Volume 2 · Number 4 · 1992
GENE EXPRESSION

Elements in the first intron of the α1(I) collagen gene interact with Sp1 to regulate gene expression

DeAnn J. Liska, Victoria R. Robinson, and Paul Bornstein

Department of Biochemistry, University of Washington, Seattle, Washington

Sequences within the first intron of the α1(I) collagen gene act both positively and negatively to regulate expression of the gene. We have further characterized a 274 bp intronic sequence that contains an orientation-specific inhibitory activity and represents a constitutive DNase I-hypersensitive site in the gene. We show that this sequence contains two tandem, unique binding elements for the transcription factor Sp1. In addition, an Sp1-like site, capable of competing for protein binding to the intronic elements, resides in the distal promoter of the collagen gene. The results of experiments with site-directed mutations that abolish binding to the intronic elements indicate that these protein–DNA interactions have an inhibitory effect on the transcriptional efficiency of α1(I) collagen-reporter gene constructs in transient transfection analysis. These data support our conclusion that the first intron plays a complex role, involving multiple protein–DNA binding interactions, in the regulation of expression of the α1(I) collagen gene.

Recent observations suggest that the mechanisms underlying the intricate spatial and temporal expression of mammalian genes are likely to be complex and involve an interplay of cis-acting enhancers and silencers, as well as DNA methylation and chromatin conformation (Gross and Garrard, 1988; Elgin, 1988). Although the molecular details remain to be elucidated, it is clear that the interaction of nuclear, trans-acting factors with specific DNA sequences is central to the regulation of transcription of mammalian genes (Ptashne, 1988; Dynan and Tjian, 1986). Of the several sequence-specific, DNA-binding transcription factors that have been described, one of the best characterized is Sp1. Sp1 is expressed in many cultured mammalian cells and is involved in the transcription of a variety of cellular and viral genes (Dynan and Tjian, 1986; Briggs et al., 1986; Kadonaga et al., 1988). Sp1 is highly regulated during development (Saffer et al., 1991), can be modified posttranslationally by phosphorylation (Jackson et al., 1990), and displays a 10- to 20-fold range of binding affinities for specific Sp1-binding sites (Kadonaga et al., 1988). These observations indicate that Sp1 is likely to play an important role in cellular processes during development and differentiation.

Functional Sp1-specific binding sites have been localized from within 100 bp to several kb both 5′ and 3′ to the start of transcription, and these distal and proximal Sp1-binding sequences are capable of acting in a synergistic manner (Courey et al., 1989). A favored mechanism invoked to explain this observation is the interaction of distally DNA-bound Sp1 with transcription factors at or near the start of transcription. In vitro experiments have demonstrated that Sp1, bound to distal DNA sequences, can interact with proximally bound Sp1, resulting

Received April 13, 1992; revision accepted June 11, 1992.

Correspondence: Paul Bornstein, Department of Biochemistry SJ-70, University of Washington, Seattle, WA 98195 Tel (206) 543-1789 Fax (206) 685-1792

 1052-2166/92/0204/379-11$2.00

in a looping out of the intervening DNA sequences (Su et al., 1991). Li et al. (1991) have also demonstrated that proximally bound Sp1 can form a stable loop with distally bound enhancer protein E2 of the bovine papillomavirus type 1 promoter. Whether this is a functional mechanism for the interaction of distal, cis-acting sequences with proximal elements in the context of chromatin is not known.

We are interested in understanding the mechanisms involved in the regulation of expression of the human α1(I) collagen gene. Several lines of evidence have implicated a role for sequences within the first intron in the regulation of transcription of the gene. (1) A region of DNase I-hypersensitivity, approximately 900 bp from the start of transcription, has been demonstrated in the first intron of both the mouse and human α1(I) collagen genes (Barsh et al., 1984; Breindl et al., 1984). (2) Germ-line integration of a retrovirus into the first intron of the murine α1(I) collagen gene abolishes transcription of the gene in most tissues (Schnieke et al., 1983). (3) Sequences within the first intron were required for activity of a transgene containing 444 bp of the 5′-flanking sequence of the human α1(I) collagen gene in transgenic mice (Slack et al., 1991). (4) The first intron of the human and rat α1(I) collagen genes displays a substantial degree of sequence conservation, including the conservation of several putative protein-DNA binding motifs, whereas a similar interspecies sequence comparison between the fifth introns shows no such conservation (Bornstein and Sage, 1989).

The presence of intronic cis-acting sequences that show both positive and negative effects in vitro has been demonstrated in both the human and mouse α1(I) collagen genes (Bornstein et al., 1987; Rossouw et al., 1987; Bornstein and McKay, 1988; Bornstein et al., 1988; Liska et al., 1990; Rippe et al., 1989; Boast et al., 1990). These investigations have indicated that intronic, cis-acting sequences show a cell-type dependent activity and are most active in cells of a primary fetal origin (Bornstein et al., 1988; Boast et al., 1990). Consonant with that notion, several studies have demonstrated that the intronic, cis-acting activities are not observed in NIH 3T3 fibroblasts (Bornstein et al., 1988; Olsen et al., 1991). Although these studies have identified regions of cis-acting sequences, few studies have characterized specific sequence elements or investigated protein–DNA interactions within these putative regulatory regions.

We have reported that a 274 bp fragment, located in the region of the intronic DNase I-hypersensitive site, inhibited transcription from the collagen promoter in an orientation-dependent manner (Bornstein et al., 1988). Furthermore, the inhibitory activity of this cis-acting sequence was alleviated by a deletion of specific sequences in the α1(I) collagen promoter. We postulated that the 274 bp sequence was involved in a promoter–intron interaction that played a role in regulation of the α1(I) collagen gene. In this study we report that the transcription factor Sp1 is able to bind specifically to two elements, A1 and A3, that are located within this intronic sequence, as well as to a sequence in the collagen promoter. Mutation of the intronic Sp1-binding sequences leads to an increase in transcriptional efficiency of transiently transfected reporter gene constructs in chicken tendon fibroblasts (CTF). These data support a role for the transcription factor Sp1 in the regulation of expression of the α1(I) collagen gene.

