A Sequence Conserved in Both the Chicken and Mouse \( \alpha_2(\text{I}) \) Collagen Promoter Contains Sites Sensitive to \( S_1 \) Nuclease*

(Received for publication, November 22, 1983)

Catherine McKeon, Azriel Schmidt, and Benoit de Crombrugghe

From the Laboratory of Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205

We have examined the \( S_1 \) nuclease sensitivity in the promoter of both the chicken and mouse \( \alpha_2(\text{I}) \) collagen genes. When these DNAs are introduced into supercoiled plasmids and digested with \( S_1 \) nuclease, a discrete region containing one or more cleavages is found in each promoter. These \( S_1 \) cleavage sites were mapped by the distance of the \( S_1 \) site from known restriction enzyme cleavage sites. In the chicken gene, the \( S_1 \)-sensitive segment is located 180 to 200 base pairs preceding the start site of transcription, whereas in the mouse promoter it is between -145 to -165 base pairs. This site in the chicken promoter maps to the segment that has previously been shown to be \( S_1 \) and DNase I hypersensitive in chromatin. Although these \( S_1 \) sites are found at different distances from the start site of transcription in the two promoters, the sequences at these sites are strongly conserved between the two species. Each sequence consists of an identical tandem repeat containing a short palindrome within each repeat. Since the DNA sequence does not exhibit the features that would favor either a left-handed Z-DNA configuration or a cruciform structure, an alternative model is discussed that could account for the \( S_1 \) sensitivity of these sequences. The conservation of these sequences and their \( S_1 \) sensitivity suggests they play a role in the activation or regulation of the \( \alpha_2(\text{I}) \) collagen gene promoters.

\( S_1 \)-sensitive sites in naked supercoiled plasmids were first described in bacterial plasmids in which perfect inverted repeats form a cruciform structure under the stress of supercoiling (1, 2). Recently, another DNA structure has been shown to be \( S_1 \) sensitive, namely the boundary between right-handed B-DNA and left-handed Z-DNA (3). The transition from right-handed B-DNA to left-handed Z-DNA is favored in certain DNA sequences by supercoiling and the bases at the transition point are probably left unpaired (3). \( S_1 \)-sensitive sites have been found in several eukaryotic DNAs when these segments were cloned in supercoiled plasmids even though no sequence with the potential to form a cruciform appeared to be present (4-6). One such sequence from a sea urchin histone gene spacer region contained a stretch of 45 bases with an alternating GA sequence which might slip, causing locally unpaired bases (4). In the case of promoter elements, \( S_1 \)-sensitive sites have been demonstrated in both chromatin and in supercoiled plasmids (5).

We have studied the \( S_1 \) sensitivity of two supercoiled plasmids containing the 5′ end of the \( \alpha_2(\text{I}) \) collagen gene from either chicken or mouse. Both promoters contain discrete \( S_1 \)-sensitive sites. In the chicken promoter, there is a major \( S_1 \)-sensitive site which is found approximately 180 bp upstream of the start site of transcription. This site maps within the DNA sequence of \( S_1 \) hypersensitive regions which are found in the promoter of this gene when the DNA is in a chromatin structure (7). The analogous mouse promoter contains an \( S_1 \)-sensitive site that maps around -145. The DNA sequences at these sites are homologous in the mouse and chicken promoters and consist of minor variations of the tandemly repeated palindrome, CCTCCCC.

**MATERIALS AND METHODS**

Three recombinant plasmids were used in this study: pCol-CAT, pCol-CAT ΔBgl and pCol-Md. Plasmid pCol-CAT is a derivative of the expression vector pSV-CAT (8) in which a 1.3-kb region of the chicken \( \alpha_2(\text{I}) \) collagen gene promoter is fused to the bacterial chloramphenicol acetyltransferase gene (CAT), replacing the SV40 early promoter (constructed by H. Okubu and C. Gorman). The rest of the plasmid consists of 2.4 kb of pBR322 sequences which contains the origin of replication of pBR322 and 1.5 kb which contains segments of the large T antigen gene and the poly(A) addition signal of the early transcription unit of SV40 (Fig. 1A). The chicken \( \alpha_2(\text{I}) \) collagen promoter region contains sequences between 1200 bp preceding the start site of transcription to 108 bp following this start site. Plasmid pCol-CAT ΔBgl is a modified version of pCol-CAT in which an internal deletion has removed 1000 bp between -103 to -1100 bp. In this construct, the TATA box and the CAT box of the chicken collagen promoter remain intact.

