Proc. Natl. Acad. Sci. USA
Vol. 78, No. 9, pp. 5334–5338, September 1981
Biochemistry

Structure of the promoter for chicken $\alpha 2$ type I collagen gene

(DNA sequencing/S1 nuclease/primer extension/dyads of symmetry/signal peptide)

GABRIEL VOGELI*†, HIROAKI OHKUBO*, MARK E. SOBEL‡, YOSHIHIKO YAMADA*, IRA PASTAN*, AND BENOIT DE CROMBRUGGHE*

*Laboratory of Molecular Biology, National Cancer Institute and ‡Laboratory of Developmental Biology and Anomalies, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland 20205

Communicated by C. B. Anfinsen, May 1, 1981

ABSTRACT **The chicken $\alpha 2$ type I collagen gene is 38 kilobases long and its coding information is subdivided into more than 50 exons. In the current study, we used primer extension and S1 nuclease mapping to determine the sequence of the 5′ end of $\alpha 2$ collagen mRNA and to locate the start site for transcription of the $\alpha 2$ collagen gene. The DNA sequence around the start site for transcription shows a typical Goldberg–Hogness sequence, 5′ T-A-T-A-A-A-T 3′, between −33 and −26 and a 5′ G-C-C-C-A-T-T 3′ sequence ("CAT" box) between −84 and −78. Three AUGs are found in the initial portion of the mRNA, the first from +54 to +56, the second from +117 to +119, and the third from +134 to +136. The first two AUGs are followed by short coding sequences that could specify a hexapeptide and a tetrapeptide, respectively. Only the third AUG is followed by an open reading frame coding for a sequence that presents considerable homology with the previously determined amino acid sequence of prepro $\alpha 1$ collagen. In the promoter region sequence there are several extensive dyads of symmetry. Three of these inverted repeats which precede the start site for transcription overlap each other and may have a role in the developmental regulation of this gene.**

The collagens are a family of extracellular proteins with common structural properties (for a review, see ref. 1). It is likely that the synthesis of each of the collagen types found in different tissues of higher vertebrates responds to a tissue-specific, developmentally regulated, differentiation program. In cultured chicken embryo fibroblasts the synthesis of type I collagen synthesis is selectively inhibited by the presence in these cells of $p60^{src}$, the transforming protein encoded by the Rous sarcoma virus genome (2). The decrease in collagen synthesis in these transformed cells is due to a decrease of more than 90% in the levels of both translatable cytoplasmic mRNA (3) and of nuclear intron-specific RNA (4) for $\alpha 2$ type I collagen. Hence, it is likely that the control of collagen synthesis by $p60^{src}$ occurs at the level of transcription. We consider the synthesis of type I collagen by chicken embryo fibroblasts as a model system in which to study a specific differentiation program and its perturbation by oncogenic growth factors.

To study the control of the collagen genes that are expressed in chicken embryo fibroblasts in appropriate *in vivo* and *in vitro* reconstituted systems, we have isolated the gene that codes for chicken $\alpha 2$ collagen (5, 6). The gene is 38 kilobases (kb) long; its coding information is subdivided into more than 50 exons. A remarkable feature of this gene is that many of the exons that code for the helical portion of $\alpha 2$ collagen have an identical length, 54 base pairs (bp), although the sequences within these exons vary. This finding has important implications regarding the assembly of the ancestral collagen gene and led us to propose that the primordial collagen gene arose by multiple duplications of a single genetic unit containing an exon of 54 bp (7, 8).

Our major interest in the $\alpha 2$ collagen gene is to study its control. We assume that this control occurs at the 5′ end of the gene. We therefore characterized the 5′ end of $\alpha 2$ collagen mRNA and the structure of the promoter for this gene.

MATERIALS AND METHODS

Plasmid Isolation and DNA Sequencing. A 3.5-kb *Eco*RI/*Bgl* II fragment from the genomic clone λ COL-323 (6, 8) was subcloned into the *Eco*RI and *Bam*HI sites of pBR322. Plasmid DNA was purified from *Escherichia coli* C600 by standard isolation methods (5). Restriction digestions were performed according to the recommendations of the manufacturers (New England BioLabs, BRL). End-labeling of DNA with [γ-^{32}P]ATP and DNA sequence determinations were done by using the methods of Maxam and Gilbert (9).