Materials and methods

Plasmid and probe construction

Site-directed mutagenesis was performed by polymerase chain reaction (PCR) as follows. 3′ primers for mutation of A1 and A2 (see Fig. 1) were a 74-mer from +924 to +997 (5′-CCCACCGCGCTAGCGGGCTCCCCAGACGGC CACTAGGTGTCGGGCAACGCTAAGCAATGG GGCCGGGTATGGGCC-3′) and a 44-mer from +953 to +997 (5′-CCCACCGCGCTAGCGGCT CCCCAGACGGTTGAGCTTCGTCGGGC-3′, respectively. Here and elsewhere underlined bases represent changes from the wild-type sequence. The 5′ primer was the top strand of oligo S1 from +588 to +616, which has been previously described (Liska et al., 1990). The plasmid pUC19E, containing the 2.3 kb Hind III-Hind III fragment of the α1(I) collagen gene, including 804 bp upstream of the start of transcription, the first exon, and most of the first intron, has been described previously (Bornstein et al., 1987). The wild-type fragment containing the A1 and A2 sites was excised from pUC19E by Bgl II-Nhe I digestion. The 400 bp PCR product containing the mutation was purified by

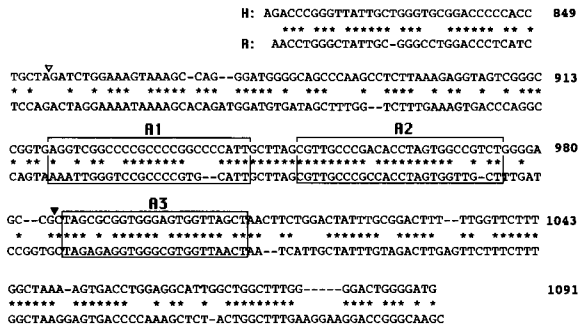


Figure 1. Alignment of the 274 bp sequence (from +820 to +1093) of the human α1(I) collagen intron (top line) with the rat α1(I) collagen intron (bottom line). Base numbers are for the human sequence, with base 1 as the start of transcription. Boxed sequences represent DNase-I protection footprints A1, A2, and A3, corresponding to bases +919 to +944, +951 to +978, and +986 to +1009 in the human sequence, respectively (Bornstein et al., 1987). The Bgl II and Nhe I recognition sites in the human sequence are indicated by the open and filled triangles, respectively. The degree of sequence identity (indicated by asterisks) is 66%.

agarose gel electrophoresis, subjected to digestion with Bgl II and Nhe I, and cloned into the Bgl II-Nhe I site of pUC19E, producing pCol(MA1) and pCol(MA2), respectively. Mutation of the A3 site was performed by PCR, with the 5′ primer (5′-GGAGCCGCTAGCGCGGT ATTAGTGGT-3′) and 3′ primer (5′-GGAGGTCCA GCCCTCATCCC-3′) corresponding to sequences +978 to +1013 and +1247 to +1266, respectively. The wild-type fragment was excised from pUC19E using Nhe I and a partial restriction digestion with BspM I and replaced with the PCR product, after Nhe I-BspM I restriction digestion, to produce pCol(MA3). pCol(MA1/MA3)hGH (human growth hormone) was prepared by excision of the Bgl II-Nhe I fragment from pCol(MA1) and ligation into the Bgl II-Nhe I site of pCol(MA3).

The 2.3 kb collagen fragment, flanked by Hind III sites, was excised from the plasmids containing mutations and was cloned into the Hind III site of pUC19-hGH (Bornstein and McKay, 1988) to produce pCol(MA1)hGH, pCol(MA2)hGH, pCol(MA3)hGH, and pCol(MA1/MA3)hGH. Orientation of the Hind III fragment was determined by Pvu II restriction digestion. All mutations were verified in the final plasmid by double-stranded DNA sequencing.

The constructs pCol-hGH, pCol(671–1440)hGH, pCol(1440–671)hGH, and pCol[ΔI]hGH have been previously described (Bornstein et al., 1988; Liska et al., 1990). Plasmids pCol(671–1440MA1/MA3)hGH and pCol(1440–671MA1/MA3)hGH were prepared analogously to pCol(671–1440)hGH and pCol(1440–671)hGH. Briefly, the +671 to +1440 fragment was isolated from pCol(MA1/MA3) by Sst II digestion and substituted in both orientations into pCol[ΔI]hGH. All of the plasmids contain an hGH minigene that includes the last two exons, an intron, and the poly(A) addition signal of the hGH gene (Bornstein and McKay, 1988).

Protein-DNA binding assays

Preparation of nuclear extracts from chicken tendon fibroblasts (CTF) and HeLa cells, mobility shift assays, and methylation interference analyses were performed as previously described (Bornstein et al., 1988; Liska et al., 1990). Mobility shift assays with purified Sp1 protein (Jackson et al., 1990) were performed in nuclear dialysis buffer (Liska et al., 1990) with the addition of 0.2 M KCl, 12 mM $MgCl_2$, 0.1% (v/v) Nonidet P-40. Oligonucleotide Sp1 (5′-GATCGATCG GGGCGGGGCGATC-3′) was obtained from Stratagene. Other double-stranded oligonucleotides used for mobility shift assays were prepared by the Howard Hughes Medical Institute Chemical Synthesis Facility, Seattle, as follows:

A1	(5′-CCGGTGAGGTCGGCCCCGCCCCGGCCCCATTG-3′)
MA1	(5′-CCGGTGAGGTCGGCCCATACCCGGCCCCATTG-3′)
A3	(5′-CGCTAGCGCGGTGGGAGTGGTTAGCTAAC-3′)
MA3	(5′-CGCTAGCGCGGTATTAGTGGTTAGCTAAC-3′)
A2	(5′-CCATTGCTTAGCGTTGCCCGACACCTAGTGG-3′)
S1	(5′-GCCGTGGGATGATTCATAAGGAAAGATTG-3′)
P1	(5′-TTTATGCCCCTCCCTTAGCT-3′)

A probe containing the protected region A1 and A2 was prepared by digestion of pUC19E with Bgl II and Nhe I. A probe containing protected region A3, and lacking A1 and A2, was prepared from A274-pUC19 (Bornstein et al., 1987) by deletion of the sequences between the Bgl II and Nhe I sites (see Fig. 1). The promoter sequences −477 to −256 of the human α1(I) collagen gene was subcloned into pUC19. Probes for protein binding were prepared from this plasmid using restriction endonuclease sites in the polylinker and an internal Xba I site at −331. Probes were 5′ end-labeled with ^{32}P on the coding and non-coding strands with T4 polynucleotide kinase after restriction endonuclease digestion.

Transient transfection and RNase protection assay

Procedures for transient transfection of CTF, purification of RNA and analysis for hGH mRNA by RNase protection assay have been previously described (Bornstein et al., 1988; Liska et al., 1990). Efficiency of transfection was monitored by cotransfection with a plasmid containing a neomycin-resistance gene driven by a metallothionein promoter.

