Characterization of the Promoter for the α1(IV) Collagen Gene

DNA SEQUENCES WITHIN THE FIRST INTRON ENHANCE TRANSCRIPTION*

(Received for publication, November 5, 1987)

Paul D. Killen‡, Peter D. Burbelo, George R. Martin, and Yoshihiko Yamada

From the Laboratory of Developmental Biology and Anomalies, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland 20892

Two overlapping clones spanning 19 kilobase pairs (kb) of the 5' end of the α1(IV) collagen gene were isolated and found to contain a single exon which encoded the 5'-untranslated sequence and 84 base pairs of the signal peptide. The 5' end of this exon was determined to be the 5' end of the transcript by S1 nuclease protection and primer extension. The nucleotide sequence of 1 kb of the 5'-flanking DNA was extremely G + C-rich (>70%) and contained two GC boxes and a putative cAMP regulatory sequence. The transcriptional regulation of the α1(IV) gene was studied with chimeric gene constructs utilizing 2.5 kb of the 5'-flanking sequence coupled to the gene for chloramphenicol acetyltransferase. Transfection of this construct into differentiating F9 cells resulted in low chloramphenicol acetyltransferase activity compared to β-actin or Rous sarcoma virus long terminal repeat promoters, although these cells produce large amounts of collagen IV. Inclusion of a 2.7-kb sequence 2.3 kb downstream from the first exon in either orientation increased the transcription of the chloramphenicol acetyltransferase construct approximately 10-fold in F9 cells, but was not active in NIH 3T3 cells, which synthesize little collagen IV. These results indicate the presence of an enhancer within the first intron, which increases the expression of this gene.

Collagen IV, a major structural component of basement membranes, is a heteropolymer, composed of two α1(IV) chains and one α2(IV) chain. The molecule is approximately 450 nm in length and contains a globular C-terminal domain (NC1) and a disulfide-rich N-terminal domain (7S), which cross-link to other collagen IV molecules to form an end-to-end-linked network in the matrix (1). Collagen IV provides a scaffolding for basement membrane components, such as laminin (2, 3), fibronectin (4), and heparan sulfate proteoglycan (5). In addition, several studies also indicate that a variety of cells, including hepatocytes (6), endothelial cells (7), and some tumor cells (8–10), bind to collagen IV.

Collagen IV plays an important role in cell migration and differentiation (11). For instance, during differentiation of parietal endoderm from the inner cell mass, there is a marked increase in the biosynthesis of several basement membrane components including laminin and collagen IV (12, 13). One frequently studied model of this process is the differentiation of F9 teratocarcinoma cells exposed to retinoic acid and dibutyryl cAMP, which show a rapid and coordinated expression of the genes for basement membrane proteins including laminin and collagen IV chains (14). Most models of gene regulation implicate control regions at the 5' end of the gene. For this reason, we have isolated murine genomic clones coding for the amino-terminal portion of the α1(IV) collagen chain. Here we describe the characterization of the clones containing the first exon and the 5'-flanking sequence of the α1(IV) chain gene of collagen IV. We have also begun to identify the sequences necessary for expression of this gene by transfecting plasmids, which contain potential regulatory sequences of this gene coupled to chloramphenicol acetyltransferase, into both undifferentiated and differentiated F9 cells. Our results indicate that noncoding DNA sequences located within 5 kb downstream of the first exon of the α1(IV) gene are required for the expression of the α1(IV) collagen gene.

EXPERIMENTAL PROCEDURES

Materials—Restriction enzymes were purchased from Bethesda Research Laboratories and New England Biolabs (Beverly, MA). Polynucleotide kinase and T4 DNA ligase were purchased from Pharmacia LKB Biotechnology Inc. [γ-32P]dATP and α-32P-labeled dATP were purchased from Amersham Corp. Avian myeloblastosis virus reverse transcriptase was purchased from Seikagaku Chemical Corp. Ltd. (St. Petersburg, FL).

Identification of First Exon—A 140-bp fragment from the 5' end of a cDNA to the N-terminal portion of the murine α1(IV) chain mRNA† was used as probe to obtain two overlapping genomic clones from a murine genomic library (the generous gift of P. Leder, Harvard University). A single exon was identified within these genomic clones by hybridization of the cDNA probe in Southern blots and by sequence analysis (see Fig. 1). The nucleotide sequence of a 1.5-kb HindIII-XhoI fragment containing the putative first exon was obtained by subcloning random fragments produced by sonication in M13 phage and sequencing the single-stranded phage DNA by the dideoxy chain termination method with universal primers (15). Ambiguous sequences were resolved by sequencing plasmid DNA with synthetic oligonucleotides and avian myeloblastosis virus reverse transcriptase (16). Residual ambiguities were resolved by chemical degradation methods (see Fig. 2).