Primer Extension by Avian Myeloblastosis Virus Reverse Transcriptase. A *Hin*fI/*Pst* I DNA fragment, labeled at its *Hin*fI 5′ end, was used as primer (see Fig. 1). This fragment was hybridized with total oligo(dT)-purified RNA (10). The primer was extended with reverse transcriptase (a gift from J. W. Beard) and each of the four deoxynucleotides at 1 mM (11). After 30 min at 42°C, the reaction was terminated with 20 mM EDTA, NaOH was added to 0.2 M, and the labeled reaction products were electrophoresed on a 7% polyacrylamide/7 M urea gel to determine the size of the extended product.

Digestion with S1 Nuclease. S1 endonuclease mapping was conducted according to Berk and Sharp (10). Oligo(dT)-purified RNA (5 μg) from chicken embryo calvaria and long bones (12) was hybridized with 0.025 μg of a *Pst* I/*Ava* II DNA fragment (see Fig. 3) labeled at its *Pst* I 5′ end. At the end of the hybridization period (3 hr; 50°C), the reaction mixture was diluted 1:10 in ice-cold S1 nuclease buffer containing 1000 units of S1 nuclease per ml (10). After 30 min at 40°C the reaction products were extracted with phenol and separated on a 7 M urea/10% polyacrylamide gel (9).

RESULTS

Determination of the 5′ End of $\alpha 2$ Collagen RNA. We previously reported the isolation of the entire chicken $\alpha 2$ collagen gene in a series of overlapping clones (5, 6, 8). The gene was isolated starting from its 3′ end by successive screenings of a library of random genomic fragments. When hybrids between $\alpha 2$ collagen RNA and the $\alpha 2$ collagen genomic clones were examined by electron microscopy, unhybridized RNA tails could be seen adjacent to those exons located at the 5′ and 3′ end of

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

Abbreviations: kb, kilobase(s); bp, base pair(s).

† Present address: National Eye Institute, Bldg. 6, Rm. 203, National Institutes of Health, Bethesda, MD 20205.

the clones. In hybrids between α2 collagen RNA and the most 5′ genomic clone that we isolated, λ COL-323, no RNA tail could be seen attached to the most 5′ exon in this clone. This exon, which was previously designated as exon 51 (6, 8), is called exon 1 in this paper. The clone λ COL-323 contains an additional 7-kb segment beyond exon 1. These sequences could contain the promoter region of the α2 collagen gene.

To examine this promoter region, a 3.5-kb *Eco*RI/*Bgl* II DNA segment located around exon 1 was subcloned in plasmid pBR322 (Fig. 1). The only exon present in this subclone is exon 1. To localize the exon more specifically within this 3.5-kb subcloned segment, we digested it with *Sma* I, fractionated the digestion products by agarose gel electrophoresis, transferred the DNA fragments to nitrocellulose paper (13), and hybridized them with a preparation of ^{32}P end-labeled (14) α2 collagen RNA. An 800-bp *Sma* I fragment (Fig. 1) was found to hybridize with the RNA (data not shown). We concluded that this fragment contains sequences present in exon 1.

To prove that exon 1 of clone λ COL-323 is the promoter proximal exon of the gene and also to localize precisely the sequences present at the extreme 5′ end of α2 collagen mRNA on the genomic DNA clone, we performed a primer extension experiment and an S1 nuclease mapping experiment.

Primer Extension Experiment. Fig. 2*A* illustrates the principle of the primer extension using reverse transcriptase (11). As template, we used a RNA preparation enriched for α2 collagen mRNA; the primer was a small DNA fragment containing sequences located in exon 1. After hybridization of the primer which was labeled at its 5′ end to the mRNA, cDNA was synthesized by using reverse transcriptase. The size of the extension product could be precisely measured and its sequence could be determined. Because the primer hybridizes to a sequence close to the 5′ end of the mRNA, the cDNA should extend to the end of the α2 collagen mRNA. If the nucleotide sequence of this cDNA is identical and colinear with the sequence present in the cloned genomic DNA, we can conclude that no introns interrupt the mRNA sequence defined by the cDNA extension product. We can also conclude that exon 1 contains the sequence present at the extreme 5′ end of the RNA.

By using a 47-nucleotide *Hin*fI/*Pst* I fragment as primer, a single discrete extension product was synthesized. No other longer cDNAs could be seen on the autoradiograph even after prolonged exposure of the gel (data not shown). This cDNA was eluted from a polyacrylamide gel and, because it contained a uniquely labeled 5′ end, its sequence was determined by the Maxam and Gilbert (9) method. The sequence could be read to about two nucleotides from the 3′ end of this cDNA (Fig. 2*B*). The sequence is identical and colinear to the DNA sequence in the genomic clone immediately adjacent to the *Hin*fI/*Pst* I fragment used as primer (Fig. 3). The genomic DNA corresponding to the extended primer therefore is not interrupted by introns. The finding of a unique cDNA product strongly suggests that this cDNA extends to the 5′ end of α2 collagen mRNA and that exon 1 is the promoter-proximal exon of the α2 collagen gene.