Results

Protein–DNA interactions in the 274 bp intronic sequence

The 274 bp sequence has been shown to contain three regions of protection when analyzed by DNase-I footprint assay with nuclear extracts prepared from HeLa, L-cell, and a lymphoid cell line (Bornstein et al., 1987). These footprinted regions contain a high-affinity Sp1 box (A1) and a GC-rich region and viral core enhancer motif (A3) and show both spatial and base conservation between rodent and human genes (Fig. 1). Although a complex pattern of protein-DNA binding was obtained by DNase-I footprint analysis, gel mobility shift analysis with the 274 bp sequence has shown only one major and one minor band after incubation with CTF or HeLa nuclear extracts (Bornstein et al., 1988; data not shown). To further define the protein–DNA interactions in this region, we performed methylation interference analysis of the specifically shifted bands obtained with the 274 bp sequence after incubation with either HeLa or CTF nuclear extract. Sequences within the A1 and A3 footprinted regions showed specific base interactions with nuclear proteins (Fig. 2, A, B, and C). Interestingly, no interactions within the A2 footprinted sequence were observed, although this protected region shows the highest degree of conservation between the human and rodent genes (Fig. 1). The GC element in A3 does not directly conform to the GC consensus motif (5′-GGGCGG-3′; Briggs et al., 1986; Jones et al., 1986) but instead is similar to an Sp1-binding sequence that has been identified in the human growth hormone promoter (5′-GGGTGG-3′; Lemaigre et al., 1990).

To determine whether the two regions, A1 and A3, were able to bind factors independently of each other, a probe was prepared that contained the A1 and A2 footprinted regions only. Figure 3A indicates that, in the absence of the A3 sequence, the A1-A2 probe is able to bind nuclear factors. A probe containing only the A3 sequence shows the same pattern of binding (Fig. 3B). Double-stranded oligonucleotides containing the A1 and A3 binding sequences were able to compete fully for binding to either the A1-A2 or A3 probes (Fig. 3A and B). These data indicate that the A1 and A3 regions bind the same nuclear factors independently of each other. Mutations in the GC-rich sequence of either A1 or A3 abolish the ability of the oligonucleotides to compete for binding (Fig. 4), further supporting the observation that bases within the viral core enhancer motif in A3 are not involved in specific nuclear protein–DNA interactions under these conditions.

Methylation interference experiments did not identify any specific protein–DNA interactions within the A2 footprint region. To investigate further whether the conserved A2 region contains sequences capable of binding nuclear factors in HeLa or CTF nuclear extracts, we prepared a double-stranded oligonucleotide to the conserved sequences between +939 and +969, designated oligo A2. No specific binding was observed to oligo A2, either alone or after cloning in multiple copies (data not shown). The oligo A2 also did not compete for binding to a larger fragment containing region A3 (Fig. 3B) or to the fragment A1-A2 (data not shown). These observations suggest that either the A2 region is not able to bind nuclear factors independently of the larger sequence, or that the A2 region is not recognized by factors in HeLa or CTF nuclear extracts. These data do not exclude the possibility that the A2 sequence interacts with developmental or cell-type specific factors.

The transcription factor Sp1 is able to interact with the A1 and A3 regions

Binding to the A1 region appears to encompass the Sp1 consensus motif. In addition, binding to the A3 region does not appear to involve bases within the viral core enhancer motif but instead involves a GC-rich region just 5′ to that motif. To investigate whether Sp1 was able to bind to the A1 and/or A3 element, we performed a mobility shift assay with purified Sp1. Figure 5A indicates that both A1 and A3 are capable of binding purified Sp1, and both produce the same mobility shift pattern. Furthermore, oligo-

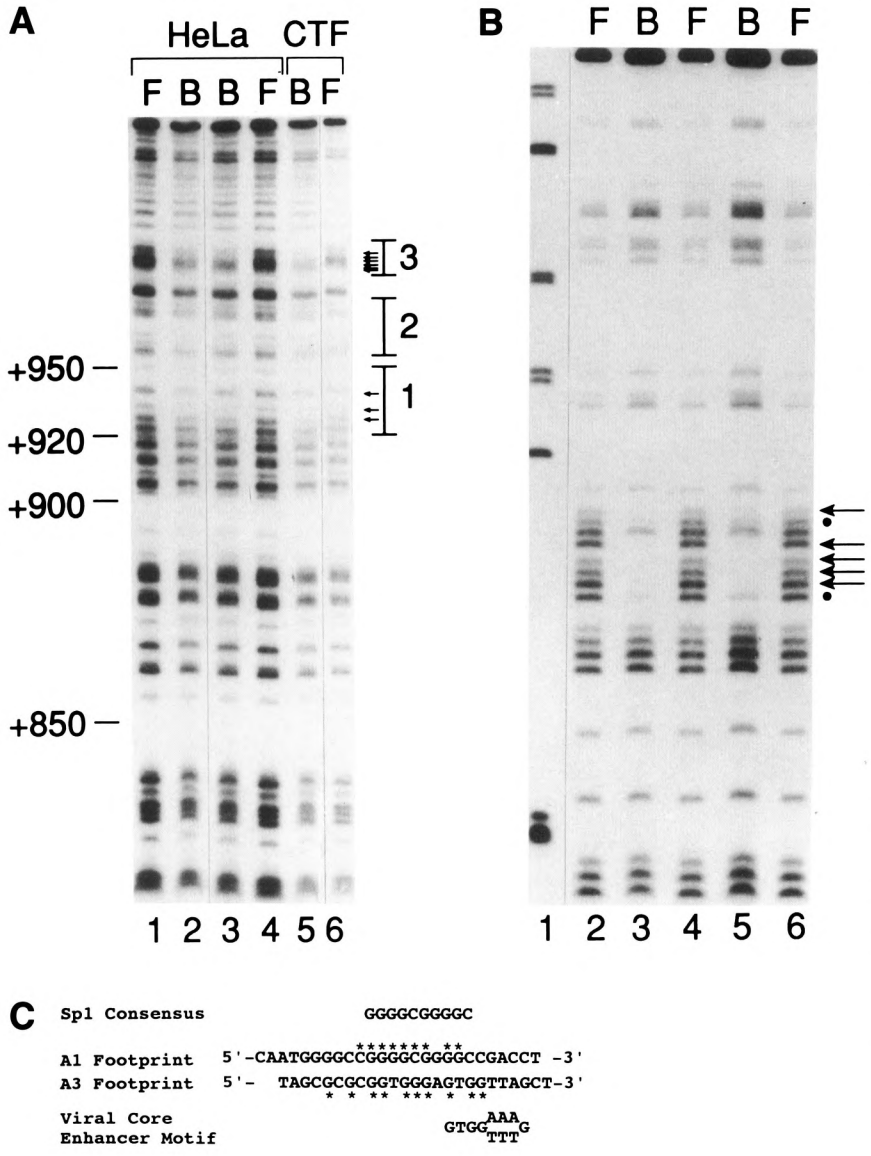


Figure 2. Methylation interference analysis of the 274 bp sequence. **A.** The 274 bp fragment was end-labeled on the coding strand and subjected to partial methylation and depurination. After incubation with either HeLa or CTF nuclear extract and gel mobility shift analysis, the bands corresponding to the protein-bound (lanes 2, 3, 5) and uncomplexed (lanes 1, 4, 6) probe were excised and processed separately. Base numbers refer to base 1 as the start of transcription. Sequences corresponding to A1, A2, and A3 are indicated. Arrows indicate the protected guanines. **B.** The non-coding strand from +854 to +984 was end-labeled and analyzed for specific binding as described. Uncomplexed probe is represented in lanes 2, 4, and 6. Protein-bound probe is represented in lanes 3 and 5. The arrows represent full protection, and the dots represent partial protection of bases within the A1 protected region. The molecular weight markers in lane 1 indicate bands of 123, 110, 90, 76, 67, and 34 bp. **C.** Summary of the methylation interference data. Shown are the non-coding and coding strand of the A1 and A3 footprints, respectively. The viral core enhancer and Sp1 consensus sequences are also shown. Protected bases are indicated by asterisks.

