Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 1659–1663, March 1984
Biochemistry

Endonuclease S1-sensitive site in chicken pro-α2(I) collagen 5′ flanking gene region

(type I α2 collagen gene/restriction digestion/topoisomers/pyrimidines)

MITCHELL H. FINER, ERIC J. B. FODOR, HELGA BOEDTKER, AND PAUL DOTY

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, MA 02138

Contributed by Paul Doty, November 28, 1983

ABSTRACT **A site that is preferentially cleaved by the single-strand-specific endonuclease from *Aspergillus oryzae* was located *in vitro* 180 base pairs upstream from the 5′ end of the chicken pro-α2(I) collagen gene. It is found in supercoiled plasmids with a negative superhelical density of −0.024 or more but not in linear DNA molecules. The nuclease S1 sensitivity is retained in plasmids containing genomic fragments extending from position +8 to −285 (where +1 is the first transcribed base) and from −147 to −351 and also in a 5.7-kilobase *Eco*RI fragment that extends 1.6 kilobases 5′ and 4.1 kilobases 3′ to the 5′ end of the gene. Analysis at the nucleotide level on a DNA sequence gel places the site at −181 to −182 on the sense strand and at −182 to −184 and −192 to −195 on the nonsense strand. These sites lie within a stretch of 42 pyrimidines interrupted by a single guanine and within the sequence T-C-C-C-T-C-C-C-T-T-C-C-T-C-C-C-C-T-C-C-C-T.**

Altered chromatin conformation has been postulated and in numerous cases observed in genes coding for proteins that are being expressed at high levels and whose expression is tightly controlled. The altered conformation is associated with the enhanced sensitivity of the gene to DNase I (1, 2) over a considerable region and with the hypersensitivity to DNase I of smaller sites located often but not always near the 5′ end of the gene (3–6). The simplest interpretation of this altered conformation is that some of the DNA in these regions is single stranded (7). Alternatively, it may exist in an altered DNA conformation, such as a cruciform structure (8–10) or in a left-handed Z helix (11), in which the loop of the cruciform or the B-to-Z junction would provide the single-stranded region responsible for the DNase I hypersensitivity. More recently, the single-strand-specific endonuclease S1 from *Aspergillus oryzae* has been used to identify an S1-sensitive site 50 to 150 base pairs (bp) 5′ to the transcription start site of the chicken β-globin gene (12). The single-stranded nature of the DNA at this site was confirmed by its reaction with bromoacetaldehyde (13).

Nuclease S1-hypersensitive sites also have been identified *in vitro* in supercoiled plasmids in regions 5′ to the *Drosophila melanogaster* heat shock genes (14) and within the adenovirus 12 early and adenovirus 2 major late promoter regions (15). Although the biological importance of these sites has not been demonstrated, their nonrandom location suggests that they serve some, perhaps tissue-specific, function.

As a first step in studying how the expression of the chicken pro-α2(I) collagen gene is regulated, we probed the 5′ flanking gene region (promoter region) of this gene for unusual secondary structure with nuclease S1. It has been reported (16, 17) that the promoter region contains several inverted repeats, which have the potential of forming cruciform structures, and short repeats of CpGp, which are potential Z-DNA sequences. Although we indeed identified a dominant nuclease S1 site, it is not located in either a potential cruciform or Z-DNA sequence but is within a stretch of 42 pyrimidines interrupted by a single guanine.

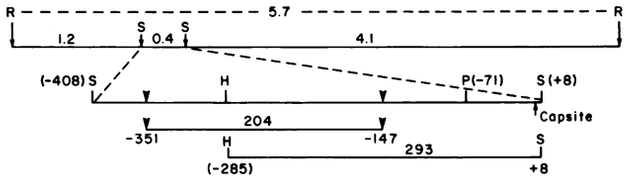

FIG. 1. Restriction map of 5.7-kb *Eco*RI fragment showing the location of the 0.4-kb *Sma* I fragment containing the pro-α2(I) collagen promoter region and the location of the 293-bp *Hinf*I–*Sma* I fragment and the 204-bp *Hpa* II fragment within the *Sma* I fragment. R, *Eco*RI sites; S, *Sma* I sites; H, *Hinf*I site; ▾, *Hpa* II sites; P, *Pst* I site.