S1 Nuclease Mapping Experiment. To define more precisely the 5′ end of the mRNA and hence the start site of transcription for α2 collagen RNA, we performed an S1 nuclease mapping experiment (Fig. 2*A*) (10). The 114-nucleotide *Pst* I/*Ava* II fragment labeled at its *Pst* I 5′ end was hybridized with a preparation of α2 collagen RNA. The single-stranded tails of the RNA·DNA hybrid were digested with S1 nuclease. The exact size of the protected labeled DNA probe was determined by coelectrophoresis with the products of the four base-specific sequence-determination reactions of the 5′-end-labeled *Pst* I/*Ava* II probe (Fig. 2*C*). Because the base-specific fractionation products are in fact one nucleotide shorter than their sequence indicates, the two protected fragments in lane 6 of Fig. 2*C* correspond to a thymidine and an adenine residue in the sequencing lanes. The microheterogeneity seen in Fig. 2*C* is probably inherent to the S1 nuclease mapping assay, although it is possible that transcription starts at two adjacent residues. The sequence of the DNA probe is complementary to the sequence of α2 collagen RNA. Therefore, the 5′ end of the α2 collagen RNA begins with the complementary adenine or uridine residues. This locates the start site for transcription of α2 collagen RNA on the genomic DNA to position +1 or −1 (Fig. 3) or both.

Sequences Around the Initiation Site for Transcription. Fig. 1 shows the restriction fragments and the strands used to determine the sequence shown in Fig. 3. The thymidine and adenine residues that correspond to the probable 5′ end of the mRNA are indicated as −1 and +1. The DNA sequence shown extends from 404 bp upstream from the mRNA start site to 177 bp downstream. There are three distinct AUG codons in the initial part of α2 collagen mRNA (see also Fig. 6). The first is found from +54 to +56, the second from +117 to +119, and the third from +134 to +136. The first AUG is followed by five

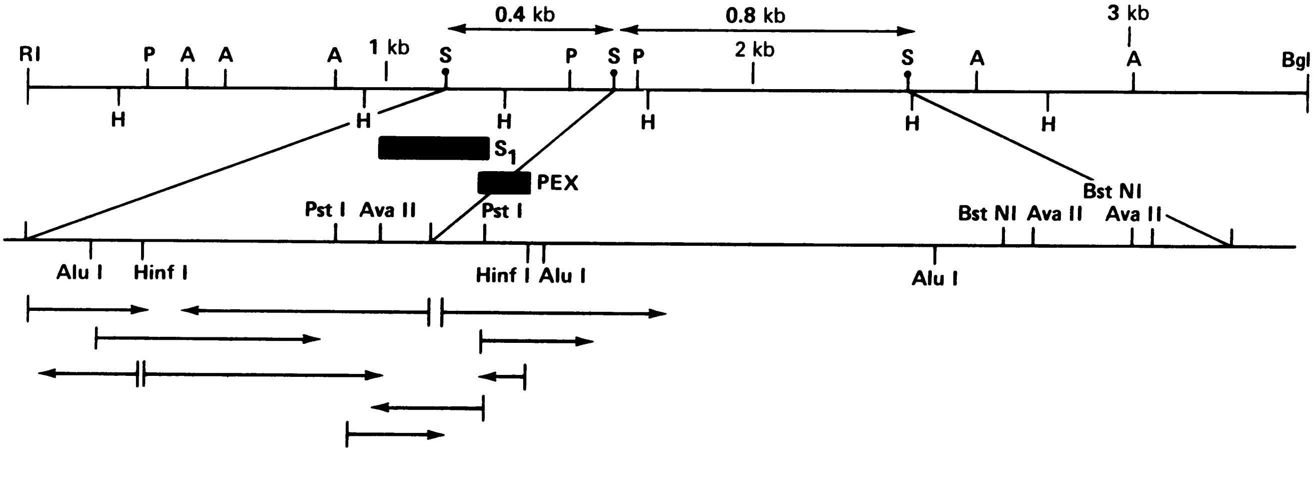

FIG. 1. Restriction map of a 3.5-kb genomic DNA fragment surrounding the most 5′ exon of the chicken α2 collagen gene. The arrows indicate the restriction fragments for which the DNA sequence was determined and the 5′ → 3′ direction of sequence analysis. The two boxes designated S1 and PEX refer to the restriction fragments used in the S1 nuclease mapping assay and the primer extension experiments, respectively. Restriction enzyme cleavage sites are indicated as follows: RI, *Eco*RI; H, *Hin*fI; P, *Pst* I; A, *Sau*3A; S, *Sma* I.

