A highly conserved intronic sequence is involved in transcriptional regulation of the $\alpha 1(I)$ collagen gene

DeAnn J. Liska, James L. Slack,
and Paul Bornstein

Departments of Biochemistry and Medicine
University of Washington
Seattle, Washington 98195

The first intron of the human $\alpha 1(I)$ collagen gene contains a positive, orientation-dependent *cis*-acting sequence located between bases +292 and +670. Transient transfection experiments indicate that this sequence is functional in both primary chicken tendon fibroblasts and in a human fibroblast-like cell line derived from SV40-transformed marrow stromal cells. DNase I footprint, methylation interference, and mobility shift analyses provide evidence for a sequence-specific binding activity and show that the region of binding corresponds to a 29-base-pair sequence that is also present in the rat $\alpha 1(I)$ collagen intron. This conserved sequence contains an AP1 consensus motif. Sequence-specific binding activity is present in nuclear extracts from HeLa and fibroblast cell lines but not in extracts from two lymphoid cell lines. Mutation of the AP1 consensus sequence indicates that this motif is required for function of the *cis*-acting element. These data indicate that transcriptional modulation of the $\alpha 1(I)$ collagen gene involves an interaction between an intronic AP1-containing sequence and its cognate transcription factors.

Introduction

The collagens are a family of proteins that function both as structural components of tissues and in morphogenetic processes. The precise transcriptional regulation of the genes coding for these proteins is therefore crucial for the determination and maintenance of tissue structure (Bornstein and Sage, 1989; Raghow and Thompson, 1989). A common feature of these genes is the presence of regulatory elements in the first intron. The genes that encode the two polypeptides of type I collagen, $\alpha 1(I)$ and $\alpha 2(I)$, both contain intronic sequences that increase transcriptional activity when placed in the correct position relative to their promoters (Bornstein *et al.*, 1987; Rossi and de Crom-

brugghe, 1987; Rossouw *et al.*, 1987; Rippe *et al.*, 1989). The $\alpha 1(II)$ collagen gene contains an intronic enhancer that is involved in the cell-specific expression of that gene (Horton *et al.*, 1987), and the $\alpha 1(IV)$ gene contains an enhancer in its first intron that may regulate both the $\alpha 1(IV)$ and $\alpha 2(IV)$ genes (Burbelo *et al.*, 1988; Killen *et al.*, 1988). How these intronic sequences are involved in the transcriptional regulation of these genes has not been defined.

Intronic *cis*-acting sequences are not unique to the collagen genes. The immunoglobulin μ heavy chain gene (Banerji *et al.*, 1983; Gillies *et al.*, 1983; Mercola *et al.*, 1983), muscle creatine kinase gene (Sternberg *et al.*, 1988), and the $\delta 1$ -crystallin gene (Hayashi *et al.*, 1987) contain intronic regulatory sequences. Recent studies suggest that intronic enhancer or cell-specific *cis*-acting sequences also exist in the genes for β -actin (Ng *et al.*, 1989), $\alpha_{2\mu}$ -globulin (Choy *et al.*, 1989), and β -globin (Galson and Housman, 1988). One mechanism by which these regulatory sequences could influence transcription involves an interaction between the *cis*-acting sequences in the intron and in the promoter of the genes. This interaction could be mediated by the *trans*-acting factors that bind to these sequences (Dyner and Tjian, 1985; Müller *et al.*, 1988).

Our studies have focused on the role of the first intron in the transcriptional regulation of the human $\alpha 1(I)$ collagen gene. We have utilized a human growth hormone (hGH)¹ fusion gene that contains the human $\alpha 1(I)$ collagen gene sequence from –804 to +1440 in the first intron. No significant reduction in transcription was seen in the presence of the first intron when a deletion was made in the promoter between bases –625 and –161 (Bornstein *et al.*, 1988). Deletion of the intron, however, resulted in a 4- to 10-fold reduction in transcription, as determined by a transient transfection assay in

¹ Abbreviations used: CTF, chicken tendon fibroblasts; DMEM, Dulbecco's modified Eagle's medium; hGH, human growth hormone; SV-MSC, SV40-transformed marrow stromal cells; TGF- β , transforming growth factor- β ; TPA, 12-O-tetradecanoylphorbol-13-acetate.

chicken tendon fibroblasts (CTF) (Bornstein and McKay, 1988; Bornstein *et al.*, 1988).

Several consensus sequences for known transcription factors are present in the $\alpha 1(I)$ intron. These include SP1 and viral core enhancer motifs (Bornstein *et al.*, 1987; Rossouw *et al.*, 1987), which are contained within a 274-bp segment, A274, in the 3' half of the intron. However, it was clear from previous studies (Bornstein and McKay, 1988; Bornstein *et al.*, 1988) that the transcriptional enhancement by the collagen $\alpha 1(I)$ intron could not be attributed to sequences present in the A274 element alone.

During our investigation of the intronic regulatory element A274, we also obtained preliminary evidence that a positively acting element was located in the 5' region of the intron, between bases +292 and +670 (Bornstein *et al.*, 1988). In this study we show that this region of the intron contains an orientation-dependent, positively acting element that functions in both CTF and in a human fibroblast-like cell, SV-MSC. We extend these observations to show that, in contrast to the A274 element, this activity is not dependent on the presence of upstream promoter sequences. This *cis*-acting sequence contains a protein-binding site that is conserved in the rat and human genes and includes an AP1 consensus motif. Furthermore, site-directed mutagenesis of this motif abolishes the transcriptional enhancement observed with the intronic sequence +292 to +670. It therefore appears that the $\alpha 1(I)$ collagen intron modulates transcription through different mechanisms, one of which may involve an interaction between an AP1-containing sequence and its *trans*-acting factors.

