Structural and Functional Analysis of the Chick Chondroitin Sulfate Proteoglycan (Aggrecan) Promoter and Enhancer Region*

(Received for publication, September 24, 1996, and in revised form, February 5, 1997)

Edward W. Pirok III‡‡, Hao Li§§, James R. Mensch, Jr.§, Judith Henry§, and Nancy B. Schwartz¶¶

From the Departments of ‡‡Pathology, §§Pediatrics, and ¶¶Biochemistry and Molecular Biology, and **Committee on Developmental Biology, University of Chicago, Chicago, Illinois 60637

Aggrecan is a large chondroitin sulfate proteoglycan, the expression of which is both tissue-specific and developmentally regulated. Here we report the cloning and sequencing of the 1.8-kilobase genomic 5′ flanking sequence of the chick aggrecan gene and provide a functional and structural characterization of its promoter and enhancer region. Sequence analysis reveals potential Sp1, AP2, and NF-I related sites, as well as several putative transcription factor binding sites, including the cartilage-associated silencers CIIS1 and CIIS2. A number of these transcription factor binding motifs are embedded in a sequence flanked by prominent inverted repeats. Although lacking a classic TATA box, there are two instances in the 1.8-kb genomic fragment of TATA-like TCTAA sequences, as have been defined previously in other promoter regions. Primer extension and S1 protection analyses reveal three major transcription start sites, also located between the inverted repeats. Transient transfections of chick sternal chondrocytes and fibroblasts with reporter plasmids bearing progressively reduced portions of the aggrecan promoter region allowed mapping of chondrocyte-specific transcription enhancer and silencer elements that are consistent with the sequence analysis. These findings suggest the importance of this regulatory region in the tissue-specific expression of the chick aggrecan gene.

During development, the extracellular matrix is a complex dynamic structure, the components and organization of which help to establish the requisite position and state of differentiation. The large chondroitin sulfate proteoglycan (CSPG),1 aggrecan, has been localized predominantly to skeletal tissue and is considered to be a hallmark of cartilage differentiation. In chick cartilage, aggrecan expression begins at embryonic day 5 in limb rudiments, continues through the entire period of chondrocyte development, and remains a biochemical marker of the cartilage phenotype thereafter. In very early embryos, aggrecan is expressed in the notochord as early as stage 16, long before chondrogenesis occurs (1).

We have extensively studied the properties and expression of aggrecan from embryonic chick cartilage. These studies include synthesis and processing (2–5), structural analysis via peptide sequencing to elucidate glycosylation motifs, and a consensus sequence for O-glycosylation and mapping of the S103L monoclonal antibody epitope (6–10). Moreover, we have conducted molecular analysis to construct the composite sequence of chick cartilage CSPG from overlapping cDNAs and to identify a defect in the aggrecan gene associated with the chondrodystrophy, nanomelia (9, 11).

This sequence, obtained from 10-day-old chick embryos, has 6464 nucleotides that include an open reading frame encoding 2109 amino acids and 16 nucleotides of the first untranslated exon (11). Another chick aggrecan cDNA sequence, obtained from embryonic chick brain, was 6597 nt in length, including 265 nt of 5′-untranslated exon sequence (12). Using chick CSPG cDNA probes, we subsequently isolated genomic clones containing exons encoding the chick CSPG core protein. The two 5′ globular domains, G1 and G2, are encoded by four and three exons, respectively, and the interglobular domain is encoded by a single exon. The chondroitin sulfate attachment domain is encoded by the largest exon, 3216 bp, which is approximately 50% of the total coding sequence. These data reveal that the chick CSPG gene contains at least 18 exons spanning more than 30 kb. No evidence was obtained for multiple genes for aggrecan in the chick genome. Elucidation of the genomic organization of chick aggrecan has allowed for a more thorough comparison with the mammalian aggrecans, as well as the avian and mammalian link proteins, with respect to origin and mechanisms of divergence. A summary of this work was published recently (13).

We have also found that aggrecan is developmentally expressed, in ovo and in limb bud cultures, on both protein and mRNA levels in a pattern commensurate with the onset of chondrogenesis. The modulation of expression of this cartilage-specific CSPG and type II collagen mRNA in stage 24 limb bud mesenchyme cells cultured in high density was examined under conditions that promote chondrogenesis in vitro (14) and mimics the same process in limb development in ovo. Morphologically, mesenchymal proliferation ceases by day 2, condensation occurs first in the formation of aggregates by days 4–5, and then of overt nodules by days 6–8, concomitant with cellular differentiation and production of matrix. Quantitatively, a 50-fold increase in aggrecan mRNA occurs from day 2 (when first detected) to day 6, followed by a slight decline (about 2-fold) by day 8 when the message reaches a plateau thereafter (15). This same pattern is observed immunologically, using the monoclonal antibody S103L, which is specific to the aggrecan protein. These studies indicate that during limb development...

---

*A The abbreviations used are: CSPG, chondroitin sulfate proteoglycan; nt, nucleotide(s); kb, kilobase(s); bp, base pair(s); SP, signal peptide; PCR, polymerase chain reaction.

**This work was supported by U. S. Public Health Service Grants AR-19622 and HD-09402, a grant from the Brain Research Foundation (to N. B. S.), Training Grant HL-07237 (to H. L.), and a Fellowship from AR-19622 and HD-09402, a grant from the Brain Research Foundation (to E. W. P.). 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”.

Deceased.

‡‡ To whom correspondence should be addressed: The University of Chicago, 5841 S. Maryland Ave., MC 5058, Chicago, IL 60637. Phone: 312-702-6426; Fax: 312-702-9234.