METHODS

Construction of Recombinant Plasmids Containing the 5′ Flanking Gene Region of the Pro-α2(I) Collagen Gene. *pXf3/CgPR*. A 416-bp *Sma* I fragment of the pro-α2(I) collagen gene containing the region from position +8 to −408 as shown in Fig. 1, where +1 is the transcription start site, was subcloned by D. Hanahan in pXf3 [a derivative of pBR322 constructed by D. Hanahan (18)] by using synthetic *Eco*RI and *Hin*dIII linkers.

pCg293. The 416-bp *Sma* I fragment was isolated and digested with *Hinf*I, and the resultant 293-bp fragment shown in Fig. 1 was subcloned into the *Hin*dIII site of pBR322 by using synthetic *Hin*dIII linkers.

pCg204. The 416-bp *Sma* I fragment was digested with *Hpa* II. The 204-bp fragment shown in Fig. 1, extending from −147 to −351, was subcloned into pBR322 by using synthetic *Hin*dIII linkers. Note that the insert lost 147 bp 5′ to the cap site, including the "CAT" box and the "TATA" box, which are implicated as essential for efficient and accurate initiation of other eukaryotic genes.

pCg5.7. The 5.7-kb *Eco*RI restriction fragment containing the first four pro-α2(I) collagen exons and 1.6 kb of 5′ flanking gene sequences (19) was subcloned into the *Eco*RI site of pBR322. The location of the 416-bp *Sma* I fragment within the 5.7-bp *Eco*RI fragment is shown in Fig. 1.

All transformations were carried out as described by D. Hanahan (18). Plasmid DNA was purified on CsCl gradients.

Digestion with Nuclease S1 and Restriction Enzymes. Plasmid DNA, either supercoiled or linear, was digested with nuclease S1 (Bethesda Research Laboratories) at 0.4 units/μg of DNA in 30 mM Na acetate, pH 4.5/80 mM NaCl/1 mM $ZnSO_4$. Digestions were carried out at 7°C for 16 hr, a modification of the procedure described by Lilley (8). After digestion, the DNA was purified by phenol extraction and ethanol precipitation. It was then either digested directly with the appropriate restriction enzyme (New England BioLabs) or

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 pair(s); bp, base pair(s).

first 5′-end-labeled by treatment with calf intestine alkaline phosphatase (Boehringer Mannheim), followed by treatment with T4 polynucleotide kinase and [γ-^{32}P]ATP (New England Nuclear). Labeled products of the digestions were analyzed on 6% polyacrylamide gels. All enzymatic reactions were carried out according to manufacturers' specifications.

Preparation of Topoisomers of pCg293. Plasmid DNA was incubated in the presence of various concentrations of ethidium bromide with turkey erythrocyte topoisomerase I (prepared by G. Pflugfelder) for 3.5 hr at room temperature. Then the samples were extracted with phenol/chloroform/isoamyl alcohol, 50:49:1 (vol/vol) three times and then dialyzed first against 2 M NaCl/10 mM Tris·HCl, pH 8, and then against 10 mM Tris·HCl, 0.2 mM Na_2EDTA, pH 8, to remove the ethidium bromide as described by Peck *et al.* (20).

DNA Sequence Analysis. In order to determine the location of the nuclease S1-hypersensitive site at the nucleotide level on each strand, supercoiled pXf3/CgPR DNA was treated with nuclease S1 as described above. Plasmid DNA linearized by nuclease S1 was 5′-end-labeled by treatment with calf intestine alkaline phosphatase, followed by treatment with T4 polynucleotide kinase in the presence of [γ-^{32}P]ATP. Labeled DNA was digested with *Eco*RI or *Hin*dIII restriction endonuclease. pCgPR DNA for sequence determination was digested either with *Eco*RI or *Hin*dIII restriction endonuclease and 3′-end-filled by using [α-^{32}P]dNTP's and the *Escherichia coli* DNA polymerase I large fragment. DNA labeled at the *Eco*RI site was digested with *Hin*dIII (and vice versa), generating asymmetrically labeled promoter DNA fragments, which were purified from 6% acrylamide gels and their sequences determined as described by Maxam and Gilbert (21). The strand for sequence assay was labeled at its 3′ end; the nuclease S1-cut fragment was labeled at its 5′ end, allowing examination of the S1 cutting at the nucleotide level on each strand. A four-nucleotide correction (increase) must be made in the nuclease S1 lanes because of the four nucleotides added by the 3′-end-filling during sequence determinations.