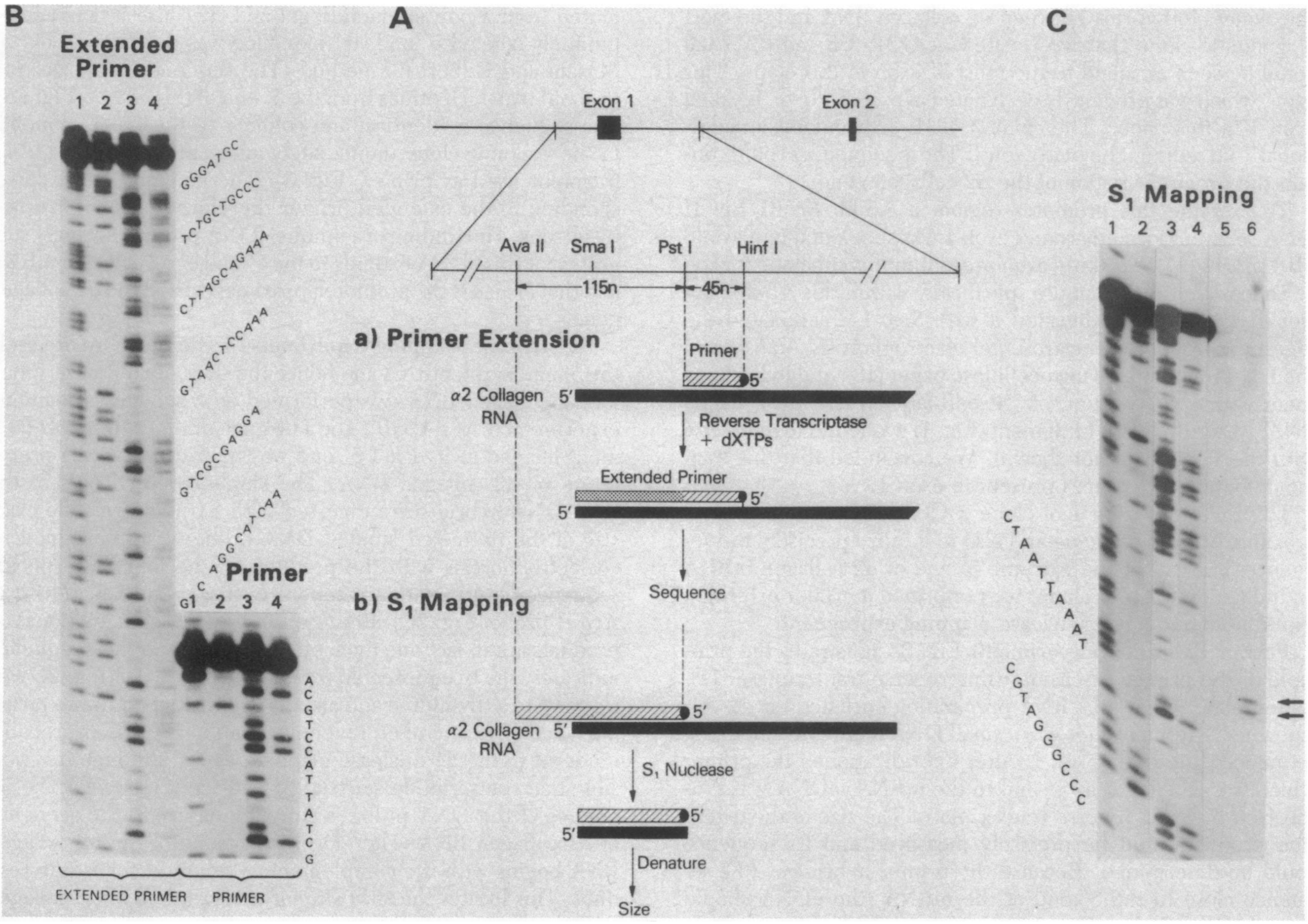

FIG. 2. Primer extension experiment and S1 nuclease mapping experiment. (*A*) The principle of the experiment. *Primer Extension.* The *Hin*fI/*Pst* I DNA fragment, labeled at its *Hin*fI 5′ end and used as primer, is hybridized to a preparation of α2 collagen mRNA and extended with reverse transcriptase. *S1 nuclease mapping.* The *Pst*I/*Ava* II fragment, labeled at its *Pst* I 5′ end, is hybridized with a preparation of α2 collagen mRNA. The hybrids are treated with S1 nuclease, denatured, and analyzed for size. (*B*) Base-specific fractionation products of the extended primer (left four lanes) and of the primer alone (right four lanes) electrophoresed on a 10% polyacrylamide/7 M urea gel. Lanes: 1, A + G reaction; 2, G reaction; 3, C + T reaction; 4, C reaction. (*C*) Size fractionation of the products of the S1 nuclease mapping reactions (lane 6) and comparison of their size with that of the base-specific fractionation product of the *Pst* I/*Ava* II fragment labeled at its *Pst* I 5′ end (reactions as lanes 1–4 in *B*). Lane 5, control reaction (S1 nuclease digestion without collagen RNA); lane 6, products of S1 nuclease digestion after hybridization of the probe with collagen RNA.

triplets coding for Pro-Ala-Ser-Lys-Arg and then by a termination codon. The second AUG is followed by codons for Ser-Ser-Lys and then by a termination codon. The third AUG, which appears in a different reading frame than the two preceding AUGs, is followed by an open reading frame. The amino acid sequence deduced from the nucleotide sequence starting from this third AUG indicates that most residues are hydrophobic (Fig. 4) and show considerable homology with the known amino acid sequence of the NH_2-terminal portion of prepro α1 collagen (15). This sequence is consistent with the incomplete amino acid sequence of the extreme NH_2-terminal region of prepro α2 collagen (16).

