Volume 1 · Number 1 · April 1991
GENE EXPRESSION

Homology between α2(V) and α1(III) collagen promoters and evidence for negatively acting elements in the α2(V) first intron and 5′ flanking sequences

Daniel S. Greenspan,[1,2] Seung-Taek Lee,[2] Bum-Soo Lee,[2] and Guy G. Hoffman[1]

[1] Department of Pathology and Laboratory Medicine and [2] Program in Cell and Molecular Biology, University of Wisconsin, Madison, Wisconsin

We have isolated a 17 kilobase pair (kb) genomic clone containing the 5′ portion of the human α2(V) collagen gene. Nucleotide sequence was determined for 1671 base pairs (bp) comprising the promoter region, first exon and 334 bp of the first intron, and the major transcriptional start site determined by primer extension and S1 nuclease analysis. Sequence comparison revealed the α2(V) promoter to be similar in structure to the promoter of the α1(III) collagen gene. This is the first instance of such similarities between promoter regions of genes encoding different fibrillar collagen chains. Homology, in 5′ flanking sequences, extends upstream to about nucleotide −120 in each gene and is particularly striking near the TATTTA sequence (TATA box) present in each promoter. Some homology also surrounds the two transcription start sites. The 5′ untranslated regions of the two genes also show strong homology. Chimeric chloramphenicol acetyltransferase (CAT) constructs were prepared with various fragments from the 5′ portion of the α2(V) gene. Transient expression assays, in human fibroblasts, localized the functional α2(V) promoter to the region of 5′ flanking sequence conserved between the α2(V) and α1(III) genes. Expression assays also identified negatively acting elements, in intron and 5′ flanking sequences, which inhibit transcription from the α2(V) promoter.

A subset of the collagens are the fiber forming or fibrillar collagens. Partial characterization of a number of fibrillar collagen genes has demonstrated a conservation of intron/exon organization, denoting close evolutionary kinship (Ramirez et al., 1985). The promoter regions of these genes, however, have been found to be widely divergent. Even the genes encoding pro-α1(I) and pro-α2(I) chains, which are coordinately expressed to form the type I collagen heterotrimer, were found to have less than 4% homology in their promoter regions (Dickson et al., 1985). Mechanisms which coordinate expression of the various fibrillar genes, therefore, remain obscure.

Type V collagen is now grouped with types I-III and XI as a fibrillar collagen (Myers et al., 1985; Weil et al., 1987; Woodbury et al., 1989). The most widely distributed form of type V collagen is a heterotrimer comprised of one pro-α2(V) and two pro-α1(V) chains (reviewed in Fessler and Fessler, 1987). Type V collagen shares a number of characteristics with type III collagen. Both are apparently capable of forming thin homotypic fibrils, or of combining with type I collagen to form heterotypic copolymers

Received August 16, 1990; revision accepted November 13, 1990.

Correspondence: Daniel S. Greenspan, Department of Pathology and Laboratory Medicine and Program in Cell and Molecular Biology, University of Wisconsin, 1300 University Avenue, Madison, WI 53706 Tel(608)262-4676 Fax (608)262-2327.

© 1991 by the University of Health Sciences/The Chicago Medical School. All rights reserved. 1052-2166/91/0101/029-11$2.00

(Fessler and Fessler, 1987). In tissues in which the three fibrillar collagens are expressed, types V and III collagen are minor constituents in relation to the much more abundant type I (Smith et al., 1986). Types III and V are also similar in the degree and nature of their post translational modifications (Fessler and Fessler, 1987). Of the six fibrillar collagen genes assigned to specific chromosomal locations, only those encoding the pro-α1(III) and pro-α2(V) chains are located on the same chromosome (Emanuel et al., 1985; Weil et al., 1987), and they may be tightly linked (Tsipouras et al., 1988).

To further examine relationships between fibrillar collagen genes and mechanisms underlying their transcriptional control, we have isolated a 17 kb genomic clone containing the 5′ portion of the human α2 gene. We report the nucleotide sequence of the 5′ portion of the α2(V) gene, identify the major transcriptional start site used in dermal and corneal fibroblasts, and also report the first example of significant homology between promoter regions of two different fibrillar collagen genes: α2(V) and α1(III). Expression assays identified the functional α2(V) promoter and indicated that sequences upstream of the promoter and sequences near the junction of the first exon and first intron, act to inhibit expression.

Materials and methods

Isolation of RNA

Cytoplasmic RNA was isolated as previously described (Greenspan and Weissman, 1985). Poly(A^+) RNA was prepared from cytoplasmic RNA by means of the PolyATtract kit (Promega).

Cell culture

AH1F normal neonatal foreskin fibroblasts have been described (Lee et al., 1988). Corneal fibroblasts, obtained by mincing corneas from normal donor eyes, followed by six passages in Hams F12 medium with 10% fetal calf serum, were kindly provided by C. Brandt (University of Wisconsin–Madison).

Construction of α2(V)-CAT fusion plasmids

The pUC19-derived vectors pCAT-Basic (Promega, technical bulletin #80) and pCAT-Enhancer (Promega, technical bulletin #82) contain the bacterial chloramphenicol acetyltransferase (CAT) gene in a 773-bp TaqI fragment (Gorman et al., 1982). Downstream of the CAT gene are the simian virus 40 (SV40) sequences 4710–4100 and 2770–2533 which encode the small t splice site and early polyadenylation signal, respectively (Tooze, 1981). Immediately downstream of the SV40 early polyadenylation signal, pCAT-Enhancer also contains SV40 sequences 270–37 which include the SV40 transcriptional enhancer. Plasmids pGGH33 and pGGH34 were obtained by insertion of a 894 bp α2(V) ClaI-MluI fragment (Fig. 1), between the AccI and XbaI sites in the polylinkers of vectors pCAT-Basic and pCAT-Enhancer, respectively. The MluI end of the insert and XbaI ends of the vectors were joined by means of a synthetic adaptor, CGCGTGAGATCTGTCTAG, which was double stranded except for a 4 base overhang at each end. Cutting at the BglII site in the adaptor, blunt-ending with Klenow fragment, and then cutting with MluI, allowed insertion of a 233bp MluI-PvuII fragment into pGGH33 and 34 to give pGGH36 and 37, respectively. The 233 bp MluI-PvuII fragment, blunted with Klenow fragment, inserted, in the reverse orientation, between blunt ended HindIII and SalI sites in the polylinker upstream of the CAT gene in pCAT-Enhancer gave pGGH52. Removal of the 894 bp ClaI-MluI fragment from pGGH36 and 37 by cutting at the MluI site and at a PstI site in the vector polylinkers, followed by blunt ending with T4 polymerase and religation, gave pGGH38 and 39, respectively. Cutting a HindIII site in the polylinker just upstream of a α2(V) sequences in pGGH39, blunt ending with Klenow fragment, and insertion of a blunt ended 353 bp BsaJI-SacI fragment containing 20 bp of the first exon and 333 bp of the first intron, gave pGGH48 (+ orientation) and pGGH49 (− orientation). Insertion of the 894 bp ClaI-MluI fragment into the pGGH39 HindIII site, in the − orientation, gave pGGH57. An EcoRI site downstream of the CAT gene and SV40 enhancer in pGGH39 was cut and blunt ended with Klenow fragment. Insertion into this site of the blunt ended 894 bp ClaI-MluI fragment containing 5′ flanking sequences gave pGGH53 (+ orientation) and pGGH54 (− orientation). Insertion into the same site of pGGH39 with the blunt ended 353 bp BsaJI-SacI intron fragment gave pGGH55 (+ orientation) and pGGH56 (− orientation).