RESULTS

Identification of the Nuclease S1-Sensitive Site in Pro-α2(I) Collagen 5′ Flanking Gene Region. To identify nuclease S1-hypersensitive sites in supercoiled plasmids, the plasmid pXf3/CgPR was first incubated with nuclease S1, 5′-end-labeled, and then cut with the appropriate restriction enzyme. When either the vector pXf3 or the recombinant plasmid pXf3/CgPR was digested with nuclease S1 and then with restriction enzyme *Pst* I, a 430-bp fragment was obtained (Fig. 2, lanes 2 and 3). This corresponds to the distance between the pBR322 *Pst* I site and the major S1 site in the vector (8, 10). In addition, *Pst* I digestion of the recombinant plasmid generated a 100-bp fragment not obtained with *Pst* I digestion of the vector alone. Because there is a *Pst* I site at −71 in the pro-α2(I) collagen flanking gene region (17), this strongly suggests the S1-hypersensitive site is located 100 bp 5′ to the *Pst* I site, or about 170-bp 5′ to the transcription start site.

The approximate location of this site was confirmed by nuclease S1 digestion of the recombinant plasmid, followed by *Eco*RI or *Hin*dIII digestion. *Eco*RI digestion generated a 190-bp fragment, whereas *Hin*dIII digestion generated a 290/300-bp doublet (Fig. 2, lanes 4 and 5); in addition, there were large fragments resulting from nuclease S1 scission of the pBR322 S1 site followed by *Eco*RI and *Hin*dIII digestion. Occasional cutting at both the pBR322 and collagen S1 sites generated minor fragments of 1360 and 2200 bp, as shown by digestion of the recombinant plasmid with nuclease S1 alone, which produced these two bands in addition to the linearized plasmid (Fig. 2, lane 6). Because the 1360-bp fragment was not found in the *Eco*RI digestion (Fig. 2, lane 4) and the 2200-bp fragment was not found in the *Hin*dIII digestion (Fig. 2, lane 5) of the S1-digested recombinant plasmid, the collagen S1 site could be located in about the middle of the 400-bp "promoter" fragment (Fig. 3). The endpoint of each of the digests did not map to precisely the same location because, unlike restriction enzymes, nuclease S1 is expected to recognize a small region rather than a specific sequence, and the region it recognizes may extend over a number of nucleotides. In addition, DNA fragments occasionally display anomalous mobilities on polyacrylamide gels (22). Hence, it is likely that the location provided in this way for the S1 site may be further refined.

