated by TLC. A representative autoradiogram is shown. Expression of CAT is increased < 15-fold in SL2 cells co-transfected with intronless pCOL-KT and an Sp1 expression plasmid, pPACSp1 (lane 2), compared with that in SL2 cells transfected with intronless pCOL-KT alone (lane 1).

antisense Sp1 did not significantly change their expression (results not shown).

100 bp of the human Pro α 1(I) collagen promoter sequences are sufficient to confer Sp1 responsiveness

In order to identify a minimal Sp1-responsive human Pro α 1(I) promoter, constructs containing increasing deletions of the 5'-upstream sequences, in addition to the intron, were co-transfected with pPACSp1 and their expression was measured. As shown in Figure 4, progressively larger deletions of the promoter in the

Figure 4 A minimal promoter of 100 bp is sufficient to confer Sp1 responsiveness to the human Pro α 1(I) collagen–CAT construct

SL2 cells were transfected with intronless pCOL-KT constructs containing variable-length deletions of the 5'-upstream sequences. CAT activity in SL2 cell extracts was measured by TLC 48 h after transfection, and detected by fluorography. Deletion of promoter sequences up to −100 bp (lane 1), −330 bp (lane 3) or −560 bp (lane 5) did not affect expression of the reporter CAT in SL2 cells; all truncated promoter constructs failed to be expressed in the SL2 cells. However, co-transfection of each of these constructs with the Sp1 expression plasmid, pPACSp1, showed that the most truncated collagen promoter–CAT construct containing promoter sequences from −100 to +117 bp (lane 2) was most responsive to the action of Sp1, compared with constructs containing sequences either from −330 to +117 bp (lane 4) or from −560 to +117 bp (lane 6).

Figure 5 Recombinant Sp1 protein binds to a minimal human Pro α 1(I) collagen promoter containing 100 bp of 5'-upstream sequences

Sp1 (5 or 15 ng), purified from HeLa cells, was incubated with a radiolabelled (10000 c.p.m.) restriction fragment containing nucleotides -100 to +117 bp of human Pro α 1(I) gene. The reaction mixtures (35 μ l) containing 7.1 mM Hepes (pH 7.8)/3.6 mM MgCl₂/100 mM KCl/5.7% (w/v) glycerol/0.03% (v/v) Nonidet P40 and 40 μ g/ml poly(dI-dC) were incubated at 10 °C for 20 min. Bromophenol Blue (0.05%, v/v; 2 μ l) in 48% glycerol were added to the reactions immediately before loading on a 5% (w/v) native polyacrylamide gel, and electrophoresing in 40 mM Tris/360 mM glycine and 0.1% Nonidet P40 at 4 °C for 1.75 h at 260 V. Lane 1 represents radiolabelled probe without any protein. Lanes 2 and 3 show binding of 5 ng and 15 ng of purified Sp1 to Pro α 1(I) sequences between -100 and +117 bp respectively. Binding can be competed out by 10- (lane 4), 50- (lane 5) or 100-fold (lane 6) molar excess of double-stranded oligonucleotides containing a consensus Sp1 binding site. Binding of purified Sp1 was also inhibited in the presence of 10- (lane 7), 50- (lane 8) or 100-fold (lane 9) molar excess of a double-stranded oligonucleotide containing the wild-type Pro α 1(I) Sp1 binding site or by a 217 bp *EcoRI*-*PstI* restriction fragment of pCOL-KT Δ I (-101 to +117), containing the wild-type Pro α 1(I) Sp1 binding site (lane 10). The major (★) and minor (■) bands representing Sp1-DNA complexes are marked.

intronless pCOL-KT constructs showed increased basal CAT expression when co-transfected with pPACSp1. The construct containing promoter sequences from -100 to +117 bp of the human Pro α 1(I) collagen gene was the most responsive to the action of Sp1 compared with any of the longer constructs (Figure 4). It is important to recall that the highly truncated -100 bp promoter sequence of the human Pro α 1(I) collagen gene contains the TATA box (at -28 to -25 bp) and one putative Sp1-binding motif (at -87 to -82 bp).