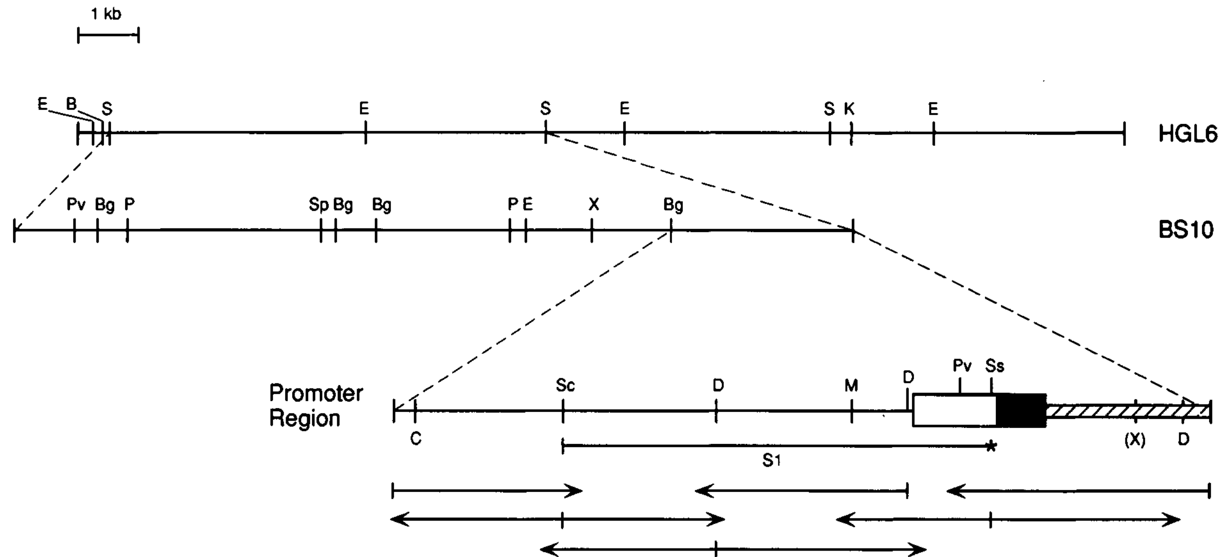

Figure 1. Restriction map of the 17 kb genomic clone and subclones containing the 5′ region of the human α2(V) collagen gene. E, EcoRI; B, BamHI; K, KpnI; S, SacI; Pv, PvuII; Bg, BglII; P, PstI; X, XbaI; Sp, SphI; C, ClaI; Sc, ScaI; D, DraI; M, MluI; Ss, SspI; (X), XbaI site overlapping a dam methylation site. White and black boxes denote untranslated and coding regions of the first exon, respectively. The cross hatched bar represents intron sequences. S1, ScaI-SspI fragment used as S1 probe; asterisk denotes ^{32}P-labeled 5′ end. Arrows indicate region and direction of sequencing.

DNA transfection and transient expression assays

AH1F cells (10^6 cells/10-cm dish) were transfected, by calcium phosphate precipitation (Lee et al., 1988), with 50 μg CAT construct, 10 μg salmon sperm DNA (Sigma), and 10 μg of the β-galactosidase plasmid pCH110 (Hall et al., 1983), used as an internal control for normalization of transfection efficiency. Nonchromatographic CAT assays (Seed and Sheen, 1988) were performed by incubating cell extracts 12 hr at 37°C with [^{14}C]chloramphenicol (Amersham) and n-butyryl Coenzyme A (Promega) followed by partitioning of butyrylated chloramphenicol into xylene and quantification by liquid scintillation counting. β-galactosidase activities were determined as described previously (Shen et al., 1987).

Results

Isolation and sequencing of human genomic clones

A 233 bp SspI-EcoRI fragment, corresponding to 5′ untranslated sequences, was isolated from a human α2(V) collagen cDNA clone (Greenspan et al., 1989), and used to screen a human leukocyte genomic library in EMBL3, purchased from Clontech. A 17 kb genomic clone (HGL6) containing the 5′ portion of the pro-α2(V) gene and about 6.9 kb of flanking sequences 5′ to the transcriptional start site was isolated. HGL6 was characterized by digestion with a number of restriction endonucleases and Southern blot analysis using the 233 bp cDNA fragment as probe. A 6.8 kb SacI fragment which hybridized to the probe was subcloned into the SacI site of pUC19 generating plasmid BS10. Partial restriction maps of HGL6 and BS10 are shown in Figure 1. Further restriction and blotting identified a 1,671 bp BglII-SacI fragment containing the 5′ portion of the α2(V) gene. This fragment was restricted, the fragments subcloned into M13mp18 and 19 to provide overlapping templates (see Fig. 1), and both strands of the 1,671 bp fragment sequenced in their entirety (Fig. 2) by dideoxy chain termination, as described (Lee et al., 1988).