nucleotides containing the A1 or A3 sequences competed for binding to an oligonucleotide containing a high-affinity site for the nuclear factor Sp1 (Fig. 5B). Oligonucleotides A1 and A3 competed for binding of either ^{32}P-labeled A1 or A3 at the same concentrations, suggesting that the two sequences bind the same factor independently and with similar affinities (data not shown). This finding is consistent with previous studies that have failed to demonstrate cooperative binding of Sp1 to adjacent sites (Gidoni et al., 1985; Anderson and Freytag, 1991).

An Sp1-binding site resides in a distal promoter sequence

We have previously reported that sequences in the promoter between bases −477 to −256 were able to compete for binding to the 274 bp sequence and have hypothesized that these two regions may interact via DNA-binding proteins. We wanted to identify the sequences within the promoter region responsible for the specific competition for the A1- and A3-binding activities. Mobility shift analysis indicated that sequences between −477 and −332 competed for

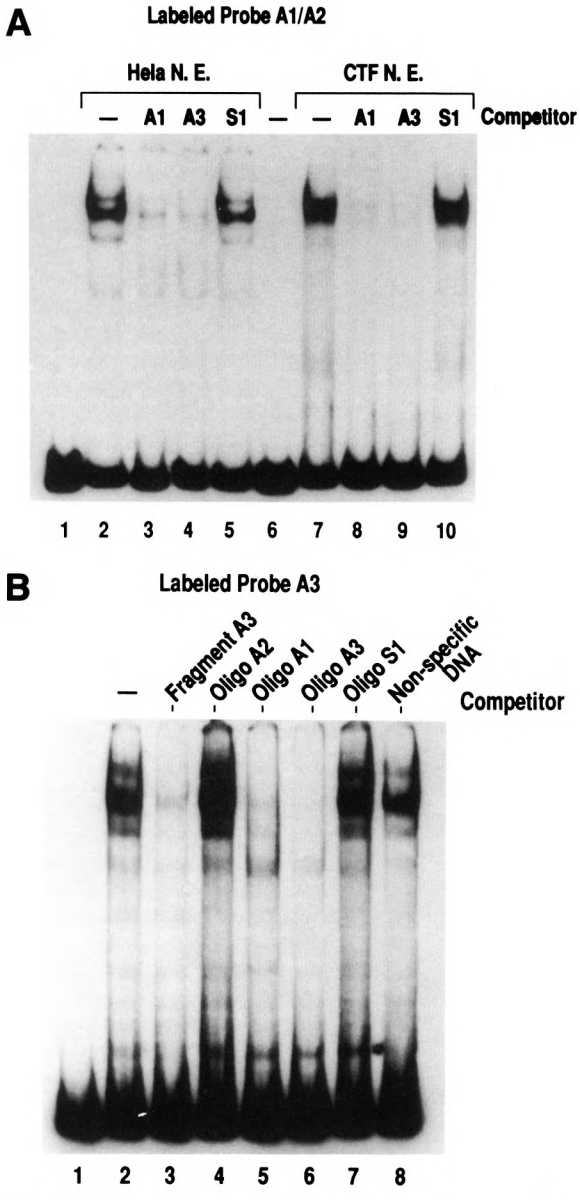


Figure 3. Nuclear protein binding of fragments A1-A2 and A3. **A.** Fragment A1-A2 was end-labeled and analyzed for protein binding in a gel mobility shift assay with either 10 μg HeLa (lanes 2–5) or 16 μg CTF (lanes 7–10) nuclear extract. Lanes 1 and 6, no nuclear extract; lanes 2 and 7, no added competitor; lanes 3 and 8, 50-fold molar excess of oligonucleotide A1; lanes 4 and 9, 50-fold molar excess of oligonucleotide A3; lanes 5 and 10, 50-fold molar excess of oligonucleotide S1. **B.** The sequences between the Bgl II and Nhe I sites, containing A1 and A2, were excised from the 274 bp sequence to produce fragment A3. Fragment A3 was end-labeled and analyzed for protein binding in a gel mobility shift assay with 10 μg HeLa nuclear extract. Lane 1, no nuclear extract added; lane 2, no added competitor; lane 3, 50-fold molar excess of fragment A3; lane 4, 50-fold molar excess of oligonucleotide A2; lane 5, 50-fold molar excess of oligonucleotide A1; lane 6, 50-fold molar excess of oligonucleotide A3; lane 7, 50-fold molar excess of oligonucleotide S1; lane 8, 80 ng salmon sperm DNA.

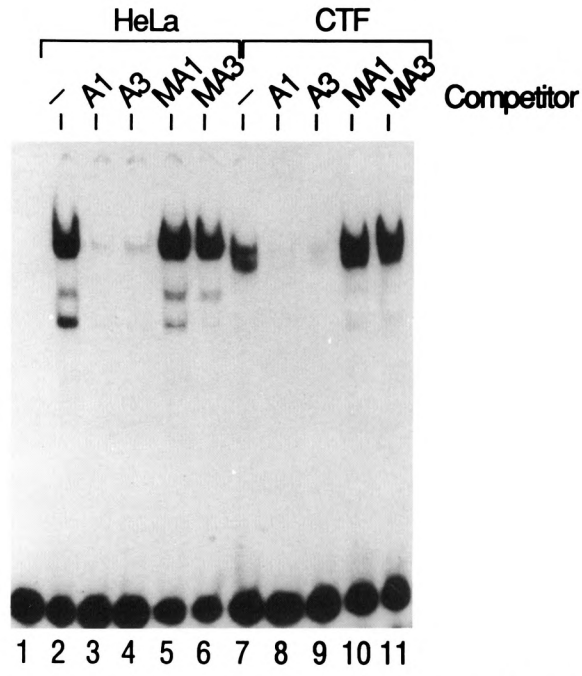


Figure 4. Mobility shift analysis with oligonucleotides A1 and A3. Oligonucleotide A1 was ^{32}P-end-labeled and assayed for protein binding activity by incubation with either 10 μg HeLa (lanes 2–6) or 24 μg CTF (lanes 7–11) nuclear extract. Lane 1, no nuclear extract was added; lanes 2 and 7, nuclear extract with no added competitor; lanes 3 and 8, 100-fold excess of oligonucleotide A1; lanes 4 and 9, 100-fold excess of oligonucleotide A3; lanes 5 and 10, 100-fold excess of oligonucleotide MA1; lanes 6 and 11, 100-fold excess of oligonucleotide MA3.