The size of exon 1 is 203 nucleotides as determined by an S1 nuclease mapping experiment (data not shown). It specifies

```
     -400                                              -350
CCCCCGGACAGCTCCCGCTCCGACAGCCGTCGCGCTTACCGGCGCGCCCGCCGCCGGCGGGCAACAAAGCAGGGCGAGGGGCGGGGAACGTCTGAAA
         -300                                             -250
AAAAAAAAAAAAATCAGACGGCGAGTCAGATTTTCCTCCTGAAAGCCTCAAAGTGTCCACGTCCTCGAAGCATGGAACCAATTTAGCGCCGCCGCCG
          -200                                              -150
CCTTCCTCTTCCCTCCCTTCCTCCCTCCCTCGCCCCCCCCTCCGACCCCGCAGCCGAGCAGCGCCGGGCTGGGGCCGGTGGGCACGTGACAGCGCTGG
           -100                                          --50
GAGCCGCGCGGGCCCCGCGGCGCCGCGCGCCCATTGCTGCAGCGCCGCCGGTGCCCGCAGCCGCGGGACCCCCTGCGGTATAAATACGGCGGAGCGGG
             -1 +1
              \ |                                           +50
GCTTGATTAATTTAGCATCCCGGGCAGCAGGTTTCTGCTAAGTTTGGAGTTACTCCTCGCGACTGTATGCCTGCGTCCTGCAGGTAATAGCCAACCACG
             ------->
             +100                                           +150
TCCGGGGGCTCTGCAACACAAGGAGTCTGCATGTCTAGCAAGTAGACATGCTCAGCTTTGTGGATACGCGGATTTTGTTGCTGCTCGCAGT ...
```

FIG. 3. DNA sequence of genomic DNA around the transcription start site, indicated as +1/−1. The Goldberg–Hogness (−33 to −26) box sequence and the CAT box sequence (−84 to −78) are underlined. The arrow indicates the direction of transcription. The sequence corresponds to the RNA sense strand.

α2 Met Leu Ser Phe Val Asp Thr Arg Ile Leu Leu Leu Leu Ala Val

α1 Met Phe Ser Phe Val Xxx Ser Arg Leu Leu Leu Leu Ile Ala Ala

FIG. 4. Comparison of the amino acid sequences at the NH_2 termini of prepro α2 collagen and prepro α1 collagen (16).

both the untranslated segment at the 5′ end of α2 collagen mRNA and a sequence coding for the first 23 amino acids of prepro α2 collagen.