Results

The first intron of the human $\alpha 1(I)$ collagen gene contains a positive, orientation-dependent element

Preliminary observations suggested that an intronic segment of the $\alpha 1(I)$ gene, containing nucleotides +292 through +670, enhanced transcription in CTF when placed in its natural position 3' to the collagen promoter (Bornstein *et al.*, 1988). We wished to validate these results with the use of human cells and thus utilized a human marrow stromal cell line, SV-MSC, to further define the transcriptional activity of constructs containing this region of the collagen intron. SV-MSC have characteristics of both fibroblasts and smooth muscle cells (Singer *et al.*, 1987) and actively transcribe the $\alpha 1(I)$ and $\alpha 2(I)$ collagen genes (Andrews *et al.*, 1989). As

shown in Figures 1 and 2A, the transfection of SV-MSC revealed relative levels of transcription nearly identical to those previously reported for the transfection of CTF with the constructs pCol-hGH, pCOL(Δ P)hGH, and pCol(Δ I)hGH (Bornstein *et al.*, 1988). That is, the inclusion of the intron resulted in a 3- to 4-fold increase in transcriptional efficiency, and deletion of promoter sequences between bases -625 and -161 showed very little effect on the level of hGH mRNA compared with the parent plasmid, pCol-hGH. In both SV-MSC and CTF, inclusion of the intronic fragment, +292 to +670, restored nearly all of the enhancement provided by the full intron. Within each experiment the results were consistent: the plasmids containing either the full intron or the positively oriented intronic fragment always showed an enhanced level of hGH mRNA compared with the intron-deleted plasmid. However, the level of enhancement varied from experiment to experiment and ranged from 2- to 5-fold with SV-MSC and 4- to 10-fold with CTF.

Different intronic elements modulate $\alpha 1(I)$ transcription by disparate mechanisms

Recently, we reported that the collagen intronic segment A274 (+820 to +1093) was inhibitory to transcription only when the 5' promoter region from -477 to -255 was present. Furthermore, mobility shift assays suggested that A274 and this promoter region bound the same protein complex (Bornstein *et al.*, 1988). We proposed that a complex between the promoter and the intron was responsible for the orientation-dependent inhibition seen with this fragment. In a set of experiments similar to those performed with A274, we asked whether the orientation-dependent enhancement seen with the intronic region +292 to +670 was also dependent on regions of the promoter. The intronic sequence +292 to +670 was transferred to plasmids in which both the upstream promoter region from -625 to -161 and the intron had been deleted, yielding the constructs pCol(Δ P;l:292-670)hGH and pCol(Δ P;l:670-292)hGH. These plasmids were assayed for transcriptional activity and compared with the plasmids containing the intact promoter sequence. The results of experiments with SV-MSC are summarized in Figures 1 and 2A. Similar results were obtained in transient transfection analysis with CTF (Figure 2B). In SV-MSC and CTF the plasmid containing the positively oriented intronic fragment was transcribed three- to five-fold more effectively than the

Figure 1. Diagrammatic representation of pCol-hGH-derived plasmids and tabulation of relative transcriptional activity. The open rectangle and heavy line designate the first collagen exon and the hGH minigene, respectively. The start of transcription at base +1 is indicated by the arrow. Unique *Sst* II sites (S) in the collagen intron are indicated. The orientation of insertion of the intronic fragment between the *Sst* II sites at +292 and +670 is indicated by the open arrow. Site-specific mutation in the AP1 motif is indicated by the solid triangle in plasmids 8, 9, and 10. Plasmids pCol(ΔP)hGH and pCol(ΔI)hGH were previously identified as pCol(Δ-625 to -161)hGH and pCol-hGHΔ292-1440, respectively (Bornstein *et al.*, 1988). The transcriptional activity of each plasmid in SV-MSC and CTF, relative to pCol(ΔI)hGH ± SEM and corrected for transfectional efficiency, is tabulated. The number of independent determinations is given in parentheses.

plasmid containing the negatively oriented fragment. When the two plasmids containing the promoter deletion were compared, the same orientation-dependent enhancement was observed. These data indicate that the promoter region between -625 and -161 is not necessary for the orientation-dependent transcriptional enhancement by this intronic sequence. These findings are in contrast to those observed with the intronic element A274 and suggest that the two regions of the intron function differently in regulating transcription of the $\alpha 1(I)$ collagen gene.

The intronic sequence +292 to +670 contains an AP1 motif that interacts with nuclear proteins

To investigate whether nuclear extracts contained a binding activity that interacted with this region of the $\alpha 1(I)$ intron, we performed a gel mobility shift assay with the fragments +292 to +474 and +475 to +670. No significant DNA-binding activity for the 5' fragment, +292 to

+474, was observed in nuclear extracts prepared from CTF (data not shown). However, with the use of either CTF or HeLa nuclear extracts, a major shifted band was observed with the 3' fragment, +475 to +670 (Figure 3, A and B, lane 2). Competition analysis indicated that this band resulted from a sequence-specific DNA-binding activity present in these extracts (Figure 3). Similar results were obtained with nuclear extracts prepared from L-cells (data not shown).

DNase protection analysis using nuclear extracts from CTF and HeLa cells was utilized to delineate the binding region. As shown in Figure 4, the fragmentation pattern clearly shows a protected region that, when compared with the co-electrophoresed products from a G + A chemical cleavage reaction, was located between bases +589 and +614 on the sense strand and between bases +596 and +618 on the antisense strand. The same protection pattern was observed with nuclear extracts from SV-MSC and L-cells (data not shown), but not with a nuclear extract prepared from Manca cells (Figure 4, see below).

Figure 2. Analysis of transcriptional efficiency of pCol-hGH-derived plasmids by transient transfection of SV-MSK (A) and CTF (B). Cultured cells were transfected with pCol-hGH (lanes 1 and 2), pCol(Δ I)hGH (lanes 3 and 4), pCol(I:292–670)hGH (lanes 5 and 6), pCol(I:670–292)hGH (lanes 7 and 8), pCol(Δ P;I:292–670)hGH (lanes 9 and 10), or pCol(Δ P;I:670–292)hGH (lanes 11 and 12). Lane 13: hGH and β -globin riboprobes. Total RNA (10 μ g) was prepared from the transfected cells, hybridized to a 32 P-labeled antisense riboprobe, and analyzed for hGH mRNA. The protected band provided by hGH mRNA is indicated by the solid arrow and that of the control β -globin mRNA by the open arrow. The average relative hGH mRNA levels for the experiment shown in B, corrected for efficiency of transfection, are: lanes 1 and 2, 3.3; lanes 3 and 4, 1.0; lanes 5 and 6, 8.0; lanes 7 and 8, 1.2; lanes 9 and 10, 4.6; lanes 11 and 12, 0.72.