nuclear factor binding to the 274 bp sequence, whereas sequences between -331 and -256 were ineffective (data not shown). Figure 6A shows that a number of protein–DNA complexes are formed after incubation of the promoter sequence -477 to -332 with HeLa nuclear extract. The A1 and A3 oligonucleotides competed for most of these interactions. Methylation interference analysis indicated that these band shifts were due to interactions with a GC-rich sequence (Fig. 6B and C). The bands that were not competed by the A1 or A3 oligos showed no specific protection pattern and appeared to result from nonspecific interactions (Fig. 6B; data not shown). These data suggest that the binding activities we have observed in the promoter between -477 to -332 are due to Sp1.

To investigate further whether the promoter sequence is capable of binding the same factor as the A1 and A3 sequences, we prepared a double-stranded oligonucleotide to bases -458 to -473, designated oligo P1. Mobility shift

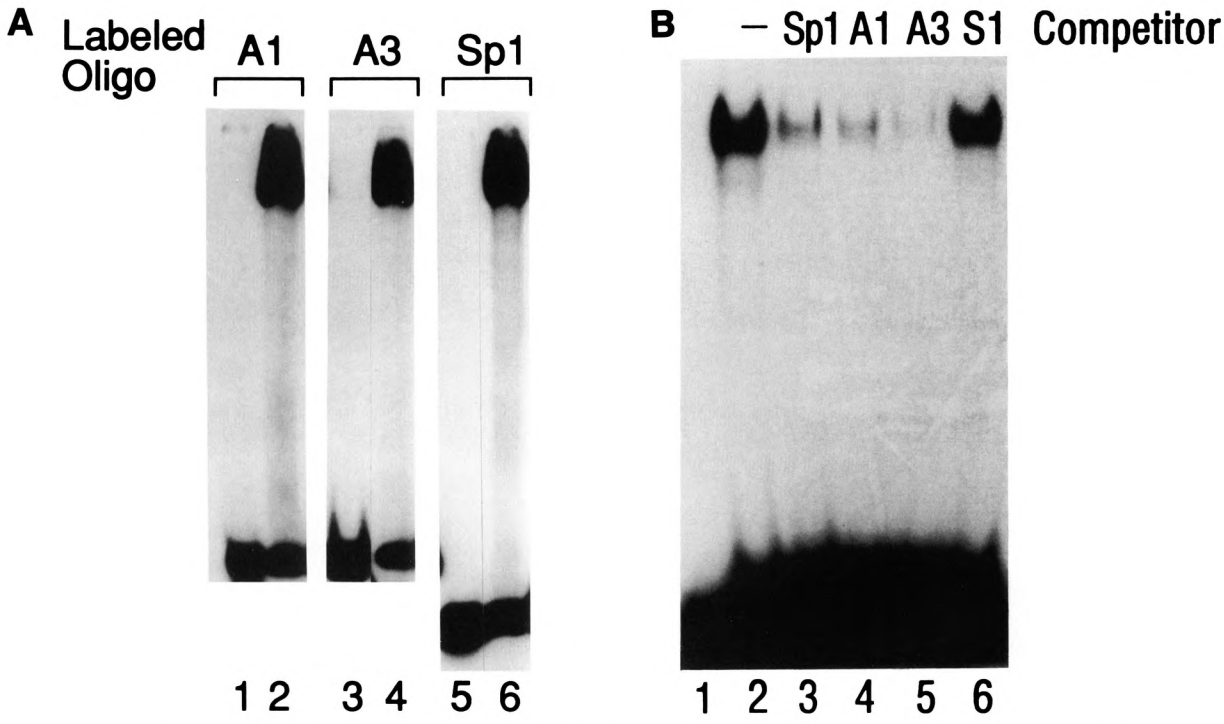


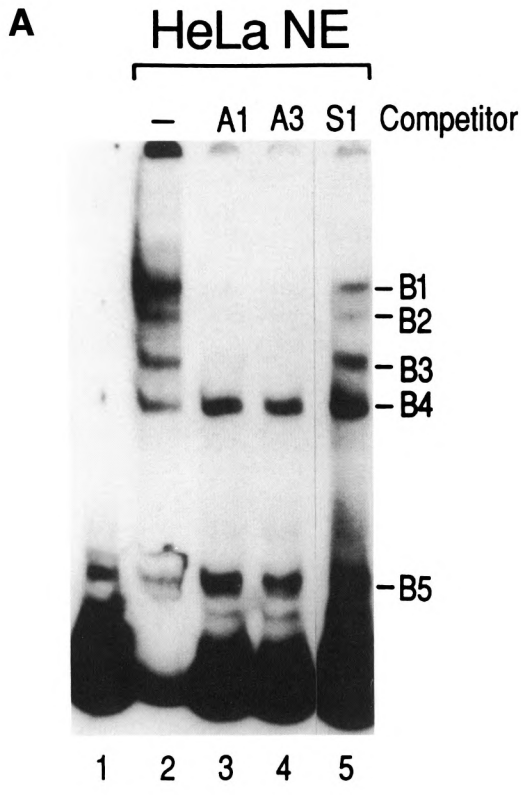
Figure 5. Binding analysis of Sp1 with the A1 and A3 oligonucleotides. **A.** The binding of purified Sp1 protein was assayed by mobility shift analysis to ^{32}P end-labeled oligonucleotide A1 (lanes 1 and 2), oligonucleotide A3 (lanes 3 and 4), and oligonucleotide Sp1 (lanes 5 and 6). Lanes 1, 3, and 5, no added protein; lanes 2, 4, and 6, 20 ng Sp1 protein was added. **B.** An oligonucleotide containing a high affinity binding site for the transcription factor Sp1 was ^{32}P-end-labeled and incubated with 10 μg HeLa nuclear extract. Lane 1, no nuclear extract; lane 2, nuclear extract with no competitor; lane 3, 50-fold excess of oligonucleotide Sp1; lane 4, 50-fold excess of oligonucleotide A1; lane 5, 50-fold excess of oligonucleotide A3; lane 6, 50-fold excess of oligonucleotide S1.