Preceding the start site for transcription at −33 to −27 is the sequence 5′ T-A-T-A-A-A-T 3′. This sequence is similar to sequences found in many other RNA polymerase II-dependent promoters at the same location (17). Between −84 and −78 the sequence 5′ G-G-C-C-A-T-T 3′ is found. A similar sequence, also called the CAT box, is found in many other RNA polymerase II-dependent promoters between −75 and −85.

We used a published computer program (18) to search for dyads of symmetry in the sequence shown in Fig. 3. In addition to a number of smaller dyads of symmetry there are five large inverted repeats which are represented in Figs. 5 and 6. Dyad 1 covers a segment between −155 and −82, dyad 2 between −125 and −68, and dyad 3 between −105 and −37. If the DNAs within these sequences formed stem and loop structures, each of these three inverted repeats would exclude the two others. A fourth inverted repeat is found between −16 and +43 surrounding the initiation site. The transcription start site (+1) would be located within the stem in a potential stem and loop structure. The fifth large dyad is found between +105 and +166. This last symmetrical sequence would thus also be reflected in the secondary structure of α2 collagen mRNA (Fig. 6).

DISCUSSION

We have determined the sequence of the 5′ end of chicken α2 collagen mRNA by using both a primer extension method and by an S1 nuclease mapping assay. The nucleotide sequence of the cDNA produced by extension of a primer which contained exon sequences located close to the 5′ end of the gene is identical with the sequence in the genomic DNA. The S1 nuclease mapping experiment which locates this 5′ end more precisely suggests that the start site for transcription corresponds to the adenine residue designated +1 or to the adjacent thymidine residue designated −1 in Fig. 3. The slight ambiguity concerning the exact start site is probably inherent to the S1 nuclease mapping assay. We favor the adenine at +1 as start site because most known eukaryotic mRNAs begin with an adenine residue (17).

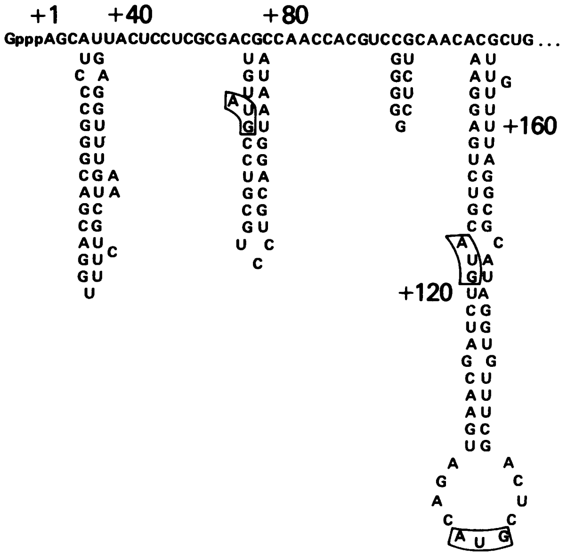

FIG. 6. Possible secondary structures of the 5′ portion of α2 collagen mRNA. The three possible start codons are enclosed.

Although our results indicate that adenine at +1, or thymidine at −1, is the likely start site for transcription, we have not yet proven by direct RNA sequence analysis that one or both of these two adjacent residues are part of a cap structure. We also cannot exclude that in the primer extension experiment the RNA template-dependent synthesis of cDNA terminates before it reaches the 5′ end of the RNA as a result of a secondary struc-