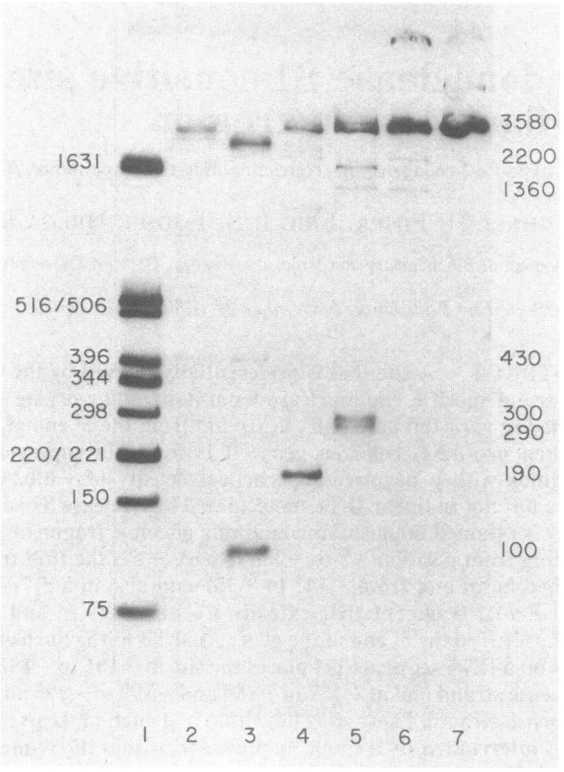

FIG. 2. Autoradiograph of a gel showing fragments resulting from nuclease S1 digestion of pXf3/CgPR followed by digestion with various restriction enzymes. Lanes: 1, pBR322 *Hin*fI size markers shown in bp; 2, vector pXf3 digested with nuclease S1 and *Pst* I; 3–5, pXf3/CgPR digested with nuclease S1 and *Pst* I, *Eco*RI, and *Hin*dIII, respectively; 6, pXf3/CgPR digestion with nuclease S1 only; 7, pXf3 digested with *Eco*RI and then digested with nuclease S1. In addition to the bands corresponding to the 3143-bp linearized vector (lane 2) and the 3580-bp linearized plasmid, pXf3/CgPR, (lanes 4–7), each of the lanes contain other high molecular weight bands corresponding to the complement of the small (110–430 bp) fragments displayed on the gels, and their sizes can be predicted from the restriction map of pXf3/CgPR shown in Fig. 3. For example, the complement (long distance from restriction site to nuclease S1 site) in lane 4 is 3400 bp, while it is 3300 bp in lane 5. Bands representing high molecular weight fragments also include the 2200-bp band for the collagen and pBR322 nuclease S1 sites (clockwise) in lanes 3, 4, and 6 and the 1360-bp band for the collagen and pBR322 nuclease S1 sites (counterclockwise in lanes 5 and 6). See text for additional explanation.

The Nuclease S1-Hypersensitive Site Is a Result of Superhelical Density of the Plasmid. The dependence of the nuclease S1-hypersensitive site on the superhelical density of the plasmid was demonstrated initially by first digesting the recombinant plasmid with *Eco*RI and then incubating it with nuclease S1. Only the linearized plasmid was produced (Fig. 1, lane 7); the sensitivity to nuclease S1 disappeared. To provide further evidence for this dependence on superhelical

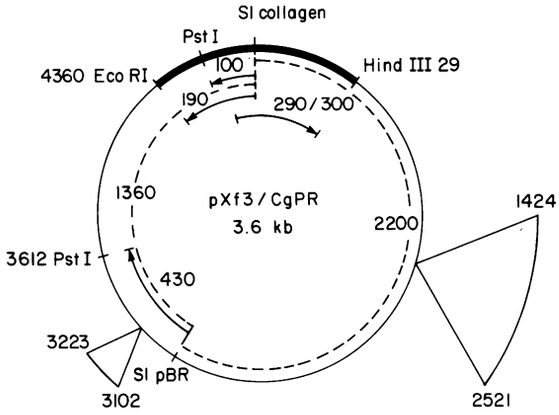

FIG. 3. Restriction map of pXf3/CgPR showing the location of the collagen promoter nuclease S1 site. The coordinates are those of pBR322; however, nucleotides 3102–3223, which contain two minor pBR322 nuclease S1 sites, and nucleotides 1424–2521, which contain the simian virus 40 poison sequence, have been removed (18). Arrows show the distances between restriction enzyme sites and nuclease S1 sites.

density, a series of topoisomers of pCg293, a derivative of the 420-bp *Sma* fragment, cloned into the *Hin*dIII site of pBR322 was prepared. When the recombinant plasmid was relaxed or had a negative superhelical density of $\sigma \leq -0.02$, only two fragments could be visualized on ethidium bromide-stained gels after incubation with nuclease S1 followed by digestion with *Pst* I. These fragments are 3.7 and 1.0 kilobase(s) (kb), corresponding to the distances between the *Pst* I site in pBR322 and in the collagen promoter fragment. The 3.7-kb fragment spans the pBR322 S1-sensitive site, while the 1.0-kb fragment spans the collagen S1-sensitive site (Fig. 4, lanes 4–7). At a superhelical density of −0.024, however, the collagen S1-hypersensitive site appeared and part of the 1.0-kb fragment was cleaved into 0.9- and 0.1-kb fragments (Fig. 2, lane 3). Thus, the hypersensitive site in pBR322 was not susceptible to nuclease S1 cleavage until a negative superhelical density of $\sigma = -0.036$ was reached. Then some of the 3.7-kb fragment was cut into 3.2- and 0.5-kb fragments (Fig. 2, lane 1). This suggests a difference in the nature of the pBR322 and the collagen S1-sensitive sites.