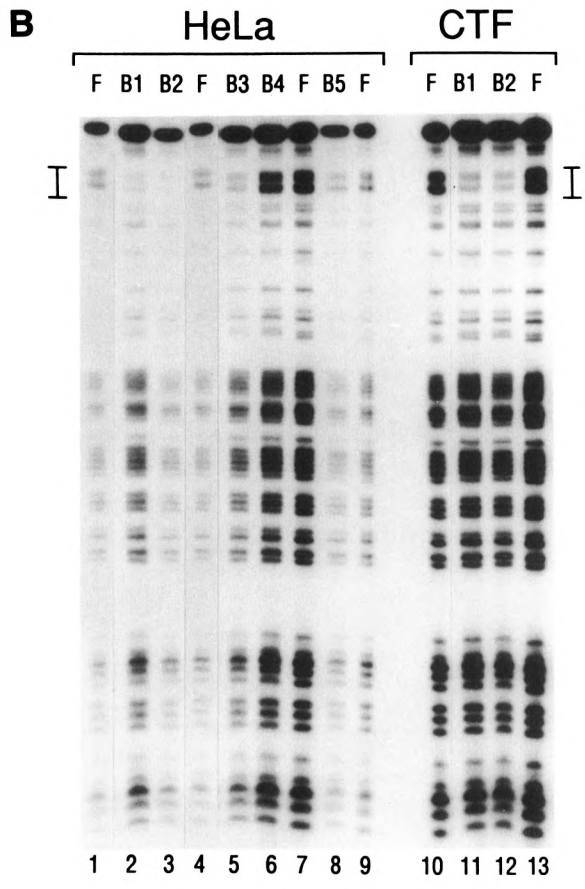
analysis with oligo P1 and nuclear extracts indicated that this sequence yielded the same shift as that observed with the high affinity Sp1-binding oligo (Fig. 6D, lanes 1 and 7). Oligos P1 and A1 were able to compete equally well for all of the binding activity to the high-affinity Sp1 binding sequence (Fig. 6D, lanes 7–11). Furthermore, oligonucleotides containing binding motifs for AP1 (S1, Fig. 6D, lane 12) or NF-1 (data not shown) did not compete for this binding activity. We conclude from these data that the binding activities we have observed in the 274 bp sequence and in the promoter between −477 to −332 are due to Sp1 or an Sp1-like factor.

Transcriptional activity of the A1, A2, and A3 intronic sequences

To investigate the role of these binding elements in the transcriptional regulation of the α1(I) collagen gene, we transferred the site-specific mutations in A1, A2, and A3 into plasmids containing the collagen promoter, first exon, and first intron. These plasmids were analyzed for relative transcriptional efficiency after transfection into CTF. The expression of the plasmids containing the mutations in either A1 or A3 was slightly higher (approx. 1.5-fold) than that of the parent construct, pCol-hGH (data not shown). We then investigated whether mutations in both of these sites would affect the transcriptional efficiency of the collagen promoter to a greater extent. As shown in Figure 7A, line 3, mutations in both A1 and A3 result in a 2-fold enhancement of transcription. A mutation in the conserved A2 sequence shows no effect on transcriptional efficiency of the collagen promoter in the context of the full intron (Fig. 7A, line 4).

The effect of the A1 and A3 mutations on transcription is modest. We tested whether removal of a previously described enhancer activity, attributed to an AP1 site located between bases +588 and +616 (Liska et al., 1990), would alter the effect observed with the A1 and A3 mutations. Plasmids containing only the intronic +671 to +1440 sequence showed an





C 5'- CCTCTTTATGCCCCTCCCTTAGCTCTTGCCA-3'
GGAGAAATACGGGGAGGGAATCGAGAACGGT
**** ***

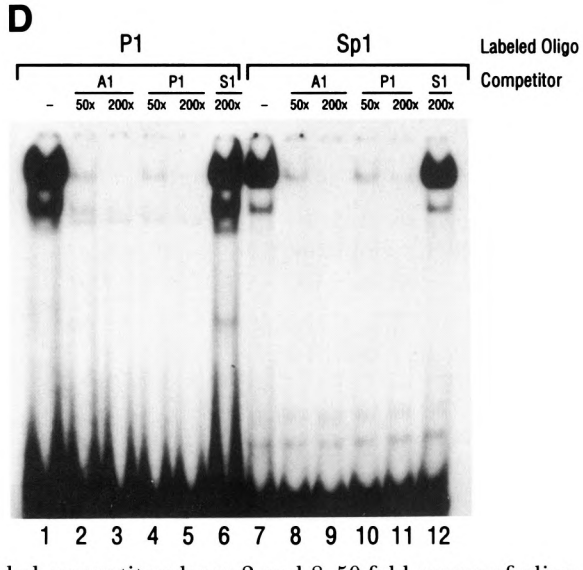
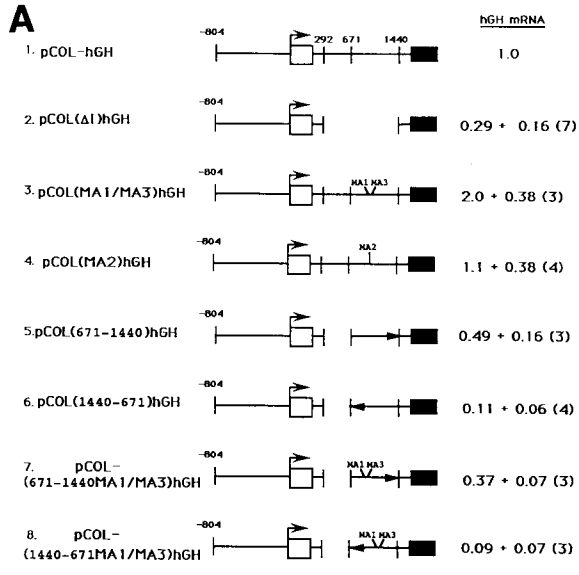


Figure 6. Nuclear protein binding to the promoter sequence −477 to −332. **A.** The collagen promoter fragment −477 to −332 was ^{32}P end-labeled and analyzed for protein binding by gel mobility shift assay with 10 μg HeLa nuclear extract. Lane 1, no nuclear extract was added; lane 2, nuclear extract with no added competitor; lane 3, 100-fold molar excess of oligonucleotide A1; lane 4, 100-fold molar excess of oligonucleotide A3; lane 5, 100-fold molar excess of oligonucleotide S1. **B.** The collagen promoter fragment −477 to −332 was ^{32}P end-labeled on the non-coding strand and subjected to partial methylation and depurination. After incubation with either HeLa (lanes 1–9) or CTF (lanes 10–13) nuclear extract, the shifted bands were excised and processed for further analysis as described. The uncomplexed probe band (F) was also excised and processed as a control. **C.** Summary of methylation interference data. Protected guanines on the lower strand are indicated. **D.** Oligonucleotides P1 (lanes 1–6) and Sp1 (lanes 7–12) were ^{32}P-end-labeled and assayed for protein binding activity by incubation with 10 μg HeLa nuclear extract. Lanes 1 and 7, nuclear extract with no added competitor; lanes 2 and 8, 50-fold excess of oligonucleotide A1; lanes 3 and 9, 200-fold excess of oligonucleotide A1; lanes 4 and 10, 50-fold excess of oligonucleotide P1; lanes 5 and 11, 200-fold excess of oligonucleotide P1; lanes 6 and 12, 200-fold excess of oligonucleotide S1.