Mutation of the Sp1 site in the minimal human Pro α 1(I) collagen promoter (-100 to +117 bp) eliminates binding of Sp1 protein

Using EMSA we found that there was a specific Sp1-binding site in the human Pro α 1(I) collagen promoter between -100 and +117 bp. As shown in Figure 5, purified Sp1 protein (from HeLa cells) binds to a radiolabelled restriction fragment containing the human Pro α 1(I) collagen promoter sequences between -100 and +117 bp forming two distinct protein-DNA complexes. Presumably these complexes represent binding of the two differentially phosphorylated forms of Sp1 that migrate on SDS/polyacrylamide gels as two polypeptide species with molecular masses of 95 and 105 kDa [32,33]. When a higher amount of recombinant Sp1 (15 ng) was included in the binding reactions, a more slowly migrating complex was observed and may reflect aggregation of Sp1 or binding of multiple Sp1 molecules [34]. Sp1 protein is thought to bind to the central CG pair of the consensus recognition site [34] by hydrogen bonding with the 4-amino group of cytosine (a potential hydrogen-bond donor) and the 6-carbonyl group of guanine (a potential hydrogen-bond acceptor). A 10-100-fold molar excess of a

Figure 6 Nuclear extracts from NIH 3T3 cells, but not from SL2 cells, bind to the proximal promoter of the human Pro α 1(I) collagen gene

A radiolabelled (10000 c.p.m.) 217 bp restriction fragment of the human Pro α 1(I) collagen promoter formed defined complexes with NIH 3T3 nuclear proteins (lane 2). Formation of this complex was inhibited in the presence of 10- (lane 3), 50- (lane 4) or 100-fold (lane 5) molar excess of a double-stranded oligonucleotide containing a consensus Sp1 binding site. Nuclear proteins (2 μ g) from SL2 cells did not form similar complexes with the human Pro α 1(I) collagen promoter (lane 6). Radiolabelled probe without protein is shown in lane 1. Reproducible major (★) and minor (■) protein-DNA complexes are denoted.

double-stranded oligonucleotide containing a consensus Sp1 site competed with the human Pro α 1(I) DNA fragment for binding to recombinant Sp1 (Figure 5), thus demonstrating specificity of the binding. Similarly, binding of Sp1 to the Pro α 1(I) minimal promoter could be competed out by a 100-fold molar excess of a restriction fragment or a double-stranded oligonucleotide fragment containing the Pro α 1(I) Sp1-binding site (Figure 5).

In a number of other Sp1-responsive genes, the GC box, located 40-70 bp upstream of the transcriptional initiation site, has been found to be the most important in regulating gene expression [29,30]. This Sp1 site is part of the transcription initiation complex by which RNA polymerase II is activated. To determine whether this is the case for the human Pro α 1(I) gene, we also mutated the GC box at -87 to -82 bp, in two ways: by a central C-to-A conversion and by a GCG-to-AAA conversion. In the SV40 promoter, conversion of GCG into AAA has been shown to prevent binding of Sp1 protein and to inhibit transcription [35]. SL2 cells reportedly do not express Sp1 [36], and as anticipated, nuclear protein from SL2 cells (2 μ g) did not form complexes with the human Pro α 1(I) collagen promoter (Figure 6). Nuclear extracts from NIH 3T3 cells, which do express Sp1, formed complexes with the human Pro α 1(I) collagen promoter (Figure 6). Furthermore, 10-100-fold molar excess of a double-stranded oligonucleotide containing a consensus Sp1-binding site was able to compete out the binding, thus demonstrating its specificity (Figure 6). As shown in Figure 7, double-stranded oligonucleotide fragments containing a G-to-C and a GCG-to-AAA conversion in the Sp1-binding site of the Pro α 1(I) collagen promoter did not compete with the wild-type Sp1 site for binding to purified Sp1 protein or to nuclear proteins from NIH 3T3 cells. Finally, mutation of the potential Sp1-binding site in the human Pro α 1(I) collagen promoter abolished expression of CAT in SL2 cells co-transfected with the Sp1-expression plasmid pPACSp1 compared with wild-type (Figure 8).