As shown in Figure 5, the region of protection defined by DNase I cleavage corresponds to a stretch of 29 bp that is also present, with only two substitutions, in the first intron of the rat α 1(I) collagen gene (+545 to +573). Furthermore, this conserved sequence contains a dyad symmetry that shows homology to the AP1 consensus sequence (Angel *et al.*, 1987; Lee *et al.*, 1987). To investigate this binding activity further, we prepared a double-stranded oligonucleotide, designated S1, that spanned the

protected region. As shown in Figure 6, gel mobility shift analysis of S1 with nuclear extracts prepared from CTF and HeLa cells showed a single shifted band. This binding activity was specific for S1 because two unrelated oligonucleotides were unable to compete for binding. The binding activity was also not competed for by either the coding or noncoding single-stranded chains of S1, nor was any specific binding detected to the single strands (data not shown), indicating that the protein(s) binds only to double-stranded DNA. Competition analysis showed that S1 was able to compete specifically for binding to the parent fragment, +475 to +670 (Figure 3B, lanes 5, 6, and 7).

These observations indicate that the major shifted band seen with the +475 to +670 fragment involves an interaction with the bases represented by S1. To verify that these bases could account for the observed *in vitro* binding activity, we performed methylation interference analysis of the fragment. As shown in Figure 7, the protected bases also lie within the protected region observed by DNase footprint analysis. These data also indicate that the AP1 consensus sequence is involved in this binding activity. No specific protection was observed with the lower migrating band seen with the HeLa nuclear extract (see Figure 3B; data not shown), suggesting that this band results from nonspecific protein-DNA interactions. We therefore conclude that the major shifted band observed after incubation of these nuclear extracts with the +475 to +670 fragment is due to a specific interaction with the sequence represented by S1.

Methylation interference experiments indicated that the AP1-like sequence, TGATTCA, in the S1 fragment is involved in the interaction of the DNA with nuclear proteins. However, the +292 to +474 fragment also contains an AP1-like sequence, TGAGACA, at +370. This sequence does not compete for binding to the +475 to +670 fragment (Figure 3, A and B, lane 3), nor was there any evidence for *in vitro* binding to the +370 AP1-like site.

As seen in Figure 4, DNase protection analysis with a nuclear extract prepared from a B-cell lymphoma line (Manca) did not show protein binding to the +475 to +670 fragment. This suggested to us that the S1-binding activity may have a limited cellular distribution. We therefore compared S1-binding activity in nuclear extracts prepared from lymphoid cell lines with that in extracts from fibroblasts and HeLa cells. As seen in Figure 8, the extracts from the B-cell line (Manca) and a leukemic T-cell line (Jurkat)

Figure 3. Gel mobility shift analysis for binding activity to the intronic fragment +475 to +670 in CTF (A) and HeLa (B) nuclear extracts. The ^{32}P -labeled intronic fragment +475 to +670 was incubated with (A) 24 μg CTF nuclear extract or (B) 12 μg HeLa nuclear extract and examined in a gel mobility shift assay. (A) Lane 1, no nuclear extract; lane 2, no added competitor; lane 3, 50-fold molar excess of unlabeled intronic fragment +292 to +474; lane 4, 50-fold molar excess of unlabeled intronic fragment +475 to +670; lane 5, 50-fold molar excess of oligonucleotide S1; lane 6, 50-fold molar excess of oligonucleotide A1; (B) lane 1, no nuclear extract; lane 2, no added competitor; lane 3, 50-fold molar excess of unlabeled intronic fragment +292 to +474; lane 4, 50-fold molar excess of intronic fragment +475 to +670; lane 5, 50-fold molar excess of oligonucleotide A1; lane 6, 50-fold molar excess of oligonucleotide S1; lane 7, 50-fold molar excess of oligonucleotide A3.

do not contain S1-binding activity. (The lower band in lanes 2–6 is due to nonspecific activity.) The lack of S1 binding activity was not due to the quality of the lymphoid extracts, because the same extracts were used in binding assays with other oligonucleotide probes (A1 and A3) and showed DNA-binding using one-third the amount of protein that was used in this assay (data not shown). These findings suggest that the protein(s) that binds to S1 is present in higher concentrations in cells that are able to transcribe collagen. However, there was no good correlation with the extent of collagen synthesis because this activity is abundant in crude nuclear extracts from HeLa and L cells, which transcribe collagen less actively than do CTF or SV-MSC.

A mutation in the AP1 motif abolishes both protein binding to S1 and transcriptional activity of the +292 to +670 fragment

The AP1-like sequence in S1 shows a six out of seven base pair match with the reported consensus sequence TGA(C/G)TCA (Angel *et al.*, 1987; Lee *et al.*, 1987). In S1, the middle base of the palindrome is a T instead of C or G. A recent report has indicated that this alteration of the AP1 consensus results in <25% binding of a *c-jun/c-fos* complex (Risse *et al.*, 1989). To investigate the interaction of this nucleotide with the S1-binding activity, we prepared an oligonucleotide containing a T to C mutation in this central base, designated M2-AP1. As

shown in Figure 9, the oligonucleotide M2-AP1 competed for S1-binding activity as well as native S1. Mobility shift analyses with M2-AP1 and HeLa or CTF nuclear extracts produced identical results to those obtained with S1 (unpublished observations). However, a mutation of four bases in the AP1 consensus, designated M1-AP1 in Figure 9B, abolished the ability of the oligonucleotide to interact with nuclear proteins (unpublished observations). M1-AP1 was not able to compete for the S1-binding activity in HeLa extracts, even at a 200-fold excess of oligonucleotide (Figure 9A). These data indicate that the AP1 consensus motif is required for the interaction of nuclear proteins with S1.