The 5' boundary of the exon was determined by S1 nuclease mapping performed according to the method of Berk and Sharp (17). A NcoI-XhoI fragment 5' end-labeled at the NcoI site (106 dpm) was precipitated with 10 μg of RNA or poly(A*) RNA from differentiated F9 cells (14); dissolved in 80% formamide, 0.4 M NaCl, 40 mM PIPES, pH 6.4, 1 mM EDTA; denatured at 80 °C for 10 min; and hybridized at 59 °C. After 3 h, three equal aliquots were diluted 10-fold in ice cold S1 buffer (0.25 M NaCl, 30 mM potassium acetate, pH 4.6, 1 mM PIPES, 1.4-piperazinediethanesulfonic acid).

The abbreviations used are: kb, kilobase pair(s); bp, base pair(s); PIPES, 1,4-piperazinediethanesulfonic acid.

† Present address: Dept. of Pathology, University of Michigan, Ann Arbor, MI 48109.

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

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J03944.

* The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J03944.
**α(IV) Collagen Enhancer**

**TABLE I**

Structure and activity of the α(IV) collagen gene promoter and first intron-chloramphenicol acetyltransferase

DNA fragments from the promoter and first intron were subcloned into the promoterless pSV0CAT-X plasmid. These constructs were transfected into F9 and NIH 3T3 cells as described under "Experimental Procedures" and chloramphenicol acetyltransferase (CAT) activity was determined 48 h later. The values shown are percentages of total chloramphenicol converted to acetylated derivatives. This percentage was calculated from data obtained by counting radioactive material in the appropriate regions of the thin-layer chromatograms. The results shown are means ± S.E. of two separate experiments done in duplicate. The level of β-actin-chloramphenicol acetyltransferase expression in F9+ and 3T3 cells was 42.47 ± 1.24 and 37.63 ± 0.32, respectively.

| PROMOTER | FIRST INTRON | % ACTIVITY |
|----------|--------------|------------|
| SV0CAT-X | CAT          | 0.50 ± .10 | .46 ± .23 |
| P47      | CAT          | 0.94 ± .12 | .43 ± .25 |
| P47A     | CAT          | 1.19 ± .11 | -          |
| P48      | CAT          | 9.56 ± .22 | .43 ± .26 |
| P52      | CAT          | 9.75 ± .27 | .65 ± .23 |
| P54      | CAT          | 1.13 ± .16 | .35 ± .19 |
| P56      | CAT          | 11.31 ± .74| .39 ± .13 |
| P57      | CAT          | 11.44 ± .63| -          |

ZnSO₄ containing 100, 500, or 1000 units/ml S1 nuclease. After incubation at 37 °C for 30 min, the digestion was stopped by phenol extraction, and the DNA fragments were precipitated.

To identify the 5' end of the murine α(IV) transcript, an analytical primer extension experiment was performed using an end-labeled synthetic oligonucleotide (5'-CATGGTGGCGCGCCCGGGG-3') corresponding to the 5' 20 bp of the genomic fragment used for S1 nuclease mapping described above. The oligonucleotide primer was annealed with 2 μg of poly(A) from differentiated F9 cells, and the primer was extended with 10 units of avian myeloblastosis virus reverse transcriptase at 43 °C in 50 mM Tris-HCl, pH 8.3, 80 mM NaCl, 40 mM KCl, 6 mM MgCl₂, 10 mM dithiothreitol. After 1 h, the reaction was stopped by phenol extraction, and the reaction products were precipitated with ethanol.

The DNA fragments from S1 nuclease digestion and primer extension were resuspended in 80% formamide; denatured at 90 °C; and analysed in parallel on a 6% polyacrylamide, 7 M urea gel with products of chemical degradation of the end-labeled NcoI-XbaI fragment used in sequencing (18).