Determination of transcription initiation sites

S1 nuclease analysis of cytoplasmic RNA from human diploid fibroblasts (AH1F) produced a 150 base protected fragment consistent with a transcription start site 158 bp upstream of the α2(V) translation initiation codon (Fig. 3A, lane 1). Primer extension of a 34 base long oligo-

```
-1081  AGATCTTGGCAAGTGTGAGGGTGGCACAAAAGTTCATCGATAATTTTCCTTAAAATGCATGTTCAACTGGGGCTATCATATTCCCTGCAAAGCTGGCTTA
 -981  AAGCTAATAATACCACCATCTGCTTTTATGTAGCTCTTTGCAATTTGGAGCTTCATTAATAATACATTACTCATTTTAGTGTCACTAAACCCTGTACACT
 -881  GGTCAGGGTAAAGCCCTTTCTTTTCTCAAATGAGAAAATTGGAGTGCACAGAGATAAAGGACTATATGCACATTTACAAAATTATTGAGTGGCAGAACCA
 -781  CTCACAAAAGGCAGAAGCAAGCCTTTTTTCATCCAAAGGTAGAGTACTTCCCATTCCTCTGCACTGCTTAGTAGACTATTTTTCAACATGGTAAAACTCT
 -681  CTGGGACAAGAGGAACCTGAAGATGTGGCCCCTTAACAAGAGATGTGGAGAAATTAGGAAAGGACATGCTTTCTTAATTCAAATTACAATGTATCACAAA
 -581  TTCATTTATCCAAAATGGTTTTCTAACCTGGACTTTACCTGGGCAACATTCTAATAATCTCCTTCCAGAGTATGGGAAAAGAAACCGAATCTGAGAAGGG
 -481  GATAAGAATACTCCCTATTCAGTCAAAAGAAAGGGTTAAAATGTGATTGTTTAAAGATTTAAGGTGGAGAAGAGAGTCTTGAATTATCTAAATGTAAAAA
 -381  AACTAGGGTAAAAGGCTGGATTTCCTTCGGAATGTCTTTTTCAGCTTATAGGGGAAAAGTTAAAGGGTGTGTGTCTGGGGGGAAGGGTTAGGGAGGGGAA
 -281  AATAAAACCCTTTTCTTCTAACAGCCTTTCTTAAAAAATAAAAGAAAAGAGAAACAAATGTTTGTTTCTGTTATTAAGCCCCATCTATGCTTAAAAGTTA
 -181  AATGAAATAGGGAAAGTTCAGGCACAGAGACGCGTGTTCTGATTTGGTTGTTTACCATCAATCAGACCGTTGCTTGGCAGACACTGGATGGTTATGAGCC
                                                                                     +|
  -81  TGAACAAGCTGAAAAGGGGCAGGAAAAGAAGTGGAGGCAGCATTCTTCCTATTTAAAGCTGCATCGCTTGAAAAAAGTTTCGCAGACTGTGCTGGAGCT
   20  GGTGCTGAAAAAGGGGGTTTGCAGAGGCTGCCCTGGGGCTGGTGCTGAAAGAAGAGCCCACAGCTGACTTCATGGTGCTACAATAACCTCAGAATCTACT
  120  TTTCACTCTCAGGAGAACCCACATGTCTAATATTTAGACATGATGGCAAACTGGGCGGAAGCAAGACCTCTCCTCATTCTTATTGTTTTATTAGGGCAAT
                                                         MetMetAlaAsnTrpAlaGluAlaArgProLeuLeuIleLeuIleValLeuLeuGlyGlnP
  220  TTGTCTCAATAAAAGCCCAGGAAGAAGACGAGGATG|GTGAGTTTGCCATTTGCTTTGCTTGTAGTGTTTATTGCACGGTTTGGGAAATAGGTGGAATGTT
       heValSerIleLysAlaGlnGluGluAspGluAsp
  320  AAGCCTCAGGAAGAGCGTGAAGAGGAGGTTCTGGTTTGAAGTTTGAGAAGGGGGCTGACGGATTTGCACCTGGAAAACGACTTGAAAGACAATTGGGATG
  420  TTTGTAAATCCGGAGCTTCTAGATCCACCGTTGGAGAGAGTACACTACTAAATTTTTCTGGGAAAGCCTACAATAAGTCCTTTTGGCTGAGTGTGAGAGT
  520  GGGACGTTTAAAATAAGAAAAGAGTGAATGCTTAAAACGGAGCACTAAAGGCATTTTAGTTAGAGGAGCTC
```

Figure 2. DNA sequence of the 5′ end of the human α2(V) collagen gene. Numbering, to the left, indicates nucleotide position relative to the major transcriptional start site which is designated +1. The predicted amino acid sequence is displayed below the corresponding nucleotide sequence of the first exon. The vertical arrow indicates the signal peptide cleavage site. A vertical line marks the junction of the first exon and first intron. Horizontal arrows underline indirect repeats. A dashed arrow, drawn above the nucleotide sequence, represents complementary sequences which comprise an oligonucleotide used in primer extension studies. Symbols over boxed sequences represent: open circle, viral core enhancer motif; closed circle, small open reading frames; open triangle, consensus sequence for Spl binding; closed triangle, TATA box; closed box, 20 bp direct repeats. Dotted lines over the nucleotide sequence indicate a potential NF-1 binding site. DNA sequences were analyzed using programs of the University of Wisconsin Genetics Computer Group (Devereux et al., 1984).

nucleotide probe (see Fig. 2) with reverse transcriptase produced a major cDNA product, 139 nucleotides in length. This result agrees with the S1 nuclease analysis and corresponds to a transcription start site assigned to a C (G in Fig. 3B, lanes X and X/2) 158 bases upstream of the AUG translational initiation codon of pro-α2(V). A minor cDNA product of 138 nucleotides may represent a secondary start site or be an artifact of the reverse transcriptase reaction (Laherty et al., 1989).

Since the content of type V collagen is high in cornea relative to other tissues in a variety of species (Birk et al., 1986), we examined the transcriptional start site of human corneal fibroblast RNA by S1 analysis (Fig. 3A, lane 2). Excision and liquid scintillation counting of bands indicated that the corneal fibroblasts contained approximately 6-fold the level of α2(V) RNA, as did dermal fibroblasts. However, the same major start site was apparently used in both types of fibroblast.

The 5′ end of a full-length pro-α2(V) cDNA previously isolated by this laboratory (Greenspan et al., 1989) lies 83 bp upstream of the major transcriptional start site identified in the present report. Since all sequences at the 5′ end of the cDNA clone are accounted for in and are collinear with genomic sequences, the cDNA clone is not the product of splicing involving an upstream noncoding exon, and is most probably the product of a rare transcript with an upstream start site. Since the cDNA clone had been isolated from a library synthesized from RNA of the SV80 line of SV40-transformed human fibroblasts, we examined the cytoplasmic RNA of SV80 cells by S1 analysis. It can be seen (Fig. 3A, lane 3) that the major transcriptional start site detected by S1 mapping was apparently the same as that utilized by normal dermal and corneal fibroblasts, supporting the view that the full-length α2(V) cDNA is probably the product of a rare transcript.