The intronic sequence, +292 to +670, containing the base alterations in M1-AP1, was prepared by oligonucleotide-based, site-directed mutagenesis. This sequence was then transferred into collagen promoter-containing plasmids, and the constructs were tested in transient transfection assays. A summary of these experiments is shown in Figure 1. The increase in hGH mRNA levels attributable to the +292 to +670 sequence was not seen with the mutated sequence (Figure 1, compare lines 4 and 9). Similarly, when the mutated sequence was incorporated into the intact intron, no increase in hGH mRNA was observed (Figure 1, compare lines 1 and 8). These data indicate that the integrity of the AP1 site is essential for the function of the *cis*-acting elements contained in the sequence +292 to +670, either alone or in the context of the full intron. The data further

Figure 4. DNase I cleavage pattern on the coding (A) and noncoding (B) strand of the intronic fragment +475 to +670. The ^{32}P end-labeled fragment was analyzed by DNase I protection. (A) Lane 1, G + A chemical cleavage sequence ladder; lane 2, naked DNA, 1 ng DNase; lane 3, 40 μg CTF nuclear extract, 25 ng DNase; lane 4, 60 μg SV-MSC nuclear extract, 75 ng DNase; lane 5, naked DNA, 2 ng DNase; lane 6, 34 μg Manca nuclear extract, 25 ng DNase; lane 7, 34 μg Manca nuclear extract, 75 ng DNase. Molecular weight markers from an *MspI* digest of pBR322 are shown to the left. (B) Lanes 1, 5, and 7, naked DNA with 1 ng DNase; lanes 2 and 3, 40 μg CTF nuclear extract, 25 ng and 50 ng of DNase, respectively; lane 4, naked DNA with 2 ng DNase; lane 6, 40 μg HeLa nuclear extract, 150 ng DNase. Protected bases extending from 5' (bottom) to 3' (top) are indicated. Coding and noncoding refer to sense and antisense strands, respectively.

suggest that the DNA sequence within S1 may be directly responsible for the increase in transcriptional activity of intron-containing plasmids.

Discussion

Several studies have suggested that the first intron of the collagen $\alpha 1(\text{I})$ gene is intricately involved in the transcription of that gene, but the mechanisms responsible for such regulation are not known. In this investigation we have located a protein-binding site, S1, in the first intron of the human $\alpha 1(\text{I})$ gene. This stretch of 29 bp is highly conserved (27/29 bases are identical when the rat and human $\alpha 1(\text{I})$ introns are compared) and contains an AP1 consensus sequence. Mutation of the AP1 motif abolishes both the *in vitro* protein-binding activity and the positive, orientation-dependent, *cis*-acting ac-

tivity of the intron in transient transfection assays in both human SV-MSC and CTF. These data suggest that the S1 element is involved in the regulation of collagen gene transcription.

Several lines of evidence support the conclusion that the S1 element affects collagen mRNA production at the level of transcription and not at a posttranscriptional step such as RNA processing. First, the S1-binding activity was only observed with double-stranded DNA, suggesting that the proteins involved do not bind to an RNA transcript. Second, when a fragment of the $\alpha 1(\text{I})$ intron containing the S1-binding site was cloned 5' to the collagen promoter, an increase in promoter activity was observed (Bornstein and McKay, 1988). The effect of the S1-containing fragment on collagen promoter activity was modest, approximately two- to fivefold. In the present study, this enhancement was shown to be reproducible using two different

Figure 5. Comparison of human and rat $\alpha 1(I)$ collagen gene intronic sequences. The region of the human $\alpha 1(I)$ collagen gene (H) between bases +447 and +679, relative to the start of transcription at base 1, was aligned with a corresponding region from the first intron of the rat $\alpha 1(I)$ gene (R), bases +407 to +640. The closed and open triangles represent the *Bss*HI and *Sst* II restriction endonuclease sites, respectively. The bar represents the sequence used for the preparation of the oligonucleotide probe, S1. The degree of sequence identity (indicated by asterisks) is 70%.

collagen-producing cell lines. These observations suggest that the effect of the $\alpha 1(I)$ intron on collagen transcription is potentially of physiological importance.

Transfection data (Figures 1 and 2) demonstrate that the activity of the *cis*-acting sequence, +292 to +670, is orientation-dependent. We have previously reported studies with a more 3' intronic *cis*-acting sequence, A274, whose activity was also orientation dependent (Bornstein *et al.*, 1987, 1988). The A274 fragment was inhibitory to transcription when placed in the negative orientation, and this inhibition was alleviated when deletions were made in the promoter between bases -477 and -255. We also observed that a fragment of the promoter containing these bases could compete for the binding of nuclear proteins to the A274 fragment, suggesting that both regions of DNA are bound in the same complex. To account for these observations, we suggested that the intron and promoter interact with each other via DNA-binding proteins (Bornstein *et al.*, 1988). Sanders and McKnight (1988) have proposed a similar mechanism of gene regulation to explain the orientation dependence of steroid inducibility of the chicken ovalbumin gene by the steroid-responsive element.

Because the +292 to +670 fragment also displayed an orientation dependence in transfection, we asked whether deletion of sequences between -625 and -161 in the collagen promoter would affect the orientation-dependent enhancement observed with this intronic frag-

ment. Deletion of the upstream promoter region did not abolish the transcriptional effect (Figures 1 and 2). In addition, there was no evidence that this promoter sequence and S1 interacted in a common nuclear protein complex (data not shown). However, we cannot exclude the possibility that an interaction occurs between the sequence represented by S1 and the proximal collagen promoter and/or the first exon, i.e., between bases -161 and +222. These data are in contrast to those previously reported for the A274 region and suggest that the regulatory effects of different parts of the intron in the $\alpha 1(I)$ collagen gene are mediated differently.

The conservation of the intronic S1 sequence between the rat and human introns suggests selection of this region for some function in the regulation of collagen transcription. Furthermore, a comparison of S1-binding activity in a number of nuclear extracts indicates that the proteins that bind to the S1 sequence are present in cells that have the ability to express collagen but not in lymphoid cell lines, which do not express collagen. That the factor is present in cells that are competent for collagen transcription suggests that this factor could be important in the tissue-specific expression of this gene.