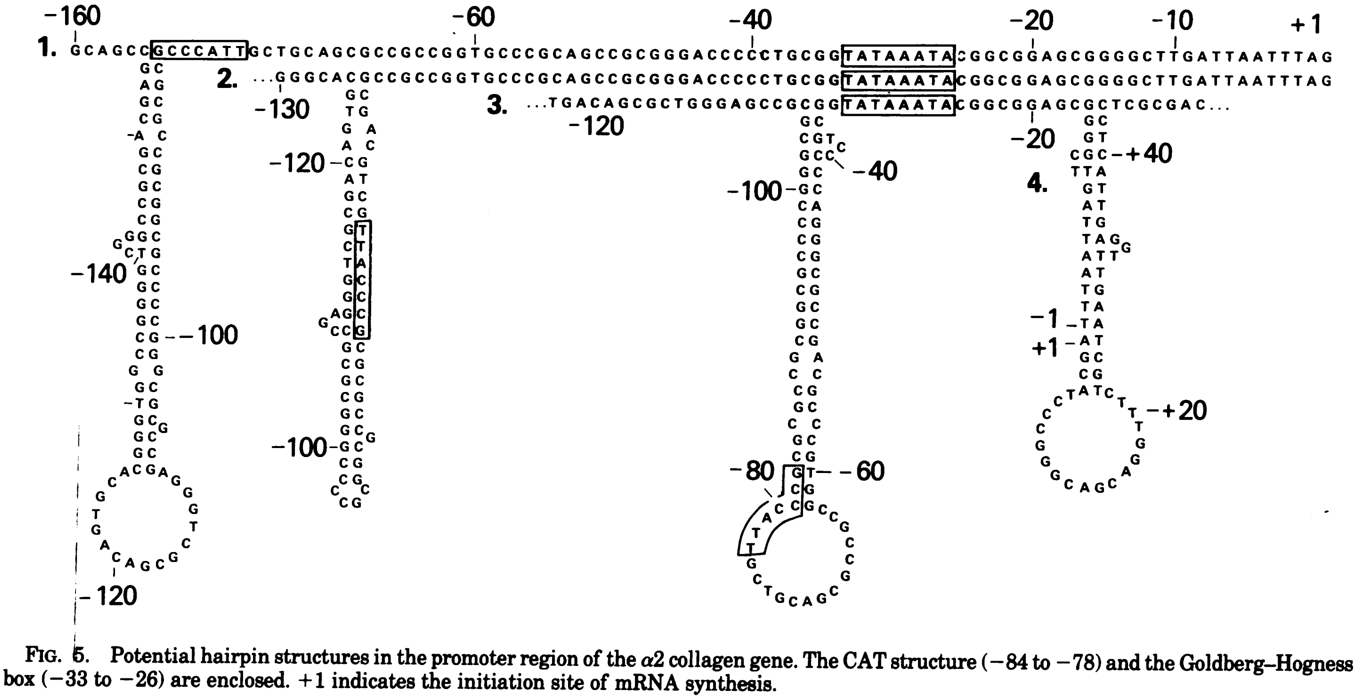

FIG. 5. Potential hairpin structures in the promoter region of the α2 collagen gene. The CAT structure (−84 to −78) and the Goldberg–Hogness box (−33 to −26) are enclosed. +1 indicates the initiation site of mRNA synthesis.

ture in the RNA. We consider this possibility unlikely because only a single discrete extension product was obtained by primer extension and also because the results of the S1 nuclease mapping experiments agree very well with the results of the enzymatic primer extension experiment in assigning the start site of transcription. Consistent with a correct assignment of the initiation site is the finding that the DNA sequence preceding the start site exhibits both a Goldberg–Hogness sequence (−33 to −26) and a 5′ G-C-C-C-A-T-T 3′ (−83 to −77) located at similar distances from the initiation site as in other eukaryotic promoters (17).

In vitro transcription experiments with extracts from HeLa cells (19) using as template cloned DNA fragments containing sequences around the +1 locus indicate that in this system transcription of the $\alpha 2$ collagen gene initiates at the same site as *in vivo* (20). This result supports our assignment of the transcription initiation site.

In this study we assume that the 5′ end of mature $\alpha 2$ collagen mRNA coincides with the 5′ end of the RNA precursor for $\alpha 2$ collagen. Such an assumption appears to be justified because for both mouse β-globin and chicken ovalbumin mRNA identity between the 5′ ends of the mature mRNA and the nuclear precursor RNA has been established (21, 22).

The first two AUG codons in the mRNA are followed by sequences coding for short polypeptides and then by a termination triplet. Only the third AUG is followed by a longer coding sequence which specifies an amino acid sequence resembling the previously determined NH_2-terminal sequence of prepro $\alpha 1$ collagen (16).

The existence of these three AUGs would constitute one of the exceptions to the scanning model for initiation of protein synthesis (23). It is interesting to note that the first two AUGs would be part of a potential base-paired hairpin whereas the third AUG would be present in the loop of such structure (Fig. 6).

If only the third AUG is recognized as a translation start, the untranslated region at the 5′ end of $\alpha 2$ collagen mRNA would be 132 nucleotides long. Such a large untranslated sequence could suggest that it may play a role in translational control. Evidence has been reported which favors the idea that the NH_2-terminal peptides of type I collagen inhibit translation of the mRNAs for type I collagen (24). It is also conceivable that the hexapeptide encoded by the sequence between +54 and +72 or the tetrapeptide encoded by the sequence between +117 and +128 plays a role in the regulation of the expression of the $\alpha 2$ collagen gene.