A second protected DNA fragment, produced in S1 nuclease analysis of cytoplasmic RNA (Fig. 3A, small arrow), is consistent with a minor transcriptional start site located 45 bp upstream of the major start site, but is still too far down-

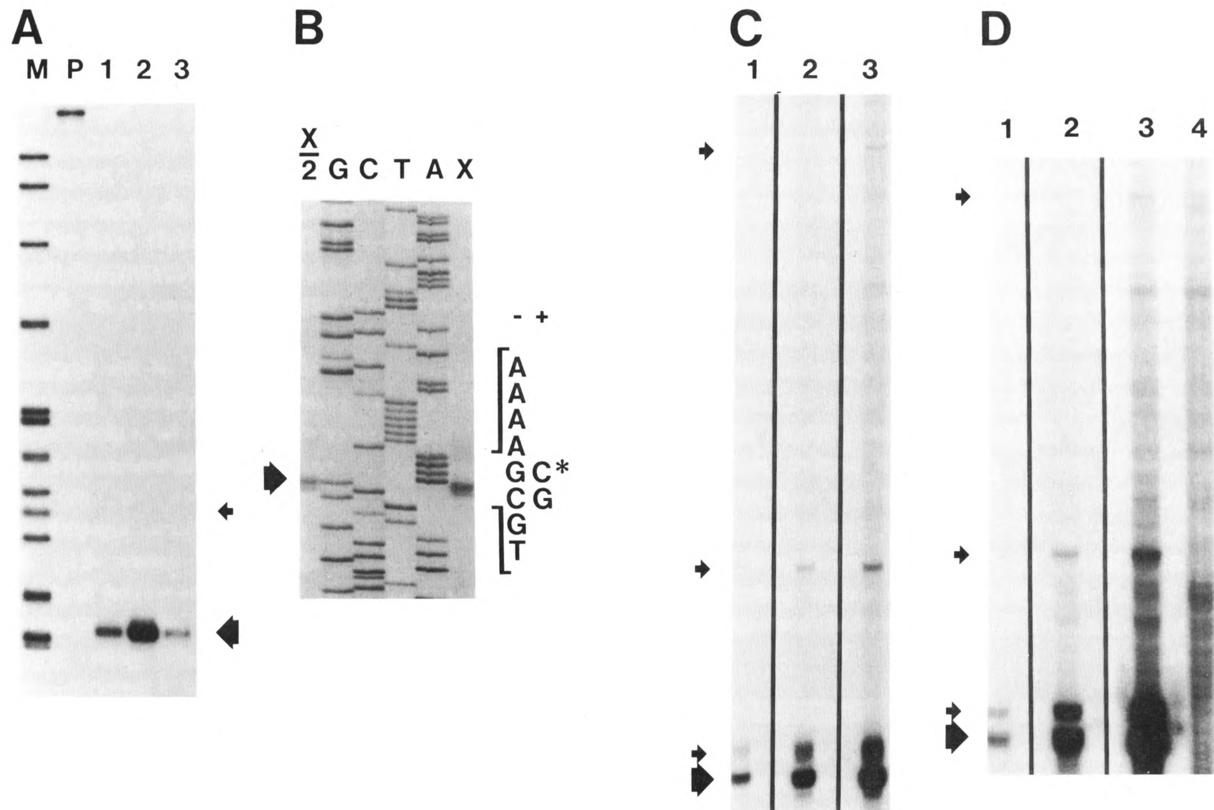

Figure 3. Determination of the transcription initiation site of the human α2(V) collagen gene. **A.** S1 nuclease analysis. Cytoplasmic RNA (25 μg) from AH1F cells (lane 1), corneal fibroblasts (lane 2) or SV80 cells was hybridized at 52°C to a 886 bp ScaI-SspI fragment from the 5′ end of the α2(V) gene (see Figure 1) and subjected to digestion with S1 nuclease as described (Lee et al., 1988). M, ^{32}P-labeled markers of MspI-digested pBR322. P, undigested probe. **B.** Primer extension analysis. A 5′ end labeled 34-base oligomer complementary to α2(V) sequences +106 to +139 (see Figure 2) was annealed, in 125 mM NaH_2PO_4, pH 6.75 and 125 mM EDTA as described (Laherty et al., 1989), to 25 μg (lane X) or 12.5 μg (lane X/2) cytoplasmic RNA, from AH1F fibroblasts. Extension with reverse transcriptase was as previously described (Greenspan et al., 1989). The same primer was employed for dideoxy sequencing (lanes G, C, T, A) in which the template was an α2(V) BglII-SspI fragment (−1081 to +150) subcloned into M13mp19. The asterisk indicates the transcription start site (+1). **C.** Primer extension of poly(A$^+$) RNA. The same 34-base primer was annealed to 10 μg of AH1F poly(A$^+$) RNA and extended as above. Autoradiograms of the same gel lane exposed for 1.5 h (lane 1), 5 h (lane 2) and 12.5 h (lane 3) are shown. **D.** RNase protection analysis of poly(A$^+$) RNA. The 886 bp ScaI-SspI fragment, used for S1 mapping, was cloned into the SmaI site of pBluescript II KS+ (Stratagene). The resultant plasmid was linearized at the SalI site and a uniformly ^{32}P-labeled, 982 base, riboprobe was generated by transcription with RNA polymerase T7. Conditions for annealing of the antisense probe (5 × 10^5 cpm per reaction) to 5 μg AH1F poly(A$^+$) RNA (lanes 1–3) or to 5 μg E. coli B RNA (Calbiochem) (lane 4) and incubation with RNases A and T1 were as described (Ausubel et al., 1987). Lanes 1–3 are autoradiograms of the same gel lane exposed for 0.25 h (lane 1), 1 h (lane 2), and 5.5 h (lane 3). Lane 4 was exposed for 5.5 h. Extension products and nuclease protected fragments were analyzed by electrophoresis on denaturing 6% polyacrylamide gels and are highlighted by arrows. DNA bands representing the major transcriptional start site are denoted by bold arrows.

stream to account for the previously described cDNA clone. In an attempt to detect other possible minor transcriptional start sites, primer extension (Fig. 3C) and RNase protection (Fig. 3D) analyses of poly (A$^+$) RNA from AH1F fibroblasts were performed. These studies confirmed the location of the major transcriptional start site and of the minor start site 45 bp upstream. In addition, longer exposures of primer extension and RNase protection gels detected a minor transcriptional start site, about 250 bp upstream of the major start site (lane 3 in Fig. 3, C and D), which could account for the previously described cDNA clone. Also detected

were heterogeneous bands, about 5–8 bases upstream of the band corresponding to the major start site (Fig. 3, C and D), which may represent additional start sites. Corresponding heterogeneous bands, not easily discerned in the S1 mapping gel, shown in Figure 3A, are more apparent in S1 mapping gels run for longer times (not shown).