There are other recent examples of interactions between the promoter and intronic elements that affect transcriptional regulation. Yutzey *et al.* (1989) have examined the regulation of transcription of the quail troponin I gene, TnI, by transient transfection into a mouse

Figure 6. Gel mobility shift analysis for S1-binding activity in CTF (A) and HeLa (B) nuclear extracts. (A) 24 μ g CTF nuclear extract or (B) 10 μ g HeLa nuclear extract was incubated with 32 P-labeled S1. (A and B) Lane 1, S1 probe with no nuclear extract; lane 2, no added competitor; lane 3, 50-fold molar excess of the oligonucleotide A1; lane 4, 50-fold molar excess of A3; lane 5, 50-fold molar excess of S1.

myogenic cell line. Their results indicate that the 5'-flanking region and the first exon are necessary, but not sufficient, to direct the correct transcriptional activation of the Tnl gene during myogenesis. From their studies the major regulatory region controlling the differential transcription pattern of the quail Tnl gene during skeletal myogenesis appears to be located within a 148-bp region of the Tnl first intron. Hallauer *et al.* (1988) produced transgenic mice carrying the quail Tnl promoter and first intron and found the same fiber-type specific expression pattern as the endogenous mouse Tnl gene, indicating that, although avian and mammalian muscle fibers show morphological and physiological differences, a homologous gene regulatory mechanism exists at the level of transcription. In our work, transient transfection of a human cell line, SV-MSC, with various collagen constructs resulted in similar relative lev-

els of transcription as in transfection of an avian cell line (CTF). These observations suggest that the mechanisms involved in the intronic enhancement of collagen transcription may be similarly conserved between species.

The presence of an AP1 site in the $\alpha 1(I)$ collagen gene raises many questions with respect to the regulation of the gene. The AP1 sequence is a recognition site for the *jun* family of *trans*-acting factors, which include *jun* B, *jun* D, and *c-jun* (Angel *et al.*, 1987; Lee *et al.*, 1987; Ryder *et al.*, 1988; Hirai *et al.*, 1989). These polypeptides form homodimers and, with the *fos* family of *trans*-acting proteins, heterodimers (Sassone-Corsi *et al.*, 1988). The *jun* family of genes appears to be regulated differently in different tissues and may mediate both transcriptional activation (Angel *et al.*, 1987; Lee *et al.*, 1987; Kerr *et al.*, 1988) and repression (Takimoto *et al.*, 1989). Recent evidence suggests that the various members of the *jun* family differ in their biological activities (Chiu *et al.*, 1989; Schütte *et al.*, 1989). In addition, Pertovaara *et al.* (1989) and Li *et al.* (1990) have reported that transforming growth factor- β (TGF- β) differentially activates members of the *jun* protein family. These observations suggest that members of the *fos* and *jun* families may play distinct roles during differentiation and at the sites of wounding. Type I collagen genes are, in part, regulated transcriptionally by TGF- β (Penttinen *et al.*, 1988; Rossi *et al.*, 1988). It is possible that the S1 sequence and a corresponding *trans*-acting factor may be involved in the regulation of the $\alpha 1(I)$ gene in developing tissue and in response to TGF- β . Preliminary experiments indicate that a member of the *fos* family represents one factor in the S1-binding activity observed in nuclear extracts (D.J. Liska and P. Bornstein, in preparation).

The AP1 site has been shown to modulate the expression of 12-O-tetradecanoylphorbol-13-acetate (TPA)-inducible genes (Angel *et al.*, 1987; Lee *et al.*, 1987). The presence of a functional AP1 site in a collagen gene is particularly interesting since the expression of type I collagen genes is not increased after TPA treatment but, instead, is inhibited both at the level of protein and mRNA (Delclos and Blumberg, 1979; Dion *et al.*, 1982; Sobel *et al.*, 1983). Whether the AP1 site described in this work is involved in this regulation will be important to determine. Certainly one would expect that this site, in the context of the collagen gene, would not function as a TPA-inducible enhancer. Experiments are currently in progress to address this issue.

Figure 7. Methylation interference analysis of the +475 to +670 fragment from the $\alpha 1(I)$ collagen intron. The intronic fragment, +475 to +670, was ^{32}P end-labeled on the coding (A) or non-coding (B) strand and then subjected to partial methylation and depurination. After incubation with nuclear extract and gel mobility shift analysis, the bands corresponding to the protein-bound (B) and uncomplexed (F) probe were excised and processed separately. (A) Lanes 1 and 3, uncomplexed probe; lane 2, complexed probe resulting from incubation with HeLa nuclear extract; (B) Lanes 1, 3, 4, and 6, uncomplexed probe; lane 2, complexed probe resulting from incubation with HeLa nuclear extract; lane 5, complexed probe resulting from incubation with L cell nuclear extract. In each case, the region of protection by DNase analysis is also shown. (C) Sequence of S1. Arrowheads indicate protected guanines on the upper and lower strands. Coding and non-coding refer to sense and antisense strands, respectively.

Materials and methods

Plasmid and probe construction

The plasmids pCol(ΔI)hGH, pCol(ΔP)hGH, and pCol-hGH (Figure 1) have been described (Bornstein *et al.*, 1988). pCol(ΔP ; ΔI)hGH was prepared from pCol(ΔP)hGH by deletion of the intronic sequence between the *Sst* II sites at +291 and +1440, followed by ligation. pCol(ΔP ; I:292–670)hGH and pCol(ΔP ; I:670–292)hGH were prepared by insertion of the intronic fragment +292 to +670, (flanked by *Sst* II sites) into pCol(ΔP ; ΔI)hGH in either the positive or negative orientation, respectively, as described for the corresponding plasmids pCol(I:292–670)hGH and pCol(I:670–292)hGH (Bornstein *et al.*, 1988). All of the plasmids shown in Figure 1 contain an hGH minigene that includes the last two exons and the poly(A) addition signal (Bornstein and McKay, 1988).

Site-directed mutagenesis was performed as described by Geisselsoder *et al.* (1987). Briefly, the fragment +292 to +670 was cloned into the phagemid pBluescript II KS(+) using the *Sst* II site in the polylinker. Single-stranded templates were obtained with the use of helper phage M13K07. Recombinant plasmids were screened by the dideoxy DNA sequencing method. Constructs containing the mutated sequence pCol(I:292–670M1-AP1)hGH and pCol(I:670–292M1-AP1)hGH were prepared as described for the native plasmids. pCol(M1-AP1)hGH was prepared with the use of the parent vector pCol(I:671–1440)hGH in which the *Sst* II site at 1440 had been destroyed by the T4 DNA polymerase reaction.