Plasmid pCol-Md contains a 1.25-kb fragment of the mouse \( \alpha_2(\text{I}) \) collagen promoter cloned between the EcoRI site and the HindIII site of pBR322. The mouse \( \alpha_2(\text{I}) \) collagen promoter region contains sequences between 507 bp preceding to 753 bp following the presumed start site of transcription (Fig. 1B). We have numbered the mouse promoter sequence by assigning the presumed start site of transcription as +1 based on sequence homologies with the chicken \( \alpha_2(\text{I}) \) collagen gene (9). In the latter, the start site of transcription has been experimentally determined (10). The sequence around this site is conserved in the mouse gene as is the distance between this conserved sequence and a TATAAATA present in both genes.

Recombinant plasmids were isolated from Escherichia coli by a sodium dodecyl sulfate lysis procedure (11). Plasmid preparations were centrifuged to equilibrium two times on an ethidium bromide/cesium chloride gradient to isolate supercoiled plasmids. Between 5 and 10 \( \mu \)g of the plasmid was digested with increasing amounts of \( S_1 \) nuclease (0-100 units) (P-L Biochemicals) in \( S_1 \) digestion buffer (250 mM NaCl, 30 mM sodium acetate (pH 4.6), 1 mM ZnSO\(_4\)) for 10 min at 25 °C. In some experiments (Fig. 4), the NaCl concentration was altered to 50, 150, and 300 mM. Samples were fractionated by electrophoresis on a 7.5% acrylamide gel in 80 mM Tris-borate buffer (pH 8.3). The gel was attuned with ethidium bromide, destained in Tris-borate buffer and photographed under ultraviolet light illumination.

* 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.

\(^1\) The abbreviations used are: bp, base pairs; kb, kilobase pairs.
RESULTS

S1-sensitive Sites in the Chicken Collagen Promoter—When the supercoiled plasmid, pCol-CAT, is cleaved by S1 nuclease, several discrete and reproducible cleavages are observed. These S1 cleavage sites can be mapped in the promoter region by cleaving the DNA with restriction enzymes subsequent to the S1 digestion and by gel electrophoretic determination of the size of the fragments that are generated.

In the plasmid containing the chicken promoter, there are two major S1-sensitive sites. The plasmid pCol-CAT (Fig. 1A) contains a unique HindIII recognition site (Fig. 2, lane A). When the plasmid is digested by S1 prior to HindIII cleavage, two additional discrete DNA species appear after gel electrophoresis (Fig. 2, lane B). The size of the slower moving DNA species (marked by an asterisk in Fig. 2, lane B) is consistent with the occurrence of an S1-sensitive site, which maps in the pBR322 segment of the plasmid (at 3060 bp in the pBR322 map near the origin of replication) as identified independently by Lilley (1) and Panayotatos and Wells (2). The other band of 290 bp corresponds to another S1-sensitive site that maps at a distance of 290 bp from the HindIII site. This 290-bp DNA fragment, like the larger one, is generated by the action of both the S1 and HindIII enzymes and appears as a doublet at higher DNA concentrations (Figs. 2 and 4). Because the S1 cleavage site could map on either side of the HindIII site, we performed two other experiments to localize it. We first cleaved the same S1-digested plasmid with SmaI. SmaI cleaves within the promoter segment at +8 and at –430, releasing a 440-bp fragment from the plasmid (Fig. 2, lane C). As seen in Fig. 2, lane D, four additional fragments are generated by S1 digestion, one resulting from an S1 cleavage in pBR322 presumably at the same site as the one observed in Fig. 2, lane B, whereas the others measure 250, 230, and 190 bp. Since the SmaI site is 104 bp from the HindIII site, the major bands in both the SmaI and HindIII digestions localize an S1-sensitive site at 180 bp preceding the start of transcription. The minor bands in the SmaI digestion suggest that another site is located more distal to the start site of transcription. This site is not easily seen after HindIII digestion since HindIII cleaves on only one side of these sites. The results of the digestion with BglII which cleaves on both sides of the S1 sites are consistent with the location of a pair of S1 sites as diagrammed in Fig. 2 (data not shown). Under the digestion conditions used, approximately 10% of the plasmid molecules show cleavages in the collagen insert. This low rate of cleavage is probably due to preferential cleavage at the cruciform structure in the pBR322 which linearizes the plasmid. To confirm the location of the S1-sensitive site, in the collagen promoter, we digested a plasmid with a deletion in the a2(I) collagen promoter which removes from –103 to –1100 in the pCol-CAT plasmid. In such a plasmid cleaved with HindIII after S1 digestion, the S1-dependent 290-bp band is no longer generated, whereas the site in pBR322 is still sensitive to S1 digestion (Fig. 2, lanes I and J).