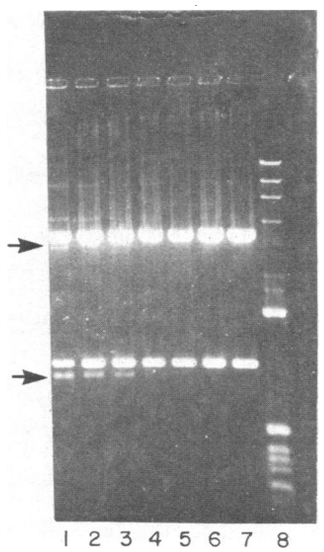

FIG. 4. Dependence of nuclease S1 sites on negative superhelical density of topoisomers of pCg293. Superhelical density in lanes 1–7 is −0.036, −0.030, −0.024, −0.019, −0.012, −0.006, and 0. pCg293 was digested with nuclease S1 and then digested with *Pst* I. Digestion with *Pst* alone would generate two fragments of 3.7 and 1 kb. The nuclease S1 site in pBR322 results in converting the 3.7-kb fragment to 3.2 and 0.5 kb. Only the larger fragment is visible on these gels in lane 1. The nuclease S1 site in the collagen promoter region converts the 1.0-kb fragment into 0.9 and 0.1 kb (lanes 1–3). Lane 8 shows *Hin*dIII-digested phage λ DNA and *Hin*fI-digested pBR322 size markers.

The Collagen Nuclease S1-Sensitive Site Is Conserved in Different Plasmids. The effector of nuclease S1 sensitivity may be colocated with the S1-sensitive site itself, or it may reside in other nearby sequences. The DNase-hypersensitive site in the 5′ flanking gene region of the *D. melanogaster* heat shock gene, *hsp 70*, located at −124 is influenced by upstream sequences (14), whereas the nuclease S1 sensitivity of cruciform structures in supercoiled plasmids has been shown to be a local property (9).

To determine which of these possibilities pertains to the collagen S1-sensitive site, the 416-bp *Sma* I fragment was either truncated or left within the 5.7-kb *Eco*RI restriction fragment as shown in Fig. 1. Digestion with *Hin*fI produced a 293-bp fragment from which the 126 bp at the 5′ end, including a stretch of 16 As, had been removed. Digestion with *Hpa* II produced a 204-bp *Hpa* fragment (as well as five smaller *Hpa* fragments) extending from −147 to −352 from which both the TATA box and CAT box had been removed. These fragments were subcloned in pBR322, and the resultant plasmids were digested with nuclease S1, followed by digestion of pCg293 with *Pst* I and of pCg204 with *Sal* I. The resultant fragments are shown in Fig. 5 *Left* and *Center*. In both plasmids the S1-sensitive site of the collagen 5′ flanking gene region remained sensitive to S1. Without S1 digestion, *Pst* I generated a 3.7-kb fragment containing the pBR322 S1 site and a 1.0-kb fragment spanning the collagen promoter S1 site. Digestion of pCg293 with S1 and *Pst* resulted in about 20% of the plasmid being susceptible to S1 at its pBR322 (3.7 kb → 3.2 + 0.5 kb) site, while at least 50% of the plasmid was