Chimeric Gene Construction—To subclone the 5'-flanking sequences of the murine α(IV) transcript, pSV0CAT (19) was linearized at the HindIII site, the site was filled in with the Klenow fragment of DNA polymerase, and the product was ligated with XhoI linkers (designated pSV0CAT-X). All subsequent constructs were obtained in this derivative of pSV0CAT. To obtain α(IV) promoter chloramphenicol acetyltransferase constructs, a plasmid containing the 2.7-kb BamHI-XhoI genomic fragment was linearized at the unique NcoI site present within the first exon. Brief digestion with Bal31 exonuclease was used to remove the translation initiation site and a short portion of the 5'-untranslated sequence. Fragments containing the residual first exon and approximately 2.5 kb of the 5'-flanking sequence were excised with BamHI, the ends were polished with S1 nuclease and the Klenow fragment of DNA polymerase, and XhoI linkers were attached. Following removal of excess linkers, the fragments were cloned in pSV0CAT-X. The orientation of the genomic fragments and the extent of Bal31 deletion were determined by nucleotide sequencing. One of these constructs (p47) containing the sequence from positions 2500 to 74 was selected for further constructs. A truncated promoter construct (at positions 715 to 74), p47A, was prepared by cutting p47 with XbaI, filling in the ends with Klenow fragment, and attaching XhoI linkers. Following removal of excess linkers, the 0.8-kb fragment was subcloned in pSV0CAT-X, and its orientation was determined as described above.

To assess the role of regulatory elements within the first intron, various restriction fragments were subcloned into the unique BamHI site located 3' to the chloramphenicol acetyltransferase gene. Derivatives of p47 and p47A containing various restriction fragments from the first intron of the gene are listed in Table I. A 5-kb XhoI-EcoRI
fragment was subcloned into p47 and p47A after attachment of BamHI linkers and designated p48 and p52, respectively. A 2.3-kb XhoI-XbaI fragment and a 2.7-kb XbaI-EcoRI fragment were prepared similarly and subcloned in both orientations in the BamHI site of p47A. All plasmids used for transfection were banded by equilibrium centrifugation in cesium chloride density gradients three times.

**RESULTS**

**Identification of First Exon**—We previously isolated a 530-bp cDNA clone for the murine α1(IV) chain from a library constructed by specific primer extension utilizing poly(A+) RNA from differentiated F9 teratocarcinoma cells. This cDNA encodes 141 bp of the 5'-untranslated sequence, a signal peptide, and a portion of the N-terminal cross-linking domain. We used a 140-bp fragment of this cDNA to obtain two distinct but overlapping genomic clones which contained the first exon and 17 kb of the 3'-flanking intron sequence. The nucleotide sequence of this exon agrees perfectly with the first 225 bp of the cDNA up to the typical splice donor consensus sequence. This exon included portions of the 5'-untranslated sequence and 84 bp coding for the first 28 amino acids of the protein (Fig. 1). By using a 128-bp fragment of the α1(IV) cDNA as probe, a third genomic clone containing the second exon was isolated. This clone contained approximately 13 kb of the 5'-flanking sequence which showed no overlap with the two genomic clones containing exon 1, suggesting that the first intron is at least 30 kb.

S1 nuclease protection and primer extension were utilized to identify the 5' boundary of the exon and the transcription initiation site(s) of the gene. For S1 nuclease analysis, a 368-bp NcoI-XbaI genomic fragment 5' end-labeled at the NcoI site was annealed to mRNA isolated from F9 teratocarcinoma cells treated with retinoic acid and dibutyryl cAMP and digested with different concentrations of S1 nuclease. Two fragments at +1 and -184 bp were protected even at high S1 nuclease concentrations (Fig. 2, lower, lanes 5–7). As expected, no protected fragments were observed when the genomic fragment was annealed with tRNA (Fig. 2, lower, lanes 2–4). Both the S1 nuclease-protected fragments and the primer extension products showed microheterogeneity around the initiation sites. The absence of primer-extended products longer than those predicted by S1 nuclease analysis excludes the possibility that S1 nuclease analysis was detecting intron boundaries rather than transcription initiation sites.