Preliminary analysis of AH1F RNA by primer extension in a previous report (Greenspan et al., 1989) had suggested a major transcriptional start site upstream of that determined in the present study. However, we have found that annealing the oligonucleotide primer from the previous report with RNA under conditions (80% formamide, 56°C) more stringent than those used previously (80% formamide, 52°C) or in aqueous buffer, as described in the legend to Figure 3 in the present study, gives results consistent with those presented here (data not shown).

Sequence analysis of the α2(V) promoter region

Comparison of the genomic sequences presented here with human α2(V) cDNA sequences (D. S. Greenspan and G. G. Hoffman, unpublished data; Woodbury et al., 1989) shows that the first exon corresponds to 158 bases of 5′ untranslated sequence, 78 bases encoding the signal peptide, another 18 bases encoding mostly acidic residues which begin the globular domain of the N-terminal propeptide, and a G residue which terminates the exon and is the beginning of a split codon for Glu (Fig. 2). The DNA sequence of the 5′ flanking region, first exon and first 334 bp of the first intron were analyzed for the presence of previously described consensus sequences and for other structural features. An enhancer "core" element, similar to those found in viral enhancers (Khoury and Gruss, 1983) and in the first introns of the mouse α2(I) (Rossi and de Crombrugghe, 1987) and human α1(I) (Bornstein et al., 1987) collagen genes, was found in the α2(V) first intron (Fig. 2). Potential binding sites for NF-1 (23) are shown, as is a consensus sequence for Sp1 binding, although it is in coding sequences (Fig. 2). In addition, a 20 bp sequence between +12 and +31 was found as a direct repeat from +52 to +71, with only two 1 bp mismatches (Fig. 2). A number of small inverted repeats, with and without imperfections, are also represented in Figure 2.

Comparison of sequences in the promoter regions of the human α2(V), α1(I) (Chu et al., 1985) and α2(I) (Dickson et al., 1985) genes, showed them to be quite dissimilar except for a segment of homology around the translation initiation codons (data not shown). Conservation of segments surrounding the translation initiation codons of the α1(I), α2(I), and α1(III) collagen genes of both birds and mammals have been reported previously (Dickson et al., 1985; Yamada et al., 1983).

In contrast, when the 5′ ends of the human α2(V) and α1(III) genes were aligned, numerous homologies were evident in their promoter regions as well as in regions downstream of the major transcriptional start site, including but not limited to the segment surrounding the translation initiation codon (Fig. 4). In the promoter region one stretch of homology begins slightly upstream of a sequence (GGGCTGGAAAG) in the α1(III) gene, identified by others as a potential NF-1 binding site (Benson-Chanda et al., 1989), and extends downstream to include the TATA box sequence, TATTTA, found in both genes. Insertion of gaps, for optimal alignment of sequences, places the major transcriptional start sites of the two genes in close proximity (Fig. 4). The start site of α2(V) is positioned 27 bases downstream of the TATTTA box, while that of α1(III) is positioned 22 bases downstream of its cognate TATTTA sequence.

Expression analysis of the α2(V) promoter region

Promoter and enhancer activity of sequences from the 5′ end of the α2(V) gene were assayed by fusing various fragments to the bacterial CAT gene (Fig. 1) and testing for CAT activity after transfection of the fusion constructs into AH1F fibroblasts. Fusion plasmids pGGH33 and 34, which contain sequences upstream of, but not including, the area of α2(V)/α1(III) homology shown in Figure 4, show no promoter activity (Fig. 5). In sharp contrast, a 233 bp MluI-PvuII fragment, from −151 to +82, shows strong promoter activity in the presence of the SV40 enhancer (Fig. 5, pGGH39). The 233 bp fragment contains the region of α2(V)/α1(III) homology and only 26 bp of additional 5′ flanking sequences upstream. These results establish the importance of this region, which contains the conserved TATTTA box and a potential NF-1 binding site, to α2(V) transcription. The level

```
                -120                        -100                        -80
α2(V)     ATCAATCAGACCGTTGCTT....GGCAGACACTGGATGGTTATGAGCCTG
α1(III)   ATTTATCA.....TTTCTTTTACTGCTGA..GGGGATGGGTGCGGCTCTC
                                   -100                         -80
                            -60                        -40
          AACAAGCTGAAAAGGGGCAGGAAAAGAAGTGGAGGCAGCATTCTTCCTAT
          ATATTTCAG.AAAGGGGCTGGAAAGTGAGGGAAGCCAAACTTTTTCCTAT
                         -60                       -40
                  -20                         +1                  +20
          TTAAAGCTGCATCGCTTGAAAAAAGTTTTCGCAGACTGTGCTGGAGCTGG
          TTAAGGCCAAAGC........AAAGGAATCTCA...............GT
            -20                                +1
                              +40                       +60
          TGCTGAAAAAGGGGGTTTGCAGAGGCTGCCCTGGGGCTGGTGCTGAAAGA
          GGCTGA........GTTTTATGAC........GGGCCCGGTGCCTGAAGG
                                              +20
                 +80                   +100                    +120
          AGAGCCCACAGCTGACTTCATGGTGCTACAATAACCTCAGAATCTACTTT
          GCAGGGAACA....ACTTGATGGTGCTACTTTGAACTGCTTTTCTTTTCT
          +40                           +60                     +80
                           +140                      +160
          TCACTCTCAGGAGAACCCACATGTCTAATATTTAGACATGATG
          CCTTTTGCACCAAGAGTCTCATGTCTGATATTTAGACATGATG
                         +100                      +120
```

Figure 4. Comparison of the 5′ terminal sequences of the human α2(V) and α1(III) genes. The α2(V) sequences were compared to α1(III) sequences (2) by computer analysis. The TATA boxes and translation initiation codons are boxed, as are the major transcriptional start sites, which are also denoted by +1. ORFs, upstream of the collagen coding sequences, are enclosed by dashed boxes. An α1(III) sequence, previously identified by Benson-Chanda et al. (1989) as a potential NF-1 binding site, is underlined. Vertical lines indicate homologies between the two sequences; dots indicate gaps introduced for optimal alignment.

of CAT expression obtained with construct pGGH39 was comparable to that obtained with the pCAT-Control plasmid, which contains both SV40 promoter and enhancer. Promoter activity was orientation-dependent, since the 233 bp fragment showed no promoter activity when placed upstream of the CAT gene in the opposite orientation (pGGH52). In addition, an RNase protection experiment (Fig. 6A) shows that transcripts which initiate within the 233 bp MluI-PvuII fragment of pGGH39 utilize the same major start site as transcripts which initiate within endogenous α2(V) genes. The protected doublet contains a 357 base fragment (Fig. 6, left lane), which corresponds to transcripts initiating at the authentic α2(V) major start site identified in Figure 3. The above data indicate that the 233 MluI-PvuII fragment, from −151 to +82, contains the functional α2(V) promoter.