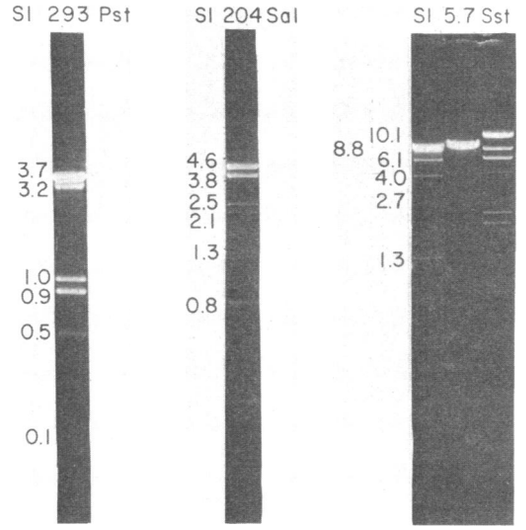

FIG. 5. Ethidium bromide-stained gels showing nuclease S1-hypersensitive site in the collagen promoter-containing plasmids. (*Left*) pCg293 digested with nuclease S1 and *Pst* I showing that at least half of the 1-kb fragment was digested by nuclease S1 to form 0.9- and 0.1-kb fragments. (*Center*) pCg204 digested with nuclease S1 and *Sal* I showing that about half of the 4.6-kb plasmid was digested into 3.8- and 0.8-kb fragments. (*Right*) pCg5.7 digested with nuclease S1 and *Sst* I (left lane); pCg5.7 digested with only *Sst* I (center lane); and λ phage *Hin*dIII fragments of 23.5, 9.6, 6.6, 2.2, and 2.1 kb (right lane). Approximately 40% of the 10-kb plasmid was cut into 8.8-kb and 1.3-kb fragments as a result of scission at the collagen S1 site. The identification of the top band in the left lane as a doublet of 10.5 and 8.8 kb is based on finding two clearly identifiable bands on an autoradiogram of ^{32}P-labeled fragments.

susceptible at its collagen S1 site (1.0 kb → 0.9 + 0.1 kb). The plasmid pCg204 has a single *Sal* I site and, hence, would only be linearized by digestion with *Sal* I alone. When preincubated with S1, however, four bands in addition to that of the linearized plasmid were clearly visible: 4.6 kb → 3.8 and 0.8 kb because of the collagen S1 site and 4.6 kb → 2.4 and 2.2 kb because of the pBR322 S1 site. The 1.4-kb fragment that resulted when both S1 sites were cut on the same plasmid was a very faint band because this was a rare event. The collagen S1-hypersensitive site was again cut in about half of the plasmids because the 3.8-kb band was almost as intense as the 4.6-kb linearized plasmid band (Fig. 5 *Center*). Hence, any sequences influencing the S1-hypersensitive site are contained within a region extending from −147 to −285.

Having established that truncated pro-α2(I) collagen 5′ flanking gene fragments retained their nuclease S1 hypersensitivity, we next determined if this site would be maintained in a 5.7-kb subclone of genomic DNA. This *Eco*RI fragment extends 1.2 kb 5′ and 4.1 kb 3′ to the 0.42-kb *Sma* I fragment and was subcloned in the *Eco*RI site of pBR322. Supercoiled plasmids were incubated with nuclease S1 and then digested with *Sst* I, as well as digested with *Sst* I alone. The latter

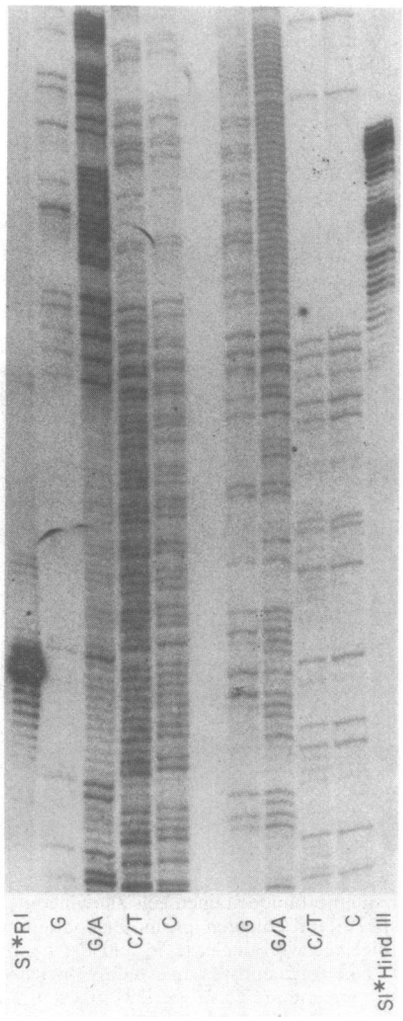

FIG. 6. Autoradiograph displaying 5′-end-labeled fragments released by nuclease S1 scission followed by digestion with *Eco*RI (sense strand) or by digestion with *Hin*dIII (nonsense strand) next to the sequence ladder. DNA sequence ladder obtained by 3′-end-filling after either *Eco*RI (*Left*) or *Hin*dIII (*Right*) digestion of pXf3/CgPR and, hence, a four-nucleotide correction must be made in lanes labeled S1. The sequence reads 5′ → 3′ (top to bottom) on each strand.