Figure 7 Binding of Sp1 to the minimal Pro α 1(I) collagen promoter is abolished by site-specific mutations in the Sp1 binding motif

Single (C₋₈₄ changed to A₋₈₄) or triple (GGC_{-83to-85} changed to AAA_{-83to-85}) mutations were introduced by site-directed mutagenesis in the proximal Sp1 binding motif of the human Pro α 1(I) promoter. Purified Sp1 (15 ng; lane 2) and NIH 3T3 nuclear proteins (5 μ g; lane 9) bind to a 217 bp *Eco*RI–*Pst*I restriction-fragment probe containing the wild-type human Pro α 1(I) Sp1-binding site at –88 to –82 bp. Binding was not inhibited by 10- (lane 3), 50- (lane 4) and 100-fold (lane 5) molar excesses of double-stranded oligonucleotides containing a C-to-A mutation or by 10- (lane 6), 50- (lane 7) or 100-fold (lane 8) excess of double-stranded oligonucleotides containing a GCG to AAA mutation. Similarly, binding of NIH 3T3 nuclear proteins was not reduced in the presence of a 10- (lane 10), 50- (lane 11) or 100-fold (lane 12) excess of a double-stranded oligonucleotide carrying a C-to-A mutation and of a 10- (lane 13), 50- (lane 14) or 100-fold (lane 15) excess of a GCG-to-AAA mutation. Note that to align the data from two different experiments illustrated in (A) and (B) the area of the autoradiograph depicting free probe has been cropped in (B).

DISCUSSION

A variety of *cis*-acting DNA sequences are found within the promoters and enhancers of eukaryotic genes and specific transcription factors interact with these DNA sequences to regulate gene expression [28,37]. There are two broad classes of transcription factors. The tissue-specific transcription factors like Oct-2, which is B-cell-specific, or Myo-D, which is specific for muscle cells, determine cellular phenotype [37]. Other transcription factors are ubiquitous; these include TFIID, which binds to a promoter element (TATA box) located 25–30 bp upstream of the transcription initiation site in genes transcribed by RNA polymerase II [28,37], CTF, which binds to a CCAAT sequence usually 40–100 bp upstream of the start site of transcription [37], and Sp1, which binds the hexanucleotide sequence GGGCGG, the so-called GC box [29]. Sp1-responsive promoters often contain multiple GC boxes: whereas the promoter of the rat insulin-like growth factor II gene contains four proximal GC boxes [38], the human multidrug-resistance promoter contains two Sp1-binding sites [39]. The GC box located 40–70 bp upstream of the RNA start site in Sp1-responsive promoters is usually considered the most important in regulating transcription [28,37].

Sp1 has also been implicated in regulating the transcription of TATA-less housekeeping genes such as the human argininosuccinate synthase gene, the promoter of which contains three GC boxes [31]. Sp1 also appears to have a role in the cell-specific expression of another TATA-less gene, the myeloid integrin CD11b gene [32]. Although nuclear extracts from myeloid (U937) and non-myeloid (HeLa) cells bind the CD11b Sp1 site in EMSA, CD11b is not expressed in HeLa cells. For this reason, it has been suggested that another myeloid-specific factor, PU.1, binds to the site in the CD11b promoter, changing the chromatin structure and enabling Sp1 to bind to its cognate sequence [32]. Sp1 may also have a regulatory function in addition to a general role in the transcription of housekeeping genes. For example, there is

now evidence that Sp1 is expressed preferentially in certain cell types such as T-cells, epithelial cells and haematopoietic cells, suggesting that increased Sp1 levels may be associated with differentiation [33,40]. Since post-translational modifications of transcription factors may also influence their ability to modulate transcription, it should be noted that Sp1 is both glycosylated [41] and phosphorylated [42].