Plasmid pGGH37, which contains an α2(V) fragment with the same 3′ endpoint as that of pGGH39, but with an additional 894 bp of 5′ flanking sequences, from −1044 to −151, showed promoter activity in CAT assays, but at 6-fold lower levels than pGGH39 (Fig. 5). RNase pro-

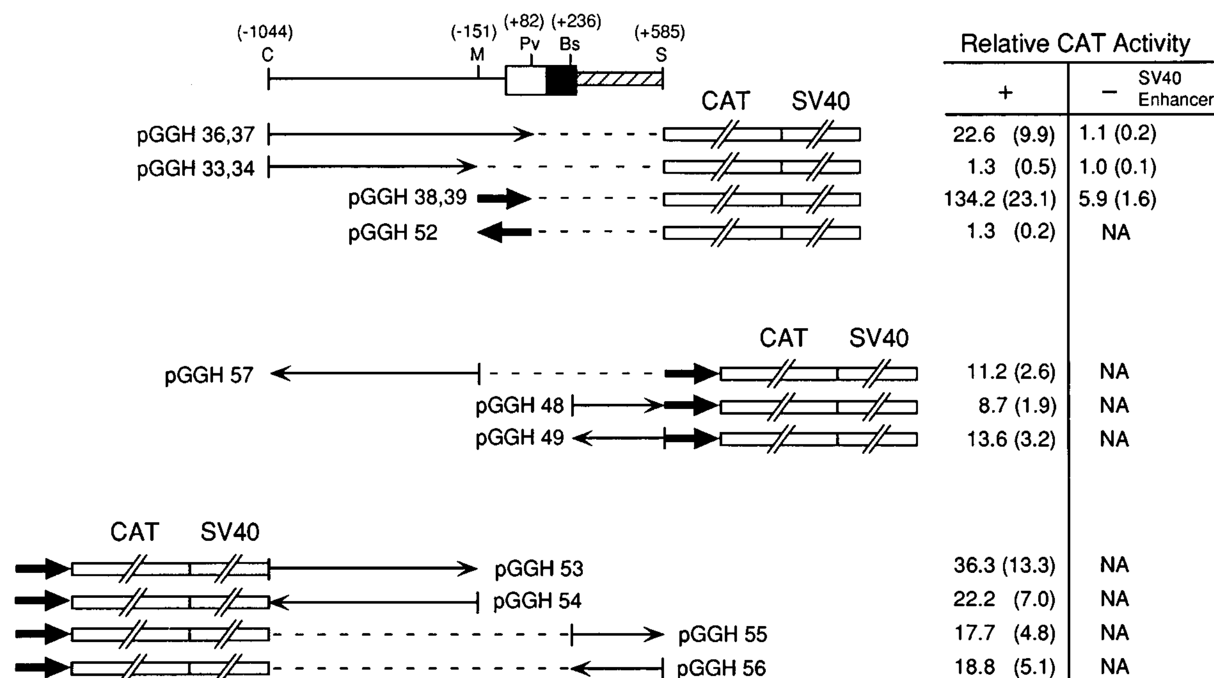

Figure 5. Promoter activities of sequences from the 5′ end of the human α2(V) gene. Arrows, beneath a partial restriction map of the 5′ portion of the α2(V) gene, denote the extent and orientation of α2(V) DNA fragments fused to the CAT gene for expression analysis. Bold arrows represent sequences, from −151 to +82, found to have strongest promoter activity. CAT activity of each plasmid is expressed relative to the conversion level achieved by the vector pCAT-control (100.0). pCAT-control (Promega, technical bulletin #81) is similar in configuration to pCAT-Enhancer but contains a 200 bp SphI-HindII fragment (128–5171) containing the SV40 early promoter, upstream of the CAT gene. (+) and (−) designate the presence or absence, respectively, of the SV40 enhancer downstream of the CAT gene in various plasmids. Data represent means ± standard deviation (in parenthesis) of from four to six independent transfection experiments. NA: not ascertained.

tection assays with RNA from pGGH37 transfected cells generated the same protected bands as assays with RNA from pGGH39 transfected cells, but at markedly lower levels (Fig. 6B). In three separate experiments involving transfection into AH1F cells and RNase protection analysis, densitometry of autoradiograms consistently found levels of pGGH37 transcripts to be 5- to 6-fold lower than levels of pGGH39 transcripts (data not shown), thus demonstrating a good quantitative correlation with CAT assay results. The same transcription start site used in pGGH39 is, therefore, used in the presence of the additional upstream sequences in pGGH37, but transcription levels are reduced. Thus, sequences from −1044 to −151 appear to inhibit transcription from the α2(V) promoter. Reversing the orientation of 5′ flanking sequences between −1044 and −151 relative to the α2(V) promoter (pGGH57), or placing these sequences downstream of the CAT gene in either orientation (pGGH53 and pGGH54) also reduced CAT expression levels relative to expression levels achieved with pGGH39 (Fig. 5). Negative regulatory elements have previously been reported upstream of the mouse α1(I) promoter (Rippe et al., 1989).

Both positive and negative transcription regulatory elements have been reported in the first introns of a number of collagen genes (Bornstein and McKay, 1988; Bornstein et al., 1987; Horton et al., 1987; Rippe et al., 1989; Rossi and de Crombrugghe, 1987). To determine how sequences within the first intron of the human α2(V) gene might affect transcription from the α2(V) promoter, a 353 bp BsaJI-SacI fragment containing 20 bp of the first exon and the first 333 bp of the first intron was inserted in both orientations upstream of the 233 bp promoter fragment (pGGH48 and 49) and downstream of the CAT gene (pGGH55 and 56). Placing the intron sequences upstream of the promoter in the direct (pGGH48) or indirect (pGGH49) orientation resulted in approximately 15- and