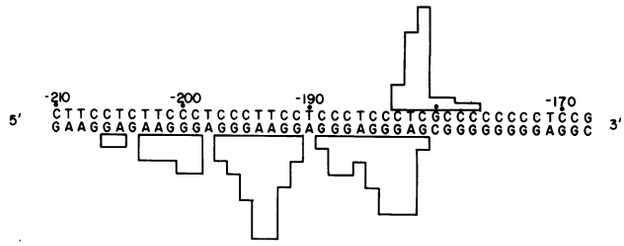

FIG. 7. Identification of collagen promoter nuclease S1-sensitive site at the nucleotide level. The height of each bar is based on a densitometer scan of the lanes containing the nuclease S1 fragments on the sequence gel, as described in the text.

resulted, as expected, in a linear 10-kb plasmid (Fig. 5 *Right*, lane 2), whereas the former produced a doublet corresponding to 8.8-kb and 1.3-kb fragments resulting from scission of the 10-kb fragment at the collagen S1 site (Fig. 5 *Right*, lane 1). The 6- and 4-kb fragments were generated by the pBR322 S1 site. A very faint 2.7-kb fragment again reflected the rare event in which the plasmid was cut at both the pBR322 and the collagen S1 sites. These results show that the signal generated by the S1 site in the pro-α2(I) collagen 5′ flanking gene region is transmissible and extends its dominance over 10 kb of DNA.

Identification of the Nuclease S1 Site at the Nucleotide Level. From the sizes of the fragments produced by nuclease S1 digestion followed by digestion with restriction endonucleases, the S1 site could be localized to within about 100 bp, some 200 bp from the transcription start site. To locate this site more precisely, the S1-generated fragments were analyzed on a DNA sequence gel. The fragments that resulted from S1 treatment, followed by 5′-end-labeling and either *Eco*RI or *Hin*dIII digestion, were analyzed on a gel containing the DNA sequence analysis products of both strands of this fragment as shown in Fig. 6 and interpreted in Fig. 7. The S1-sensitive site was clearly found within a stretch of 42 pyrimidines interrupted by a single G, located 169 bp 5′ to the transcription start site. The nuclease S1 appeared to cut the sense strand at essentially one site (−181 to −182), whereas it clearly cut the other strand at two major sites (−182 to −184 and −192 to −195) separated by 10 bp. The base composition of this region is 73% G+C. It contains two direct repeats of the symmetrical sequence T-C-C-C-T-C-C-C-T, separated by T-C-C. The most S1-sensitive site on the nonsense strand is on the boundary of the first of these, and that on the sense strand is at the boundary of the second one.

Mace *et al.* also observed that the nuclease S1 site in the *D. melanogaster* gene was located at the boundary of two repeats of C-T sequences (14). They suggested a slippage mechanism similar to the one first suggested by Hentschel (23), in which the first hexamer on one strand would pair with the second hexamer on the other strand, causing the original complements of the two hexamers to loop out and partly pair with each other. If this concept were sufficient, slippage would be expected to be even more probable at a site 100 nucleotides 5′ to the pyrimidine stretch where a tract of 16 adenine residues are located (17), which is not cut by nuclease S1. Although it is thermodynamically favorable for A+T-rich regions to be denatured in tortionally strained circular DNA molecules (24), the fact that nuclease S1 cuts within the higher-melting pyrimidine tract suggests that S1 is recognizing an altered DNA structure rather than low-melting regions of DNA.

DISCUSSION

The existence of a nuclease S1-hypersensitive site has been verified within the pro-α2(I) collagen 5′ flanking gene region in supercoiled plasmids *in vitro*. It is located 180 bp 5′ to the transcription start site. A DNase I-hypersensitive site 200 bp

upstream from the 5′ end of the α2(I) gene has been identified in chromatin isolated from expressing tissues and very young embryos in which type I collagen gene expression is detectable but minor (25). The fact that the site maps to a pyrimidine stretch and not one of the palendromic sequences was surprising at first. However, other S1-hypersensitive sites have been mapped to pyrimidine-rich regions (14, 23). Moreover, nucleosomes do not readily form on poly(dC)-poly(dG) or poly(dA)-poly(dT) (26): this might explain why some sites that are S1 sensitive in supercoiled plasmids are also S1 sensitive in chromatin. Of course, we do not know to what extent an oligopyrimidine-oligopurine mimics synthetic homopolymeric polydeoxyribonucleotides. It seems reasonable to assume that, even if the pyrimidine-rich region can be part of a nucleosome structure, the resulting structure would be less stable than those formed in other regions of the gene.