The human Pro α 1(I) gene contains several putative binding sites for Sp1, including two sites in the promoter (up to –806 bp), another four in the distal promoter (up to –2300 bp) and four in the first intron (between +700 and +1300 bp). Two of these intronic sites (from +919 to +944 and from +986 to +992) have been previously shown to bind purified Sp1, and proteins present in nuclear extracts from chick tendon fibroblasts and HeLa cells [34]. However, the precise contribution of the intronic Sp1 is somewhat unclear. Deletion of the intron has previously been shown to inhibit expression of a Pro α 1(I) collagen gene–hGH construct [5,6]. Mutations in the Sp1-binding sites which abolished their affinity for binding in EMSA resulted in 1.5–2-fold increased expression when transfected into chick tendon fibroblasts [34]. In contrast, our data suggest that, at least in the context of expression in SL2 cells, the human Pro α 1(I) gene contains negative regulatory sequences in the intron, quite apart from the four Sp1-binding sites, since basal expression (in the absence of Sp1) was increased 10–20-fold in intronless constructs. The identity of these additional regulatory sequences is not known at present.

An obligatory role for Sp1 in transcriptional regulation of the human Pro α 1(I) collagen gene through a site in the proximal promoter at –87 to –82 bp is demonstrated clearly by our data. Deletion of the second potential Sp1-binding site in the Pro α 1(I) promoter, located at –221 to –216 bp, did not affect the ability of Sp1 to activate transcription. Sp1 has also been shown to bind to a site in the promoter of the mouse Pro α 1(I) collagen gene [5]. It was shown that over-expression of Sp1 in NIH 3T3 cells decreased both basal and NF1-induced expression of a mouse

Figure 8 Mutation of the potential Sp1 site in the minimal human Pro α 1(I) promoter abolishes the Sp1 response of the reporter gene in transient expression assays

Extracts from SL2 cells transfected with pCOL-KT, pCOL-KT Δ I or pCOL-KT Δ II (–100 to +117) plasmids either alone (lanes 1, 3 and 5 respectively) or co-transfected with pPACSp1 (lanes 2, 4 and 6 respectively) were assayed for CAT activity. Constructs containing mutated Sp1 binding site (C_{–84} changed to A_{–84} or GCG_{–8310–85} changed to AAA_{–8310–85}) were also transfected into SL2 cells either alone (lanes 7 and 9 respectively) or with 1.0 mg of pPACSp1 (lanes 8 and 10 respectively). SL2 cells do not express CAT when transfected with the intronless minimal Pro α 1(I) promoter (–100 to +117 bp; lane 1) but co-transfection with the Sp1 expression plasmid, pPACSp1, enables significant expression of CAT (lane 2). This response of the minimal promoter is similar to that occurring with the longer constructs with or without the intron (compare lanes 3 and 4, and 5 and 6). Constructs containing a C_{–84} to A_{–84} mutation (lane 7) or GCG_{–8310–85} to AAA_{–8310–85} mutation (lane 9) in Sp1 motif also failed to be expressed in SL2 cells even when co-transfected with the Sp1 expression plasmid, pPACSp1 (lanes 8 and 10 respectively). The extracts from SL-2 cells transfected with pPACSp1 (lane 11) or pUC19 DNA (lane 12) or mock-transfected (lane 13) cells were also negative for CAT activity.

Pro α 1(I)–luciferase construct. Based on these data, it has been suggested that the binding of Sp1 inhibited binding of NF1 to the promoter, thereby attenuating the ability of NF1 to activate transcription. Karsenty and de Crombrughe [18] have described another inhibitory transcription factor, IF2, distinct from Sp1, which binds to the analogous region of the mouse Pro α 1(I) promoter. Interestingly, our results suggest that the purified Sp1 was unable to bind to this sequence.