Transient transfection analysis was performed as described (Bornstein *et al.*, 1988), using the calcium phosphate coprecipitation procedure for introduction of DNA into cells. The preparation of total RNA from transfected cells and analysis for hGH mRNA by RNase protection assay with a ^{32}P -labeled antisense hGH riboprobe have been described (Bornstein and

Figure 8. Gel mobility shift analysis for S1-binding activity in nuclear extracts from fibroblast and lymphoid cell lines. Lane 1, S1 probe with no nuclear extract; lane 2, 3 μ g HeLa extract; lane 3, 24 μ g CTF extract; lane 4, 24 μ g SV-MSC extract; lane 5, 20 μ g Manca extract; lane 6, 12 μ g Jurkat extract; lane 7, 5 μ g L cell extract.

McKay, 1988). Efficiency of transfection was monitored by cotransfection with a plasmid containing the β -galactosidase gene under the control of the Rous sarcoma virus long terminal repeat (Bornstein and McKay, 1988). In experiments using SV-MSC, transfection efficiency was assessed by cotransfection with a plasmid (pAN 132) that contains the β -globin gene and its promoter driven by two copies of the SV40 72-bp direct repeat and by measurement of β -globin mRNA with the use of a ribonuclease protection assay.

Probes for analysis of DNA-protein binding were prepared from pCol(I:292–670)hGH and pCol(I:670–292)hGH utilizing a unique *Bss*HI restriction site at +474 and a *Hind*III site 3' to the intronic insertion to yield the fragments +292 to +474 and +475 to +670, respectively. The probes were modified for methylation interference analysis by reaction with dimethyl sulfate, as described by Maxam and Gilbert (1980). Double-stranded oligonucleotides corresponding to the intronic sequences +913 to +944, +983 to +1111, and +588 to +616 (designated A1, A3, and S1, respectively), with added *Bam*HI cohesive ends, were synthesized by the Howard Hughes Medical Institute Chemical Synthesis Facility, Seattle, WA. The structures of these oligonucleotides are as follows

A1 (5'-CCGGTGAGGTCGGCCCCGCCCGGCCCATTTG-3')

A3 (5'-CGCTAGCGCGGTGGGAGTGGTTAGCTAAC-3')

S1 (5'-GCCGTGGGATGATTTCATAAGGAAAGATTG-3')

Oligonucleotides were annealed in 20 mM tris(hydroxymethyl)aminomethane (Tris, pH 8.0), 1 mM EDTA at 20°C for 1 h. Fragments were 5' end-labeled with 32 P by the use of T4 polynucleotide kinase and purified by electroelution from acrylamide gels (Bornstein *et al.*, 1988). Oligonucleotides used for site-specific mutagenesis were synthesized without added cohesive ends.

Nuclear extract preparation

HeLa and L cells were grown in suspension in Dulbecco's modified Eagle's medium (DMEM) with 10% calf serum; Manca cells (SK-DHL2A; Saito *et al.*, 1983) were grown in RPMI 1640 medium with 5% calf serum; and Jurkat E6 cells (Weiss *et al.*, 1984) were grown in RPMI 1640 with 10% fetal calf serum. Cells in suspension were harvested at a density of $0.8\text{--}1.0 \times 10^6$ cells/ml. Primary cultures of CTF and SV-MSC (Singer *et al.*, 1987) were maintained in DMEM with 10% calf serum and were harvested at confluence by scraping. L-glutamine, streptomycin, and penicillin were added to all growth media. Nuclear extracts were prepared by the procedure of Shapiro *et al.* (1988) with 1 mM phenylmethylsulfonyl fluoride added to all buffers. Nuclear extracts were stored at -70°C in nuclear dialysis buffer (20 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid [HEPES], pH 7.9, 20% glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.2 mM ethylene glycol-bis(β -aminoethyl ether)-*N,N,N',N'*-tetraacetic acid [EGTA], 2 mM dithiothreitol).

Figure 9. Mutation of the AP1 consensus sequence in the S1-binding site. (A) HeLa nuclear extract was incubated with S1 probe as described in Figure 3, either without added competitor (lane 1) or in the presence of 50- or 200-fold excess, respectively, of S1 (lanes 2 and 3), M2-AP1 (lanes 4 and 5), or M1-AP1 (lanes 6 and 7). (B) Sequences of S1 and related oligonucleotides. The parent sequence S1 is shown, and the mutations in the oligonucleotides M1-AP1 and M2-AP2 are indicated. The AP1 consensus sequence is also shown.

DNA-protein binding analysis

Binding reactions were performed in nuclear dialysis buffer with the addition of 0.2 M KCl, 6 mM MgCl₂, 1 mM ZnSO₄ in a final volume of 20 μ L. After a preincubation of the nuclear extract with 4 μ g poly[dl-dC] · poly[dl-dC], labeled probe (0.1–0.5 ng; 15 000 cpm) was added and the incubation was continued for 20 min at 20°C. Reaction mixtures were then loaded directly onto a 5% polyacrylamide gel, and electrophoresis was performed as described (Bornstein *et al.*, 1988). Competitor DNA fragments were prepared by restriction enzyme digestion and purified by electroelution from agarose gels.

DNase protection assays were performed by preincubation of nuclear extracts with 80–100 ng sonicated salmon sperm DNA in nuclear dialysis buffer containing 2.5 mM CaCl₂, 40 mM KCl, and 5 mM MgCl₂ for 10 min at 4°C in a volume of 10 μ L. The ³²P-labeled fragment (40 000 cpm) was added in a 1- μ L volume and the incubation was continued for 20 min at 20°C. DNase (freshly diluted into nuclear dialysis buffer from a 5 mg/ml stock solution stored in 0.15 M NaCl, 50% glycerol) was then added in 1 μ L and allowed to react for 90 s before the reaction was quenched with 250 μ L of 0.15 M NaCl, 0.3% sodium dodecyl sulfate. The reaction products were extracted with phenol/chloroform and chloroform and then precipitated from ethanol in the presence of 15 μ g glycogen. The products were analyzed on a polyacrylamide gel containing 50% urea.