B MUTATIONS IN THE 274 BP FRAGMENT

A1:	AGGTCGGCCCCGCCCCGGCCCCATTG
MA1:	----------ATA----------
A2:	GTTGCCCGACACCTAGTGGCCGTCTGG
MA2:	----------GAAGCTCAA--------
A3:	TAGCGCGGTGGGAGTGGTTAGCTAA
MA3:	---------ATT----------

Figure 7. Diagrammatic representation of pCol-hGH derived plasmids and tabulation of relative transcriptional activity. **A.** The open and filled rectangles represent the first exon of the α1(I) procollagen gene and the hGH minigene, respectively. The start of transcription at base +1 is designated by the arrow. Unique Sst II sites at +292, +671, and +1440 in the procollagen first intron are indicated. The orientation of insertion of the intronic fragment between +671 and +1440 (lines 5–10) is indicated by the arrow. Site-specific mutations in the A1, A2, or A3 elements are indicated as MA1, MA2, and MA3, respectively. The transcriptional activity of each plasmid in CTF, relative to pCOL-hGH ± SEM and corrected for transcriptional efficiency, is tabulated. The number of independent determinations is given in parentheses. **B.** Mutations in the A1, A2, and A3 elements are shown. Dashes indicate that the sequence is the same as wild-type.

orientation-dependent enhancement relative to the intronless plasmid, consistent with previous reports (Bornstein et al., 1988; Fig. 7A, lines 5 and 6). No significant difference in transcriptional efficiency was observed with plasmids containing mutations in the A1 and A3 site (Fig. 7A, lines 7 and 8). These plasmids continued to show the orientation-dependent effect, which suggests that a modest enhancer activity resides elsewhere in the intronic sequence. Furthermore, these mutations do not abolish the orientation-dependent inhibition of transcription previously reported for the 274 bp fragment. These findings suggest that additional regulatory elements may be present that account for the inhibitory activity of the 274 bp fragment.

Discussion

We have previously reported that a 274 bp sequence of the human α1(I) collagen first intron, when inverted, was highly inhibitory to transcription of a collagen-hGH fusion gene in transient transfection experiments (Bornstein and McKay, 1988; Bornstein et al., 1988). When oriented positively, the 274 bp sequence was essentially neutral. Interestingly, this orientation-dependent 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 274 bp fragment, suggesting that both regions of the DNA were bound in the same complex. To account for these observations, we proposed that the orientation-specific inhibitory effect involved interactions of complex arrays of DNA-binding proteins which, when incorrectly positioned, would disrupt the normal spatial geometry of the transcriptional complex. It was therefore of interest to further characterize the protein-DNA binding elements within these two regions of the α1(I) collagen gene.

In the present study, we demonstrate that two regions contained within the 274 bp intronic sequence, A1 and A3, bind, with a similar affinity, Sp1 or an Sp1-like factor in nuclear extracts. Although the A1 and A3 regions are closely approximated, they do not bind Sp1 in a cooperative manner but instead display independent binding. We have also identified an Sp1-binding sequence in the promoter (bases −466 to −459; Bornstein et al., 1987) which competes for binding to the A1 and A3 sequences in the 274 bp fragment. However, the orientation-dependent inhibitory effect observed with this DNA fragment may not be due directly to protein–DNA interactions involving A1 or A3, since mutation

of these elements does not abolish the effect. Recently, two Sp1-binding sequences have been identified in the basal promoter of the mouse α1(I) collagen gene (Nehls et al., 1991). These Sp1 binding sequences overlap NF1 consensus sequences and were shown to bind Sp1 or NF1 in a mutually exclusive manner. Overexpression of Sp1 in NIH 3T3 fibroblasts reduced basal promoter activity by less than 2-fold, whereas overexpression of NF1 resulted in enhanced transcription from the procollagen promoter. Coexpression of NF1 and Sp1 resulted in inhibition of the enhancement observed with NF1 alone, indicating that Sp1 modulates expression of the collagen promoter by inhibiting the binding of NF1.

Previously, we reported that site-directed mutagenesis of a conserved AP1 motif abolished the transcriptional enhancement by the first intron, after transient transfection into CTF (Liska et al., 1990). Unlike the AP1 sequence, site-specific mutations that abolish the binding of nuclear factors to the A1 and/or A3 regions do not compromise the ability of the first intron to enhance transcription in a transient transfection assay. Instead, a 2-fold enhancement of activity was observed in CTF, suggesting that these sequences, and their respective binding factors, are involved in down-regulation of the α1(I) collagen gene. These data also indicate that the AP1 motif and the A1 and A3 regions of the 274 bp sequence regulate expression of the collagen gene by separate mechanisms. It seems likely that the extent to which mutations in the A1 and A3 regions augment transcription is limited by the high transcriptional activity of collagen genes in CTF. It is possible that the same mutations in other cells, or at other times in development in which other modulatory influences exist, could lead to higher degrees of transcriptional enhancement.

Alternatively, Sp1-binding sequences may not be involved in a direct interaction with the transcriptional complex in vivo but may influence expression of the gene by another mechanism, possibly by alteration of chromatin conformation. The 274 bp sequence containing A1 and A3 is the site of a constitutive DNase I-hypersensitive site in the first intron of the human α1(I) collagen gene (Barsh et al., 1984). The vast majority of DNase I-hypersensitive sites described thus far are gene-associated and appear to play a role in determining the position and frequency of transcriptional initiation or termination (Gross and Garrard, 1988; Elgin, 1988). Constitutive DNase I-hypersensitive sites have also been observed in association with inducible protein–DNA interactions (Herrera et al., 1989; Weih et al., 1990; Hapgood, 1991). Furthermore, the Sp1 binding sequences of the SV40 genome are known to be one important determinant in the induction of the DNase I-hypersensitivity of SV40 chromatin (Jongstra et al., 1984; Buchanan and Gralla, 1987; McKnight and Tjian, 1986). Thus, the Sp1-binding sequences in the α1(I) collagen intron may be important in the formation or function of the constitutive DNase I-hypersensitive site. This DNase I-hypersensitive site may be involved in the establishment of a transcriptionally active DNA conformation at the gene locus.