At present the mechanism by which Sp1 regulates Pro α 1(I) collagen gene transcription in SL2 cells is unclear. In the mouse, deletion of the GC boxes results in the loss of 85% of the promoter activity [18]. We have introduced point mutations into the Sp1 site in the proximal human Pro α 1(I) promoter and have shown that these mutations abolish Sp1 binding *in vitro*. Furthermore, mutation of Sp1 in the minimal promoter significantly reduces transcriptional activity of the reporter gene. There is considerable heterogeneity in the known Sp1-binding sites in the Pro α 1(I) collagen gene; the two previously identified sites located in tandem in the first intron and the one found in the distal promoter were not identical [34]. Although the precise mechanisms of Sp1-binding and the role of other accessory molecules involved in this process are currently unknown, it is possible that the over-expression of Sp1 protein in SL2 cells after transfection with pPACSp1 enables Sp1 to bind to degenerate sites. This has been demonstrated in CV1 cells transfected with plasmids in which CAT expression was driven by the human β -globin or Rous sarcoma virus long terminal repeat promoters (which do not contain consensus Sp1 sites but have some GC-rich sequences); over-expression of Sp1 in CV1 cells did indeed cause enhanced transcription from these promoters [40]. It would therefore be of interest to extend these studies to a cell line that does express Sp1, such as NIH 3T3 cells, and to determine whether mutation of the Sp1-binding site in the proximal promoter of the human Pro α 1(I) gene ameliorates expression of our CAT constructs in the absence of over-expression of Sp1.

This work was supported by a Specialized Center of Research grant P50AR 39166 from the National Institute of Arthritis and Musculoskeletal Diseases (NIH) and by a Department of Veterans Affairs (DVA) Merit Review Grant. We thank Dr. Carl Wu of N.I.H., Bethesda, MD, U.S.A. for the *Drosophila* SL2 cells and Dr. Robert Tjian of University of California Berkeley, CA, U.S.A. for Sp1 constructs. Dr. Jane D. Stephenson kindly provided us with the intronless collagen promoter–CAT construct used in these studies. Competent secretarial help of Donna Jackett is also gratefully acknowledged. R.R. is a Research Career Scientist of the DVA.