Methylation interference experiments were performed by incubation of the chemically modified probe with the nuclear extract, followed by gel mobility analysis. The bands corresponding to the free and protein-complexed probes were excised after exposure of the wet gel. The DNA was eluted, subjected to piperidine cleavage, and electrophoresed on a 6% sequencing gel.

Acknowledgments

We thank John Geiser for assistance in the early phases of this work, Elizabeth Enari for technical assistance, Jack W. Singer for the SV-MSC cells, James Roberts for Manca cells, Margaret Allen for Jurkat cells, and Kathleen Doebling for preparation of the manuscript. We are indebted to David Rowe (University of Connecticut) for making the rat $\alpha 1(I)$ intronic sequence available to us before publication, to Nicholas Anagnou for the pAN 132 plasmid, and to Helene Sage for a critical reading of the manuscript.

This work was supported by Public Health Service grants AM-11248 and DE-08229 from the National Institutes of Health. D.J. Liska was supported by Training Grant HL07312.

Received: November 30, 1989.

Revised and accepted: April 17, 1989.

References

- Andrews, D.F., III, Nemunaitis, J., Tompkins, C., and Singer, J.W. (1989). Effect of 5-azacytidine on gene expression in marrow stromal cells. *Mol. Cell. Biol.* 9, 2748–2751.
- Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R.J., Rahmsdorf, H.J., Jonat, C., Herrlich, P., and Karin, M. (1987). Phorbol ester-inducible genes contain a common *cis* element recognized by a TPA-modulated *trans*-acting factor. *Cell* 49, 729–739.
- Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* 33, 729–740.

- Bornstein, P., McKay, J., Morishima, J.K., Devarayalu, S., and Gelinas, R.E. (1987). Regulatory elements in the first intron contribute to transcriptional control of the human $\alpha 1(I)$ collagen gene. *Proc. Natl. Acad. Sci. USA* 84, 8869–8873.

- Bornstein, P., and McKay, J. (1988). The first intron of the $\alpha 1(I)$ collagen gene contains several transcriptional regulatory elements. *J. Biol. Chem.* 263, 1603–1606.

- Bornstein, P., McKay, J., Liska, D.J., Apone, S., and Devarayalu, S. (1988). Interactions between the promoter and first intron are involved in transcriptional control of $\alpha 1(I)$ collagen gene expression. *Mol. Cell. Biol.* 8, 4851–4857.

- Bornstein, P., and Sage, H. (1989). Regulation of collagen gene expression. *Prog. Nucleic Acid Res. Mol. Biol.* 37, 66–106.

- Burbelo, P.D., Martin, G.R., and Yamada, Y. (1988). $\alpha 1(IV)$ and $\alpha 2(IV)$ collagen genes are regulated by a bidirectional promoter and a shared enhancer. *Proc. Natl. Acad. Sci. USA* 85, 9679–9682.

- Chiu, R., Angel, P., and Karin, M. (1989). Jun-B differs in its biological properties from, and is a negative regulator of, c-Jun. *Cell* 59, 979–986.

- Choy, H.A., McLaughlin, M., and Feifelson, P. (1989). Nuclear factors from expressing tissues interact *in vitro* with a rat $\alpha 2u$ globulin gene intron. *J. Biol. Chem.* 264, 1754–1759.

- Delclos, K.B., and Blumberg, P.M. (1979). Decrease in collagen production in normal and Rous Sarcoma virus-transformed chick embryo fibroblasts induced by phorbol myristate acetate. *Cancer Res.* 39, 1667–1672.

- Dion, L.D., Bear, J., Bateman, J., De Luca, L.M., and Colburn, N.H. (1982). Inhibition by tumor-promoting phorbol esters of procollagen synthesis in promotable JB6 mouse epidermal cells. *J. Natl. Cancer Inst.* 69, 1147–1153.

- Dynan, W.S., and Tjian, R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature* 316, 774–778.

- Galson, D.L. and Housman, D.E. (1988). Detection of two tissue-specific DNA-binding proteins with affinity for sites in the mouse β -globin intervening sequence 2. *Mol. Cell. Biol.* 8, 381–392.

- Geisselsoder, J., Witney, F. and Yuckenberg, P. (1987). Efficient site-directed *in vitro* mutagenesis. *Biotechniques* 5, 786–791.

- Gillies, S.D., Morrison, S.L., Oi, V.T., and Tonegawa, S. (1983). A tissue-specific transcriptional enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* 33, 717–728.

- Hallauer, P.L., Hastings, K.E.M., and Peterson, A.C. (1988). Fast skeletal muscle-specific expression of a quail troponin I gene in transgenic mice. *Mol. Cell. Biol.* 8, 5072–5079.

- Hayashi, S., Goto, K., Okada, T.S., and Kondoh, H. (1987). Lens-specific enhancer in the third intron regulates expression of the chicken $\delta 1$ -crystallin gene. *Genes Dev.* 1, 818–828.

- Hirai, S.-I., Ryseck, R.-P., Mechta, F., Bravo, R., and Yaniv, M. (1989). Characterization of *jund*: a new member of the *jun* proto-oncogene family. *EMBO J.* 8, 1433–1439.

- Horton, W., Miyashita, T., Kohno, K., Hassell, J.R., and Yamada, Y. (1987). Identification of a phenotype-specific enhancer in the first intron of the rat collagen II gene. *Proc. Natl. Acad. Sci. USA* 84, 8864–8868.