**Nucleotide Sequence of Promoter**—The nucleotide sequence of the 5'-flanking region of the murine α1(IV) gene is shown in Fig. 3. The major transcription initiation sites are shown. The 5'-flanking DNA has a high G + C content (75%). DNA sequence analysis identified numerous inverted repeats clustered close to exon 1. GC boxes, i.e. GGGCGG and CCGCC, occur at positions -4 and -63. Other repeats such as the sequence from positions -301 to -316 showed an inverse...
**DISCUSSION**

The genes coding for the chains of collagen IV are large and complex. Since some of the introns of these genes are

---

3 P. D. Killen, unpublished data.
very large (>30 kb), we decided to obtain the promoter region in genomic clones utilizing a specific cDNA probe representing the 5' portion of the a1(IV) chain mRNA. We obtained overlapping genomic clones that contains 19 kb of sequence including 2.5 kb of the 5'-flanking region, the 234 bp of the first exon, and at least 15 kb of the first intron. Primer extension and S1 mapping were used to confirm that we had isolated the first exon of the murine a1(IV) collagen gene and the promoter for the gene. About 1.5 kb of genomic DNA containing the first exon were sequenced. Unlike the promoters for the interstitial collagens (23-26), the a1(IV) collagen gene does not contain a TATA box. The lack of a TATA box may explain the occurrence of multiple transcription initiation sites observed from primer extension and S1 protection experiments since this sequence is important for the precise localization of the RNA polymerase on the promoter (27). Three CAAT boxes were found on the noncoding strand within the promoter for the a1(IV) collagen gene and could be of functional significance. Several other genes, including Rous sarcoma virus (28), hsp 70 (29), and the a2(I) chain of Rous sarcoma virus (28), hsp 70 (29), and the a2(I) chain of hypoxanthine gsh 70 (30), hydroxymethylglutaryl-CoA reductase (31), and the epidermal growth factor receptor (32). The a1(IV) promoter also contains two SP1 binding regions. In addition, this promoter contains a sequence at position -368 which matches the cAMP consensus sequence, found necessary for the CAMP modulation of the phosphoenolpyruvate carboxykinase (22), somatostatin (34), and fibronectin (35) genes. It should be noted that cAMP increases collagen IV synthesis for CAMP, which suggests that the basement membrane proteins laminin, fibronectin, and collagen IV may be regulated by CAMP.

To test the putative promoter for transcribing activity, chimeric chloramphenicol acetyltransferase expression plasmids containing the a1(IV) promoter region were transfected into F9 teratocarcinoma and NIH 3T3 cells. The studies showed that the 5'-flanking region of the a1(IV) gene acts as a very weak promoter, not unlike the observations made with the promoter of the a2(I) (37), a1(II) (38, 39), and a1(II) (40) collagen chain genes. Constructs with portions of the first intron plus the promoter coupled to the chloramphenicol acetyltransferase gene were prepared, and these showed a much higher level of transcriptional activity in teratocarcinoma cells, but not in NIH 3T3 cells. The active portion of the intron was localized to a 2.7-kb fragment 2.3 kb downstream from the first exon. This fragment was active when placed in either orientation in the construct, suggesting that it could be an enhancer element. Since enhancers have been identified in the first intron in the alpha(I) and alpha(II) chain genes, this regulatory element could be a general feature of collagen genes.

Interestingly, the transcriptional activity of the active construct was as high in undifferentiated as in differentiated F9 teratocarcinoma cells, which differ greatly in the amount of collagen IV they synthesize. Possibly, undifferentiated F9 cells produce factors which bind to the enhancer region necessary for the high level of transcription. The endogenous gene may fail to respond to these factors, perhaps because of altered chromatin structure and/or methylation state (41, 42). Retinoic acid and CAMP treatment do induce changes in chromatin structure and methylation pattern in the regulatory region of the gene which allow for the high levels of expression. As shown here, the collagen IV gene contains a number of potential regulatory regions. Alternatively, as in other genes, negative regulatory factors may limit the transcription of collagen IV genes in undifferentiated F9 cells, and these are not present in the portions of the gene used in our constructs. The identification of such elements as well as the detailed characterization of the regulatory regions described here should help clarify the factors controlling the synthesis of collagen IV.

REFERENCES

1. Kuhn, K., Glanville, R. W., Babel, W., Qian, R. Q., Dieringer, H., Voss, T., Siebold, B., and Oberhauser, I. (1986) Ann. N. Y. Acad. Sci. 44-54
2. Timpl, R., Rohde, H., Bogue, P., Benders, S., Laidt, J.-M., and Gabrilove, J. L. (1979) J. Biol. Chem. 254, 5653-5659
3. Timpl, R., Engel, J., and Martin, G. R. (1985) Trends Biochem. Sci. 207-209.
4. Yamada, K. M. (1983) Annu. Rev. Biochem. 2, 71-79.
5. Laurie, G. W., Bing, J. T., Kleinman, H. R., Hassell, J. R., Aumailley, M., Retinoic acid, and CAMP treatment did induce changes in chromatin structure and methylation pattern in the regulatory region of the gene which allow for the high levels of expression. As shown here, the collagen IV gene contains a number of potential regulatory regions. Alternatively, as in other genes, negative regulatory factors may limit the transcription of collagen IV genes in undifferentiated F9 cells, and these are not present in the portions of the gene used in our constructs. The identification of such elements as well as the detailed characterization of the regulatory regions described here should help clarify the factors controlling the synthesis of collagen IV.