S1-sensitive Sites in the Mouse Collagen Promoter—When the supercoiled plasmid (Fig. 1B) containing a 1.25-kb fragment of the mouse collagen promoter cloned into pBR322 is treated with S1, specific cleavages also occur. Treatment of plasmid pCol-Md with either BstEII, EcoRI, or BglII which cleave only once in the plasmid, produce a linear
plasmid of 5.5 kb (Fig. 3, lanes A, C, and G, respectively). When plasmid pCol-Md is treated with S1, first and then cut with BstEII (Fig. 3, lane B) an additional species of 225 bp is visible. Treatment of plasmid pCol-Md with S1 and then with EcoRI produces DNA species of around 900 and 1600 bp (Fig. 3, lane D). The 1600-bp species is due to the combined digestion of pBR322 by S1 and EcoRI and its length is again consistent with the location of an S1-sensitive site mapping to the same location in the plasmid as the site marked by an asterisk in Fig. 2, lanes E and J. The other S1-sensitive site can be mapped to -145 in the mouse promoter by its distance from the unique BstEII site and unique EcoRI site as diagrammed in Fig. 3. When the plasmid is treated by S1 and digested by BglII which cleaves on the other side of these sites, a band of 190 bp is observed (Fig. 3, lane H). This maps at -165. Taken together, the location of the S1 site in these three digestions suggests that, as in the chicken promoter, there is a region of 20 bp that is digested by S1 nuclease in the mouse promoter.

Characterization of the S1-sensitive Sites—The S1 sites reported by other investigators are altered in their nuclease sensitivity by changing the salt concentration or the torsional stress imposed by supercoiling (4,5). The S1 sensitivity that results from a probable cruciform structure in DNA is dependent on supercoiling and occurs at relatively low salt (50 mM) (1,2), whereas the S1 sensitivity at junctions between right-handed B-DNA and left-handed Z-DNA is increased by high salt and supercoiling (Fig. 4). The cleavages observed at 50 mM NaCl that map from 70 to 180 bp upstream of the start site of transcription probably represent slippage and mispairing in these regions where there are many CG residues (Fig. 4, lane F). At 300 mM NaCl, slippage is minimized and a stable intermediate is maintained by high salt and supercoiling (Fig. 4, lanes J and K). These findings are different from those of Hentschell (4) and may reflect the differences in the degree of slippage in each model.

A comparison of the structural elements of the two promoters is depicted in Fig. 5A. The sequence in the region between -179 and -202, where the S1-sensitive sites of the

![Fig. 3. Map of the S1-sensitive site in the mouse α2(I) collagen promoter.](image)

![Fig. 4. The characterization of the S1-sensitive site in the chick α2(I) collagen promoter.](image)
**DISCUSSION**

We have identified discrete sites in both the chicken and mouse α2(I) collagen promoters that are sensitive to S1 nuclease when the promoter is in a supercoiled plasmid. The sites are found in a segment that is almost exclusively composed by pyrimidines on one strand. In each promoter, the pyrimidine-rich sequence contains a tandemly repeated sequence and each repeat itself contains a short palindromic sequence. These pyrimidine-rich sequences do not show the characteristics of sequences expected to adopt a left-handed Z-DNA configuration, *i.e.*, alternating purine-pyrimidine sequences. Since these sequences do not contain an inverted repeat, they would also not be expected to adopt a cruciform structure. During the course of this work, a structure was proposed by Mace et al. (6) to account for the S1 sensitivity of a *Drosophila* heat shock gene promoter. We suggest a similar model whereby the first repeat on one strand base-pairs with the second repeat on the other strand (or vice versa) producing two staggered single-stranded loops as depicted in Fig. 5C. The formation of a staggered loop would relieve the stress introduced by DNA supercoiling and could account for the S1 sensitivity of this segment in supercoiled plasmids. Once the staggered loop structure was formed to relieve strain on the molecule, high salt would be expected to stabilize the structure by neutralizing the charges on the DNA.