A number of observations suggest that the first intron is necessary for expression in vivo. Jaenisch and coworkers have shown that transcription of the α1(I) collagen gene is abolished by the insertion of a Moloney leukemia proviral gene into the first intron (Schnieke et al., 1983). Investigation into the mechanism of this inhibition has shown that the provirus prevents the developmentally regulated formation of a DNase I-hypersensitive site in the 5′-flanking sequence (Breindl et al., 1984). These observations have suggested that the transcriptional defect in the α1(I) collagen allele containing the proviral insertion is most likely due to disruption or displacement of essential cell-type or developmental-specific elements within the first intron of the α1(I) collagen gene (Chan et al., 1991; Barker et al., 1991). In line with these studies, an hGH minigene driven by 444 bp of promoter, the first exon, and the first intron of the human α1(I) collagen gene, was expressed in transgenic mice in the major collagen-producing tissues, but expression of the transgene was barely detectable when sequences within the first intron were deleted (Slack et al., 1991). Recently we have shown that a transgene driven by a longer (to −2300) collagen promoter, but lacking the first intron, is well expressed in the dermis of transgenic animals but fails to be expressed in dermal fibroblasts grown in cell culture (M. Reed, D. J. Liska, E. H. Sage, and P. Bornstein, unpublished data). Thus, the first intron of the α1(I) collagen gene may be involved in developmentally specific collagen expression.

Acknowledgments

We thank S. Jackson and R. Tjian for the gift of purified SP1 protein. We also thank J. L. Slack and P. Framson for a critical reading of the manuscript.

This work was supported by National Institutes of Health Grant AR11248. D. J. Liska was supported by National Research Service Award AR08095.

The costs of publishing this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC Section 1734 solely to indicate this fact.

References

G. M. Anderson and S. O. Freytag (1991), Mol Cell Biol 11, 1935–1943.

D. D. Barker, H. Wu, S. Hartung, M. Breindl, and R. Jaenisch (1991), Mol Cell Biol 11, 5154–5163.

G. S. Barsh, C. L. Roush, and R. E. Gelinas (1984), J Biol Chem 259, 14906–14913.

G. R. Bellomy and M. T. Record, Jr. (1990), Prog Nucleic Acid Res Mol Biol 37, 81–128.

S. Boast, M. W. Su, F. Ramirez, M. Sanchez, and E. V. Avvedimento (1990), J Biol Chem 265, 13351–13356.

P. Bornstein and J. McKay (1988), J Biol Chem 263, 1603–1606.

P. Bornstein, J. McKay, J. K. Morishima, S. Devarayalu, and R. E. Gelinas (1987), Proc Natl Acad Sci USA 84, 8869–8873.

P. Bornstein, J. McKay, D. J. Liska, S. Apone, and S. Devarayalu (1988), Mol Cell Biol 8, 4851–4857.

P. Bornstein and H. Sage (1989), Prog Nucleic Acid Res Mol Biol 37, 67–106.

M. Breindl, K. Harbers, and R. Jaenisch (1984), Cell 38, 9–16.

M. R. Briggs, J. T. Kadonaga, S. P. Bell, and R. Tjian (1986), Science 234, 47–52.

R. L. Buchanan and J. D. Gralla (1987), Mol Cell Biol 7, 1554–1558.

H. Chan, S. Hartung, and M. Breindl (1991), Mol Cell Biol 11, 47–54.

A. J. Courey, D. A. Holtzman, S. P. Jackson, and R. Tjian (1989), Cell 59, 827–836.

W. S. Dynan and R. Tjian (1986), Nature 316, 774–778.

S. C. R. Elgin (1988), J Biol Chem 263, 19259–19262.

D. Gidoni, J. T. Kadonaga, H. Barrera-Saldana, K. Takahashi, P. Chambon, and R. Tjian (1985), Science 230, 511–517.

D. S. Gross and W. T. Garrard (1988), Annu Rev Biochem 57, 159–197.

J. Hapgood, S. Cuthill, P. Soderkvist, A. Wilhelmsson, I. Pongratz, R. H. Tukey, E. F. Johnson, J. Gustafsson, and L. Poellinger (1991), Mol Cell Biol 11, 4314–4323.

R. E. Herrera, P. E. Shaw, and A. Nordheim (1989), Nature 340, 68–70.

S. P. Jackson, J. J. MacDonald, S. Lees-Miller, and R. Tjian (1990), Cell 63, 155–165.

K. A. Jones, J. T. Kadonaga, P. A. Luciw, and R. Tjian (1986), Science 232, 755–759.

J. Jongstra, T. L. Reudelhuber, P. Oudet, C. Benoist, C.-B. Chae, J.-M. Jeltsch, D. J. Mathis, and P. Chambon (1984), Nature 307, 708–714.

J. T. Kadonaga, A. J. Courey, J. Ladika, and R. Tjian (1988), Science 242, 1566–1570.

F. P. Lemaigre, S. J. Courtois, D. A. Lafontaine, and G. G. Rousseau (1989), Eur J Biochem 181, 555–561.

F. P. Lemaigre, D. A. Lafontaine, S. J. Courtois, S. M. Durviaux, and G. G. Rousseau (1990), Mol Cell Biol 10, 1811–1814.

R. Li, J. D. Knight, S. P. Jackson, R. Tjian, and M. Botchan (1991), Cell 65, 493–505.

D. J. Liska, J. L. Slack, and P. Bornstein (1990), Cell Regul 1, 487–498.

S. McKnight and R. Tjian (1986), Cell 46, 795–805.

M. C. Nehls, R. A. Rippe, L. Veloz, and D. A. Brenner (1991), Mol Cell Biol 11, 4065–4073.

S. A. Olsen, A. E. Geddis, and D. J. Prockop (1991), J Biol Chem 266, 1117–1121.

M. Ptashne (1988), Nature 335, 683–689.

R. A. Rippe, S. I. Lorenzen, D. A. Brenner, and M. Breindl (1989), Mol Cell Biol 9, 2224–2227.

C. M. S. Rossouw, W. P. Vergeer, S. J. du Plooy, M. P. Bernard, F. Ramirez, and W. J. de Wet (1987), J Biol Chem 262, 15151–15157.

J. D. Saffer, S. P. Jackson, and M. B. Annarella (1991), Mol Cell Biol 11, 2189–2199.

A. Schnieke, K. Harbers, and R. Jaenisch (1983), Nature 304, 315–320.

J. L. Slack, D. J. Liska, and P. Bornstein (1991), Mol Cell Biol 11, 2066–2074.

W. Su, S. Jackson, R. Tjian, and H. Echols (1991), Genes Dev 5, 820–826.

W. P. Tansey and D. F. Catanzaro (1991), J Biol Chem 266, 9805–9813.

F. Weih, F. Stewart, M. Boshart, D. Nitsch, and G. Schutz (1990), Genes Dev 4, 1437–1449.

C.-Y. Yu, J. Chen, L.-I Lin, M. Tam, and C.-K. J. Shen (1990), Mol Cell Biol 10, 282–294.