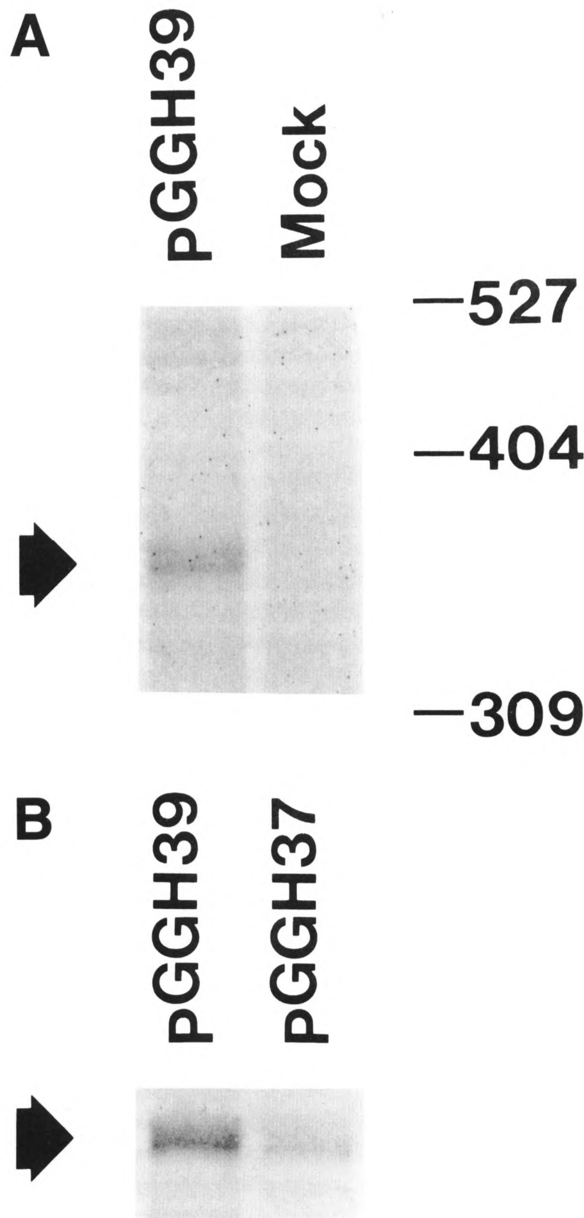

Figure 6. RNAse mapping of the transcription start site of pGGH39. A 519 bp HindIII-EcoRI fragment containing 233 bp of α2(V) sequences (–151 to +82), 248 bp of CAT sequences, and 38 bp of linker and adapter sequences was excised from pGGH39 and placed between the HindIII and EcoRI sites of pBluescript II KS+ (Stratagene). The resultant plasmid was linearized at the HindIII site and a 582 base riboprobe was generated and annealed to 10 μg of poly(A+) RNA from AH1F cells **A**, transfected with pGGH39 or mock-transfected or **B**, transfected with either pGGH39 or pGGH37 as indicated. Digestion with RNase and subsequent analysis was as described in the legend to Figure 3. Arrow indicates position of RNase-protected doublet. **A.** Numbers on right represent position and size (bp) of ^{32}P-labeled markers from MspI-digested pBR322.

10-fold reductions in expression, respectively. Placing the intron-containing fragment downstream of the CAT gene in the direct (pGGH55) or indirect (pGGH56) orientation resulted in 8- and 7-fold reductions in expression, respectively (Fig. 5). Thus, sequences from +236 to +585, which include 333 bp of the α2(V) first intron also appear to contain elements which inhibit transcription from the α2(V) promoter in a somewhat position- and orientation-independent manner.

Discussion

This study describes the isolation and characterization of the promoter region of the human α2(V) collagen gene. The major transcriptional start site was established by primer extension analysis and by RNase and S1 nuclease mapping. Thus, the predominant α2(V) mRNA species in fibroblasts have 5′ untranslated regions 158 bases long. Earlier estimates of a longer α2(V) 5′ untranslated region, based on primer extension (Greenspan et al., 1989) were, apparently, due to hybridization of α2(V)-specific primers to abundant non-α2(V) transcripts under conditions less stringent than those used in the present study.

Functional assays with α2(V)-CAT fusion constructs located elements upstream of the α2(V) promoter (between nucleotides –1044 and –151), and within sequences from +236 to +585, which include the first 333 bp of the α2(V) first intron, which inhibit transcription from the α2(V) promoter in a somewhat position- and orientation-independent manner. Sequences which inhibit transcriptional activity have previously been reported in the 5′ upstream sequences of mouse α1(I) (Rippe et al., 1989) and rat α1(II) (Savagner et al., 1990) collagen genes and in the first introns of human (Bornstein et al., 1987) and mouse (Rippe et al., 1989) α1(I) genes. Although sequences which inhibit transcription in the rat α1(II) gene, contain consensus motifs found in the transcriptional silencers of other genes (Savagner et al., 1990), silencer consensus motifs are not found in the α2(V) 5′ flanking or first intron sequences reported here. Transcriptional silencers are likely to be binding sites for nuclear factors, and some sequences which inhibit transcription in the rat α1(II) gene were shown to bind such factors (Savagner et al., 1990). Interestingly, a viral core enhancer motif in the hu-

man α1(I) first intron which, like the core enhancer motif found in the human α2(V) first intron, is located within sequences shown to inhibit transcription, has been shown to bind nuclear factors (Bornstein et al., 1987). Since cis-acting elements identified as transcriptional activators can also be involved in transcriptional inhibition (Shore and Nasmyth, 1987; Takimoto et al., 1989), it will be of interest to determine whether the enhancer motif in the α2(V) first intron binds nuclear factors and whether this motif is necessary for the inhibitory effect of the α2(V) intron sequences. Binding sites for nuclear factors will also be sought in α2(V) 5′ upstream sequences which inhibit transcription.

None of the α2(V)-CAT constructs produced significant CAT activity in the absence of the SV40 enhancer. This is in contrast to studies which have shown that the α1(I) collagen promoter, fused to the CAT gene, yields high levels of expression in the absence of heterologous enhancers (Bornstein et al., 1987; Rippe et al., 1989), and that this level of expression is increased in the presence of positive transcriptional elements from the α1(I) first intron (Bornstein and McKay, 1988; Rippe et al., 1989). Similarly, a transcriptional enhancer has been identified in the first intron of the rat α1(II) gene (Horton et al., 1987). It is possible that positive regulatory elements are present in the α2(V) sequences examined here, but that these are in close proximity to negative elements which cancel their effects in the functional assays (Bornstein and McKay, 1988); or that positive regulatory elements exist in α2(V) sequences, which flank those characterized in the present study. Alternatively, the relative weakness of the α2(V) promoter in functional assays, the relative paucity of consensus sequences for various transcription factors in the α2(V) sequence, and the apparent absence of enhancing elements and presence of elements which inhibit transcription in the α2(V) sequences examined here, may all reflect the low levels of α2(V) expression, in vivo, relative to expression levels of type I collagen genes. Clearly, further studies attempting to characterize α2(V) regulatory elements are warranted.