- Kerr, L.D., Holt, J.T., Matrisian, L.M. (1988). Growth factors regulate transin gene expression by *c-fos*-dependent and *c-fos*-independent pathways. *Science* 242, 1424–1427.
- Killen, P.D., Burbelo, P.D., Martin, G.R., and Yamada, Y. (1988). Characterization of the promoter for the $\alpha 1(\text{IV})$ collagen gene: DNA sequences within the first intron enhance transcription. *J. Biol. Chem.* 263, 12310–12314.
- Lee, W., Mitchell, P., and Tjian, R. (1987). Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell* 49, 741–752.
- Li, L., Hu, J-S., and Olson, E.N. (1990). Different members of the *jun* proto-oncogene family exhibit distinct patterns of expression in response to type β transforming growth factor. *J. Biol. Chem.* 265, 1556–1562.
- Maxam, A.M., and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65, 499–559.
- Mercola, M., Wang, X., Olsen, J., and Calame, K. (1983). Transcriptional enhancer elements in the mouse immunoglobulin heavy chain locus. *Science* 221, 663–665.
- Müller, M.M., Gerster, T., and Schaffner, W. (1988). Enhancer sequences and the regulation of gene transcription. *Eur. J. Biochem.* 176, 485–495.
- Ng, S-Y., Gunning, P., Liu, S-H., Leavitt, J., and Kedes, L. (1989). Regulation of the human β -actin promoter by upstream and intron domains. *Nucleic Acids Res.* 17, 601–615.
- Penttinen, R.P., Kobayashi, S., and Bornstein, P. (1988). Transforming growth factor β increases mRNA for matrix proteins both in the presence and in the absence of changes in mRNA stability. *Proc. Natl. Acad. Sci. USA* 85, 1105–1108.
- Pertovaara, L., Sistonen, L., Bos, T.J., Vogt, P.K., Keski-Oja, J., and Alitalo, K. (1989). Enhanced *jun* gene expression is an early genomic response to transforming growth factor β stimulation. *Mol. Cell. Biol.* 9, 1255–1262.
- Raghow, R., and Thompson, J.P. (1989). Molecular mechanisms of collagen gene expression. *Mol. Cell. Biochem.* 86, 5–18.
- Rippe, R.A., Lorenzen, S., Brenner, D.A., and Breindl, M. (1989). Regulatory elements in the 5'-flanking region and the first intron contribute to transcriptional control of the mouse $\alpha 1$ type I collagen gene. *Mol. Cell. Biol.* 9, 2224–2227.
- Risse, G., Jooss, K., Neuberg, M., Brüller, H-J. and Müller, R. (1989). Asymmetrical recognition of the palindromic AP1 binding site (TRE) by fos protein complexes. *EMBO J.* 8, 3825–3832.
- Rossi, P., and de Crombrughe, B. (1987). Identification of a cell-specific transcriptional enhancer in the first intron of the mouse $\alpha 2$ (type I) collagen gene. *Proc. Natl. Acad. Sci. USA* 84, 5590–5594.
- Rossi, P., Karsenty, G., Roberts, A.B., Roche, N.S., Sporn, M.B., and de Crombrughe, B. (1988). A nuclear factor 1 binding site mediates the transcriptional activation of a type I collagen promoter by transforming growth factor- β . *Cell* 52, 405–414.
- Rossouw, C.M.S., Vergeer, W.P., duPlooy, S.J., Bernard, M.P., Ramirez, F., and de Wet, W.J. (1987). DNA sequences in the first intron of the human pro- $\alpha 1(\text{I})$ collagen gene enhance transcription. *J. Biol. Chem.* 262, 15151–15157.
- Ryder, K., Lau, L.F., and Nathans, D. (1988). A gene activated by growth factors is related to the oncogene *v-jun*. *Proc. Natl. Acad. Sci. USA* 85, 1487–1491.
- Saito, H., Hayday, A.C., Wiman, K., Hayward, W.S., and Tonegawa, S. (1983). Activation of the *c-myc* gene by translocation: a model for translational control. *Proc. Natl. Acad. Sci. USA* 80, 7476–7480.
- Sanders, M.M., and McKnight, G.S. (1988). Positive and negative regulatory elements control the steroid-responsive ovalbumin promoter. *Biochemistry* 27, 6550–6557.
- Sassone-Corsi, P., Sisson, J.C., and Verma, I.M. (1988). Transcriptional autoregulation of the proto-oncogene *fos*. *Nature* 334, 314–319.
- Schütte, J., Viallet, J., Nau, M., Segal, S., Fedorko, J., and Minna, J. (1989). *jun-B* inhibits and *c-fos* stimulates the transforming and *trans*-activating activities of *c-jun*. *Cell* 59, 987–997.
- Shapiro, D.J., Sharp, P.A., Wahli, W.W., and Keller, M.J. (1988). A high-efficiency HeLa cell nuclear transcription extract. *DNA* 7, 47–55.
- Singer, J.W., Chabord, P., Keating, A., Nemunaitis, J., Raugi, G., Wight, T.N., Lopez, J.A., Roth, G.J., Dow, L.W., and Fialkow, P.J. (1987). Simian virus 40-transformed adherent cells from human long-term marrow cultures: cloned cell lines produce cells with stromal and hematopoietic characteristics. *Blood* 70, 464–474.
- Sobel, M.L., Dion, L.D., Vuust, J., and Colburn, N.H. (1983). Tumor-promoting phorbol esters inhibit procollagen synthesis at a pretranslational level in JB-6 mouse epidermal cells. *Mol. Cell. Biol.* 3, 1527–1532.
- Sternberg, E.A., Spizz, G., Perry, W.M., Vizard, D., Weil, T., and Olson, E.N. (1988). Identification of upstream and intragenic regulatory elements that confer cell-type-restricted and differentiation-specific expression on the muscle creatine kinase gene. *Mol. Cell. Biol.* 8, 2896–2909.
- Takimoto, M., Quinn, J.P., Farina, A.R., Staudt, L.M., and Levens, D. (1989). *fos/jun* and octamer-binding protein interact with a common site in a negative element of the human *c-myc* gene. *J. Biol. Chem.* 264, 8992–8999.
- Weiss, A., Wiskocil, R.L., and Stobo, J.D. (1984). The role of T3 surface molecules in the activation of human T cells: a two-stimulus requirement for IL2 production reflects events occurring at a pre-translational level. *J. Immunol.* 133, 123–128.
- Yutzey, K.E., Kline, R.L., and Konieczny, S.F. (1989). An internal regulatory element controls troponin I gene expression. *Mol. Cell. Biol.* 9, 1397–1405.