The amino acid sequences of prepro $\alpha 2$ collagen deduced from the DNA sequence shows considerable similarities with the published amino acid sequence of the signal peptide of prepro $\alpha 1$ collagen (16). These similarities are more pronounced than when signal peptides of different proteins are compared to each other (25). This could reflect an evolutionary relationship between the signal peptides of the two different chains of type I collagen. In addition the similarities of these two signal peptides could cause the two chains to be secreted into the endoplasmic reticulum via a common entry site in order to facilitate the assembly of collagen molecules.

The promoter sequence shows four very large dyads of symmetry. Three of these inverted repeat sequences are overlapping and would be mutually exclusive if they were to form a stem and loop structure (Fig. 5). Each of these three potential cruciform structures has about the same stability (−100 kcal). They could be interaction sites for regulatory proteins and hence play an important role in the control of the $\alpha 2$ collagen gene.

We thank Cary L. Queen for his help with the computer program. We also thank Mary Lee Lanigan and Linda S. Harris for their typing assistance.

1. Bornstein, P. & Sage, H. (1980) *Annu. Rev. Biochem.* **49,** 957–1003.
2. Sobel, M. E., Yamamoto, T., de Crombrugghe, B. & Pastan, I. (1981) *Biochemistry* **20,** 2678–2684.
3. Adams, S. L., Sobel, M. E., Howard, B. H., Olden, K., Yamada, K. M., de Crombrugghe, B. & Pastan, I. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 3399–3403.
4. Avvedimento, E., Yamada, Y., Lovelace, E., Vogeli, G., de Crombrugghe, B. & Pastan, I. (1981) *Nucleic Acids Res.* **9,** 1123–1131.
5. Vogeli, G., Avvedimento, E. V., Sullivan, M., Maizel, Jr., J. V., Lozano, G., Adams, S. L., Pastan, I. & de Crombrugghe, B. (1980) *Nucleic Acids Res.* **8,** 1823–1837.
6. Ohkubo, H., Vogeli, G., Mudryj, M., Avvedimento, E. V., Sullivan, M., Pastan, I. & de Crombrugghe, B. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 7059–7063.
7. Yamada, Y., Avvedimento, V. E., Mudryj, M., Ohkubo, H., Vogeli, G., Irani, M., Pastan, I. & de Crombrugghe, B. (1980) *Cell* **22,** 887–892.
8. Vogeli, G., Ohkubo, H., Avvedimento, V. E., Sullivan, M., Yamada, Y., Mudryj, M., Pastan, I. & de Crombrugghe, B. (1981) *Cold Spring Harbor Symp. Quant. Biol.* **45,** 777–783.
9. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
10. Berk, A. J. & Sharp, P. A. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 1274–1278.
11. Wickens, M. P., Buell, G. N. & Schimke, R. T. (1978) *J. Biol. Chem.* **253,** 2483–2495.
12. Sobel, M. E., Yamamoto, T., Adams, S. L., DiLauro, R., Avvedimento, V. E., de Crombrugghe, B. & Pastan, I. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 5846–5850.
13. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
14. Craig, E. A., McCarthy, B. J. & Wadsworth, S. C. (1979) *Cell* **16,** 575–588.
15. Palmiter, R. D., Davidson, J. M., Gagnon, J., Rowe, D. W. & Bornstein, P. (1979) *J. Biol. Chem.* **254,** 1433–1436.
16. Graves, P. N., Olsen, B. R., Fietzek, P. P., Prockop, D. J. & Monson, J. M. (1981) *Eur. J. Biochem.*, in press.
17. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980) *Nucleic Acids Res.* **8,** 127–142.
18. Queen, C. L. & Korn, L. J. (1980) *Methods Enzymol.* **65,** 595–609.
19. Manley, J. L., Fire, A., Cano, A., Sharp, P. A. & Gefter, M. L. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3855–3859.
20. Merlino, G. T., Vogeli, G., Yamamoto, T., de Crombrugghe, B. & Pastan, I. (1981) *J. Biol. Chem.*, in press.
21. Weaver, R. F. & Weissmann, C. (1979) *Nucleic Acids Res.* **7,** 1175–1193.
22. Roop, D. R., Tsai, M. J. & O'Malley, B. W. (1980) *Cell* **19,** 63–68.
23. Kozak, M. (1978) *Cell* **15,** 1109–1123.
24. Paglia, L., Wilczek, J., de Leon, L. D., Martin, G. R., Horlein, D. & Muller, P. (1979) *Biochemistry* **18,** 5030–5034.
25. Davis, B. D. & Tai, P. C. (1980) *Nature (London)* **283,** 433–438.