An important finding of the present study is the first instance of significant homology between promoters of different fibrillar collagen genes. This homology cannot be ascribed to a recent gene duplication event, since the α1(III) and α2(V) genes are thought to have evolved earlier than other fibrillar collagen genes with the exception of the α1(XI) gene (Bernard et al., 1988). The promoter regions of the more recent fibrillar collagen genes share little, or no, homology. Thus, the conserved similarities between the α1(III) and α2(V) promoters suggest that the two genes may be regulated by common transcription factors. Consistent with this possibility, transient expression assays involving chimeric α2(V)-CAT plasmids, established the importance of this conserved region to α2(V) transcription. Coordinated transcription of the two genes would be consistent with similarities in levels and tissue distribution of expression of collagen types III and V. The fact that other fibrillar collagen genes which are coordinately expressed, such as the α1(I) and α2(I) genes, do not show similarities in their promoter regions implies different mechanisms for the coordinate transcription of the different sets of fibrillar collagen genes. In this regard, it is of interest that the α2(V) and α1(III) genes are the only fibrillar collagen genes yet shown to be syntenic (Emanuel et al., 1985; Weil et al., 1987), and that they may be closely linked (Tsipouras et al., 1988). This unique clustering of the two collagen genes may also reflect mechanisms for coregulation distinct from mechanisms coordinating expression of the other fibrillar collagen genes.

Recently, two cis-acting elements important to transcriptional activity have been identified in the mouse α1(III) promoter (Ruteshouser and de Crombrugghe, 1989). Both elements lie within the region of homology between the α1(III) and α2(V) promoters identified in the present study, but these elements contain sequences which are somewhat divergent compared to corresponding α2(V) sequences. Parallel studies of α1(III) and α2(V) promoters, in a variety of systems, should clarify how differences and similarities in the two promoters reflect differences and similarities in expression of the two genes.

Acknowledgments

We thank Dr. Burton Goldberg for critical reading of the manuscript, and Ellen Pluim for excellent technical assistance.

This work was supported by National Institutes of Health Grant GM-38544 (D. S. G.) and by a Grant-

in-Aid from the American Heart Association (D. S. G.). S.T. Lee is a predoctoral fellow of the March of Dimes.

The nucleotide sequence data reported in this paper have been submitted to GenBank and have been assigned accession number M58529.

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

F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (1987), Current Protocols in Molecular Biology. Greene Publishing Associates and Wiley Interscience, John Wiley and Sons, New York.

V. Benson-Chanda, M.-W. Su, D. Weil, M.-L. Chu, and F. Ramirez (1989), Gene 78, 255–265.

M. Bernard, H. Yoshioka, E. Rodriguez, M. van der Rest, T. Kimura, Y. Ninomiya, B. R. Olsen, and F. Ramirez (1988), J Biol Chem 263, 17159–17166.

D. E. Birk, J. M. Fitch, and T. F. Linsenmayer (1986), Invest Ophthalmol Vis Sci 27, 1470–1477.

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

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

M.-L. Chu, W. de Wet, M. Bernard, and F. Ramirez (1985), J Biol Chem 260, 2315–2320.

J. Devereux, P. Haeberli, and O. Smithies (1984), Nucl Acids Res 12, 387–395.

L. A. Dickson, W. de Wet, M. Di Liberto, D. Weil, and F. Ramirez (1985), Nucl Acids Res 13, 3427–3438.

B. S. Emanuel, L. A. Cannizzaro, J. M. Seyer, and J. C. Myers (1985), Proc Natl Acad Sci USA 82, 3385–3389.

J. H. Fessler and L. I. Fessler (1987), in Structure and Function of Collagen Types (R. Mayne and R. E. Burgeson, eds.), Academic Press, Orlando, FL, pp. 81–103.

C. M. Gorman, L. F. Moffat and B. H. Howard (1982), Mol Cell Biol 2, 1044–1051.

D. S. Greenspan, G. G. Hoffman, and B. S. Lee (1989), J Biol Chem 264, 20683–20687.

D. S. Greenspan and S. M. Weissman (1985), Mol Cell Biol 5, 1894–1900.

C. V. Hall, P. E. Jacobs, G. M. Ringold, and F. Lee (1983), J Mol Appl Genet 2, 101–109.

W. Horton, T. Miyashita, K. Kohno, J. R. Hassell, and Y. Yamada (1987), Proc Natl Acad Sci USA 84, 8864–8868.

G. Khoury and P. Gruss (1983), Cell 33, 313–314.

C. D. Laherty, T. M. Gierman, and V. M. Dixit (1989), J Biol Chem 264, 11222–11227.

S.-T. Lee, B. D. Smith, and D. S. Greenspan (1988), J Biol Chem 263, 13414–13418.

J. C. Myers, H. R. Loidl, J. M. Seyer, and A. S. Dion (1985), J Biol Chem 260, 11216–11222.

F. Ramirez, M. Bernard, M.-L. Chu, L. Dickson, F. Sangiorgi, D. Weil, W. de Wet, C. Junien, and M. Sobel (1985), Ann NY Acad Sci 460, 117–129.

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

P. J. Rosenfeld and T. J. Kelly (1986), J Biol Chem 261, 1398–1408.

P. Rossi and B. de Crombrugghe (1987), Proc Natl Acad Sci USA 84, 5590–5594.

E. C. Ruteshouser and B. de Crombrugghe (1989), J Biol Chem 264, 13740–13744.

P. Savagner, T. Miyashita, and Y. Yamada (1990), J Biol Chem 265, 6669–6674.

B. Seed and J.-Y. Sheen (1988), Gene 67, 271–277.

R.-F. Shen, Y. Li, N. Sifers, H. Wang, C. Hardick, S. Tsai, and S. Woo (1987), Nucl Acids Res 15, 8399–8415.

D. Shore and K. Nasmyth (1987), Cell 51, 721–732.

L. T. Smith, K. A. Holbrook and J. A. Madri (1986), Am J Anat 175, 507–521.

M. Takimoto, J. P. Quinn, A. R. Farina, L. M. Staudt, and D. Levens (1989), J Biol Chem 264, 8992–8999.

J. Tooze (1981), DNA Tumor Viruses, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

P. Tsipouras, R. C. Schwartz, A. C. Liddell, C. S. Salkeld, D. Weil, and F. Ramirez (1988), Genomics 3, 275–277.

D. Weil, M. Bernard, S. Gargano, and F. Ramirez (1987), Nucl Acids Res 15, 181–197.

D. Woodbury, V. Benson-Chanda, and F. Ramirez (1989), J Biol Chem 264, 2735–2738.

Y. Yamada, M. Mudryj, and B. de Crombrugghe (1983), J Biol Chem 258, 14914–14919.