If the staggered loop structure is important for the expression or regulation of the collagen gene, one would expect the sequence necessary to form such a structure to be evolutionarily conserved. In fact, these segments show a high degree of sequence homology in the tandemly repeated region between the chicken and the mouse, species which are separated by 250 million years of evolution (15). In our model, the conserved sequences are those which would pair to form the staggered loop. In contrast, the sequences flanking the tandem repeats, which would not be involved in the formation of the staggered loop structure, have diverged considerably.

The importance of the sequences upstream of the chicken α2(I) start site of transcription is being investigated by creating deletions in the promoter and by transferring these deletions to an expression vector in which the promoter region is fused to the bacterial chloramphenicol acetyltransferase gene. After transfection of animal cells, the levels of chloramphenicol acetyltransferase enzyme can be measured in cell extracts and are a reflection of the strength of the promoter located upstream of the CAT gene (8). Deletions upstream of -103, such as pCol-CAT *Bgl* (Fig. 2, H, I, and J) which delete the S1-sensitive site, decrease CAT expression severalfold even though the TATA and CAT box regions of the promoter remain intact, suggesting that regions upstream of -103 are important for the expression of the α2(I) collagen gene. 3 One of the segments responsible may be the S1-sensitive region.

The location of the S1-sensitive sites in the chicken promoter falls within the -100 to -300 region, where we have previously localized a DNase I and S1-sensitive site in chromatin (7). As has been shown for other genes, the presence of a DNase I hypersensitive site in the α2(I) collagen gene promoter is tissue-specific since it is found in chromatin from cells known to synthesize collagen, such as fibroblasts, but is absent in brain where collagen is not synthesized. 2 In other genes as well, the S1-sensitive site that is found in supercoiled recombinant plasmid DNAs maps to the location of the DNase I hypersensitive site in chromatin (5). A pyrimidine-rich sequence similar to the one present in the collagen promoters is conserved in the three *Drosophila* heat shock genes and has been shown to be S1-sensitive. In addition, this sequence falls within the region of the heat shock promoter shown to be DNase I sensitive in chromatin (16, 17). In the promoter of the chicken β-globin gene as well, pyrimidine-rich sequences are present within the DNase I hypersensitive region of the promoter (18) which contain S1 cleavage sites in both chromatin and in naked supercoiled DNA (5). In the promoters of these different genes, S1 sites fall within segments that are strikingly CT-rich, in one strand over a 50-bp region. The percentages of CT in S1-sensitive regions ranges from 67% for chicken β-globin (18), 84% for *Drosophila* heat shock (19), to 96% for the chicken α2(I) collagen gene. Because the S1-sensitive sites of these different promoters in naked supercoiled DNA map to the same region as those in chromatin, the proposed structure of the S1-sensitive sites for supercoiled plasmid DNA may be analogous to their structure in chromatin.

Pyrimidine-rich stretches similar to those found in the α2(I)

---

3 C. McKeon, unpublished results.
collagen promoters might not bind nucleosomes as tightly as other DNA segments. Indeed, nucleosomes do not bind well to long tracts of either poly(T)/poly(A) or poly(C)/poly(G) oligomers (20, 21). If the DNase I hypersensitivity results from the dissociation of one or a few nucleosomes, these long pyrimidine-rich stretches would highly favor such dissociation (18, 22). As a consequence, the segments where the nucleosomes would have dissociated would come under a higher tension due to supercoiling (22). In an area that is free of nucleosomes, the sequences such as those described here could then adopt a single-stranded staggered loop configuration which might explain their S1 sensitivity in chromatin. The staggered loops shown in Fig. 5C could, of course, be stabilized by a DNA binding protein. The similarity between these regions in several genes from different species suggests that the region may be an important regulator of transcription.

Acknowledgments—We wish to thank Glenn T. Merlino and Gene Liau for critical reading of the manuscript, Ann Schombert for typing, and Ray Steinberg for photography.

REFERENCES
A sequence conserved in both the chicken and mouse alpha 2(I) collagen promoter contains sites sensitive to S1 nuclease.

C McKeon, A Schmidt and B de Crombrugghe


Access the most updated version of this article at [http://www.jbc.org/content/259/10/6636](http://www.jbc.org/content/259/10/6636)

Alerts:

- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 0 references, 0 of which can be accessed free at [http://www.jbc.org/content/259/10/6636.full.html#ref-list-1](http://www.jbc.org/content/259/10/6636.full.html#ref-list-1)