The existence of DNase I-hypersensitive sites in chromatin correlates well with gene expression in most studies made (5, 6) and also correlates with a nuclease S1-hypersensitive site in the chicken major β-globin gene and adenovirus chromatin (12) and with the site mapped *in vitro* in supercoiled plasmids (11, 13). However, it must be noted that in the *D. melanogaster* heat shock genes, the system most extensively studied to date (3, 4, 14), the DNase I hypersensitive site located 5′ to the genes, is present in chromatin isolated from normal embryonic cells that have never been subjected to heat shock (4). Moreover, DNase I-hypersensitive sites were identified 5′ to the adult β-globin gene in chicken embryo fibroblasts transformed with Rous sarcoma virus in which the embryonic globin gene, not the adult gene, was being expressed (7). Thus, the existence of a DNase I-hypersensitive site 5′ to the cap site is neither a requirement for expression nor is it evidence that the gene is being expressed. With this limitation in mind, it is still of interest to determine whether the nuclease S1- and DNase I-hypersensitive site in the pro-α2(I) collagen 5′ flanking gene region plays any role in regulating the expression of this gene.

We thank Gert Pflugfelder and Larry Peck for their help in the preparation of topoisomers, Sirpa Aho for another DNA sequence analysis of the 416-bp *Sma* I fragment, Douglas Hanahan for the plasmid pXf3/CgPR, and Elizabeth Levine for her excellent technical assistance. Finally, we want to thank Nancy Pegg for her assistance in preparing this manuscript. This research was supported by grants from the National Institutes of Health.

1. Weintraub, H. & Groudine, M. (1976) *Science* **193,** 848–858.
2. Garel, A. & Axel, R. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 3966–3970.
3. Wu, C. (1980) *Nature (London)* **286,** 854–860.
4. Keene, M. A., Corces, V., Lowenhaupt, K. & Elgin, S. C. R. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 143–146.
5. Elgin, S. C. R. (1981) *Cell* **27,** 413–415.
6. Igo-Kemenes, T., Horz, W. & Zachau, H. G. (1982) *Annu. Rev. Biochem.* **51,** 89–121.
7. Groudine, M. & Weintraub, H. (1982) *Cell* **30,** 131–139.
8. Lilley, D. M. J. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 6468–6472.
9. Lilley, D. M. J. (1981) *Nucleic Acids Res.* **9,** 1271–1289.
10. Panayotatos, N. & Wells, A. D. (1981) *Nature (London)* **289,** 466–470.
11. Singleton, C., Keysik, J., Stirdivant, S. M. & Wells, R. D. (1982) *Nature (London)* **299,** 312–316.
12. Larsen, A. & Weintraub, H. (1982) *Cell* **29,** 609–622.
13. Weintraub, H. (1983) *Cell* **32,** 1191–1203.
14. Mace, H. A. F., Pelham, H. R. B. & Travers, A. A. (1983) *Nature (London)* **304,** 555–557.
15. Goding, C. R. & Russel, W. C. (1983) *Nucleic Acids Res.* **11,** 21–36.
16. Vogeli, G., Ohkubo, H., Sobel, M. E., Yamada, Y., Pastan, I. & de Crombrugghe, B. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 5334–5338.
17. Tate, V., Finer, M., Boedtker, H. & Doty, P. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **47,** 1039–1049.
18. Hanahan, D. (1983) *J. Mol. Biol.* **166,** 557–580.
19. Tate, V. E., Finer, M. H., Boedtker, H. & Doty, P. (1983) *Nucleic Acids Res.* **11,** 91–104.
20. Peck, L. J., Nordheim, A., Rich, A. & Wang, J. C. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4560–4564.
21. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
22. Maniatis, T., Jeffrey, A. & Van de Sande, H. (1975) *Biochemistry* **14,** 3787–3794.
23. Hentschel, C. C. (1982) *Nature (London)* **295,** 714–716.
24. Wang, J. C. (1974) *J. Mol. Biol.* **87,** 797–816.
25. Merlino, G. T., McKeon, C., de Crombrugghe, B. & Pastan, I. (1983) *J. Biol. Chem.* **258,** 10041–10048.
26. Simpson, R. T. & Künzler, P. (1979) *Nucleic Acids Res.* **6,** 1387–1415.