REFERENCES

- Prockop, D. J. and Kivirikko, K. I. (1995) *Annu. Rev. Biochem.* **64**, 403–434
- Vuorio, E. and de Crombrughe, B. (1990) *Annu. Rev. Biochem.* **59**, 837–872
- Slack, J. L., Liska, D. J. and Bornstein, P. (1993) *Am. J. Med. Genet.* **45**, 140–151
- Raghow, R. and Thompson, J. P. (1989) *Mol. Cell. Biochem.* **86**, 5–18
- Bornstein, P. (1996) *Matrix Biol.* **15**, 1–6
- Bornstein, P., McKay, J., Morishima, J. K., Devarayalu, S. and Gelinas, R. E. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8869–8873
- Sherwood, A. L. and Bornstein, P. (1990) *Biochem. J.* **265**, 895–897
- Boast, S., Su, M. W., Ramirez, F., Sanchez, M. and Avvedimento, E. V. (1990) *J. Biol. Chem.* **265**, 13351–13356
- Simkevich, C. P., Thompson, J. P., Poppleton, H. and Raghow, R. (1992) *Biochem. J.* **286**, 179–185
- Wang, Q. and Raghow, R. (1996) *Mol. Cell. Biochem.* **158**, 33–42
- Olsen, A. S., Geddis, A. E. and Prockop, D. J. (1991) *J. Biol. Chem.* **266**, 1117–1121
- Niederreither, K., D'Souza, R. N. and de Crombrughe, B. (1992) *J. Cell Biol.* **119**, 1361–1370
- Goldberg, H., Helaakoski, T., Garrett, L. E., Karsenty, G., Pellegrino, A., Lozano, G., Maity, S. and de Crombrughe, B. (1992) *J. Biol. Chem.* **267**, 19622–19630
- Slack, J. L., Liska, D. J. and Bornstein, P. (1991) *Mol. Cell. Biol.* **11**, 2066–2074
- Ritzenthaler, J. D., Goldstein, R. H., Fine, A. and Smith, B. D. (1993) *J. Biol. Chem.* **268**, 13625–13631
- Ritzenthaler, J. D., Goldstein, R. H., Fine, A., Lichtler, A., Rowe, D. W. and Smith, B. D. (1991) *Biochem. J.* **280**, 157–162
- Schneider, I. (1972) *J. Embryol. Exp. Morphol.* **27**, 353–365
- Karsenty, G. and de Crombrughe, B. (1990) *J. Biol. Chem.* **265**, 9934–9942
- Courey, A. J. and Tjian, R. (1988) *Cell* **55**, 887–898
- Thompson, J. P., Simkevich, C. P., Holness, M. A., Kang, A. H. and Raghow, R. (1991) *J. Biol. Chem.* **266**, 2549–2556
- Heiermann, R. and Pongs, O. (1985) *Nucleic Acids Res.* **13**, 2709–2730
- Ahearn, J. M., Bartolomei, M. S., West, M. L., Cisek, L. J. and Corden, J. L. (1987) *J. Biol. Chem.* **262**, 10695–10705
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1988/1994) *Curr. Protocols Mol. Biol.* **1**, 9.6.2–9.6.4
- Letowsky, J. and Dynan, W. S. (1989) *Nucleic Acids Res.* **17**, 2639–2653
- Katai, H., Stephenson, J. D., Simkevich, C. P., Thompson, J. P. and Raghow, R. (1987) *Mol. Cell. Biochem.* **118**, 119–129
- Santoro, C., Mermod, N., Andrews, P. C. and Tjian, R. (1988) *Nature (London)* **334**, 218–224
- Heberlein, U., England, B. and Tjian, R. (1985) *Cell* **41**, 965–977
- Goodrich, J. A., Cutler, G. and Tjian, R. (1996) *Cell* **84**, 825–830
- Kadonaga, J. T., Jones, K. A. and Tjian, R. (1986) *Trends Biochem. Sci.* **11**, 20–23
- Kadonaga, J. T., Carner, K. R., Masiarz, F. R. and Tjian, R. (1987) *Cell* **51**, 1079–1090
- Anderson, G. M. and Freytag, S. O. (1991) *Mol. Cell. Biol.* **11**, 1935–1942
- Chen, H.-M., Pahl, H. L., Scheibe, R. J., Zhang, D.-E. and Tenen, D. G. (1993) *J. Biol. Chem.* **268**, 8230–8239
- Saffer, J. D., Jackson, S. P. and Annarella, M. B. (1991) *Mol. Cell. Biol.* **11**, 2189–2199
- Liska, D. J., Robinson, V. R. and Bornstein, P. (1992) *Gene Exp.* **2**, 379–389
- Gidoni, D., Kadonaga, J. T., Barrera-Saldana, H., Takahashi, K., Chambon, P. and Tjian, R. (1985) *Science* **230**, 511–517
- Reue, K., Leff, T. and Breslow, J. L. (1988) *J. Biol. Chem.* **263**, 6857–6864
- Zewel, L. and Reinberg, D. (1995) *Annu. Rev. Biochem.* **64**, 533–561
- Evans, T., DeChiara, T. and Efstratiadis, A. (1988) *J. Mol. Biol.* **199**, 61–81
- Cornwell, M. M. and Smith, D. E. (1993) *J. Biol. Chem.* **268**, 19505–19511
- Saffer, J. D., Jackson, S. P. and Thurston, S. J. (1990) *Genes Dev.* **4**, 659–666
- Jackson, S. P. and Tjian, R. (1988) *Cell* **55**, 125–133
- Jackson, S. P., MacDonald, J. J., Lees-Miller, S. and Tjian, R. (1990) *Cell* **63**, 155–165