4. Y. Yamada, manuscript in preparation.

5. Graham, F. L., and Van der Eb, A. I. (1973) Virology 52, 456-467.
6. Weiler, H., König, H., and Grom, P. (1985) Science 219, 626-630.
7. Sambrook, J., Short, J. M., Wynshaw-Boris, A., and Pelham, H. R. B. (1983) Cell 36, 426-443.
8. Bienz, M., and Pelham, H. R. B. (1986) Cell 46, 735-733.
9. Reynolds, G. A., Basu, S. K., Osborne, T. H., Chien, D. J., Gill, G., Brown, M. S., Goldstein, J. L., and Lusky, K. L. (1984) Cell 37, 275-285.
10. Melton, D. W., McEwan, C., McEl, A. B., and Reid, A. M. (1986) Cell 44, 319-328.
11. Watt, R., T., Yagi, J. S., Fagan, J. B., and Pelham, H. R. B. (1986) J. Biol. Chem. 261, 9721-9726.
12. Vogeli, G., Oshkuro, H., Sobel, M., Yamada, Y., Pastan, I., and de Crombrugghe, B. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 3533-3538.
13. Kohno, K., Sullivan, M., and Yamada, Y. (1985) J. Biol. Chem. 260, 4411-4417.
14. Schmidt, A., Yamada, Y., and de Crombrugghe, B. (1984) J. Biol. Chem. 239, 741-745.
15. Dickson, L. A., deWet, W., DiLiberto, M., Weil, D., and Ramirez, F. E. (1987) Nucl. Acids Res. 15, 3427-3438.
16. Fromm, M., and Berg, P. (1987) J. Mol. Genet. Genet. 127, 135.
17. Yamamoto, T., Yagi, J. S., Fagan, J. B., and Pelham, H. R. B. (1986) J. Biol. Chem. 261, 9721-9726.
18. Vogeli, G., Oshkuro, H., Sobel, M., Yamada, Y., Pastan, I., and de Crombrugghe, B. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 3533-3538.
19. Kohno, K., Sullivan, M., and Yamada, Y. (1985) J. Biol. Chem. 260, 4411-4417.
20. Schmidt, A., Yamada, Y., and de Crombrugghe, B. (1984) J. Biol. Chem. 239, 741-745.
21. Dickson, L. A., deWet, W., DiLiberto, M., Weil, D., and Ramirez, F. E. (1987) Nucl. Acids Res. 15, 3427-3438.
22. Fromm, M., and Berg, P. (1987) J. Mol. Genet. Genet. 127, 135.
23. Ishii, S., Xu, Y., Stratton, R. H., Ro, A. E., Merlino, G. T., and Pastan, I. (1993) Proc. Natl. Acad. Sci. U. S. A. 88, 4209-4214.
24. Motomura, M., Sevarino, K. A., Wagner, J. A., Mangel, D., and Goodman, R. H. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 6885-6886.
25. Dean, D. C., Boyleau, C. L., and Bourgeois, S. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 1876-1880.
26. Bush, G. L., and Bensc, C. S. (1987) Connect. Tissue Res. 19, 95-107.
27. Rovira, R., and de Crombrugghe, B. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 5590-5594.
28. Rumyantsev, C. M., Vergeer, W. P., de Ploey, S. J., Bernard, M. P., Ramirez, F., and de Wet, W. J. (1987) J. Biol. Chem. 262, 15151-15157.
29. Bornstein, P., McKay, J., Morishima, J. K., Derarayahn, S., and Gelinas, R. C. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 8684-8685.
30. Godfraind, T., Ingram, P., and de Crombrugghe, B. (1986) Mol. Cell. Biol. 6, 477-487.
31. Krumlauf, R., Hanner, R. E., Tlieghman, S. M., and Brinster, R. L. (1985) Mol. Cell. Biol. 5, 1659-1666.