1 The abbreviations used are: CSPG, chondroitin sulfate proteoglycan; nt, nucleotide(s); kb, kilobase(s); bp, base pair(s); SP, signal peptide; PCR, polymerase chain reaction.
the expression of these two differentiation-specific proteins are stringently controlled until the establishment of the cartilage phenotype. Thereafter, aggrecan continues to be synthesized and deposited in the extracellular matrix, perhaps to effect a decrease in cell adhesion necessary for maintenance of the chondrogenic state.

Concurrent with studies of mechanisms that control the temporal-spatial aspects of cartilage differentiation are structural and functional analyses of expression of the differentiation-specific products of the extracellular matrix. For instance, significant work has been done to understand the tissue-specific expression of collagen genes and the mechanisms that regulate their distinct transcriptional programs (16–18). In contrast, there have been no studies of the transcriptional regulation of the aggrecan gene that examine its tissue-specific expression during development. Mouse aggrecan has been cloned; however, no functional analysis has been performed to examine its tissue specificity (19). A preliminary characterization of the rat aggrecan promoter has also appeared, describing a 120-bp sequence containing transcription start sites (20). It is not clear whether this 120-bp genomic fragment contains tissue-specific control elements, because the 5′ promoter/enhancer region is probably larger or may contain additional regulatory elements. The same report described promoter assays on a larger isolate containing an additional 520 bp of 5′ flanking sequence, but the sequence data were not presented.

Therefore, to begin to elucidate the mechanisms that govern aggrecan expression in chondrocytes, we have cloned the promoter region of the embryonic chick S103L-reactive CSPG (aggrecan). The aim of the present study was to identify and characterize the cell- and stage-specific elements in the 5′ genomic flanking region of the aggrecan gene, which could regulate the expression of this extracellular macromolecule during embryonic development.

**EXPERIMENTAL PROCEDURES**

**Materials**—Oligonucleotides were made with an Applied Biosystems 3800 DNA synthesizer. Reagents for biochemical and molecular cloning experiments were of the highest quality available from commercial vendors. Restriction endonucleases were from New England Biolabs unless otherwise stated. T4 DNA ligase, T4 kinase, S1 nuclease, avian myeloblastosis virus reverse transcriptase, and Klenow polymerase were from Promega. Taq polymerase was from Perkin-Elmer. A chick genomic library was purchased from CLONTECH Laboratories.

**Preparation of Probe and Screening of Chick Genomic Library**—A chick aggrecan cDNA fragment comprising 260 bp of the 5′ untranslated exon was inserted into the vector pGEM-4Z by standard methods (21). The insert was excised from pGEM-4Z by Bgl II/Bss H1 restriction enzyme cutting sites. The PCR fragments were gel-purified, digested with Bgl II and Kpn I, and ligated directly into the pGEM-4Z vector. The 1.8-kb insert was excised from pGEM-4Z by EcoRI-Kpn I digestion, treated with Klenow polymerase, and blunt-end ligated into the reporter vector pGL2-Basic (Promega), which had been linearized with the restriction enzyme Nhe I and treated with Klenow. The reporter vector pGL2-Basic does not contain any eukaryotic promoter or enhancer elements. Sequences of the amplified fragments were of the highest quality available from commercial vendors. Taq polymerase was from Perkin-Elmer. A chick genomic library was purchased from CLONTECH Laboratories. The same report described promoter assays on a larger isolate containing an additional 520 bp of 5′ flanking sequence, but the sequence data were not presented.

**Sequence Determination and Analysis**—Dideoxynucleotide chain termination sequencing (22) of the Bgl II/Bss H1 DNA fragments subcloned into pGEM-4Z plasmids was performed using the U. S. Biochemical Sequenase (version 2.0) system. Primers were T7 or SP6 promoter primers (Promega) or 18–20-mer oligonucleotides synthesized according to the obtained sequence. Multiple sequence determinations were made for each primer used. Ambiguities in sequencing were resolved by using a different polymerase (e.g. avian myeloblastosis virus reverse transcriptase), sequencing the complementary strand, or both. All residues were confirmed by at least two separate sequence determinations. DNA sequence analysis was performed using the Wisconsin Package (23). Searching for palindromic sequences was done using the program COMPARE to find inverted repeats by comparing the sequence to its own complement (24), and the results were displayed via the program DOTPLOT. Putative transcription factor binding sites were located with the program FINDPATTERNS using the pattern file tfsite.dat, which comprises the Transcription Factor Database (25).

**Analysis of the Chick Aggrecan Promoter Region**

**Cell Cultures**— Cultures of day-14 chick sternal chondrocytes were established according to the procedures described by Cahn et al. (26) and as modified by Campbell and Schwartz (3). Cultures of fibroblasts were established from skin of day-10 chick embryos following trypsinization (3). Cells were plated at an initial density of 1.5 × 10^5/100-mm tissue culture dishes (Falcon) in either F-12 medium (chondrocytes) or Dulbecco’s modified Eagle’s medium (fibroblasts) and supplemented with 10% fetal calf serum. The cells were permitted to attach to the dishes, and subsequent growth (2–3 days) was maintained by a complete change of the medium every 2 days (2). On the day of transfection, chondrocyte cultures were trypsinized, and single cells were suspended in F-12 medium, replated, and allowed to attach to the dishes for 3–4 h before treatment as described below.

**Transfection**—Standard methods were followed for transient calcium phosphate transfections (21). Duplicate plates containing approximately 5 × 10^5 cells (either chondrocytes or fibroblasts) received 20 pmol of a given plasmid construct to be assayed. Five μg of a β-galactosidase plasmid reporter plasmid pGL2-Basic (Promega) were cotransfected with each experimental construct to correct for cell loss. Duplicate transfection sets were repeated three times, each time yielding similar results. The transfections were allowed to proceed for 36 h. The relative efficiency of transfecting the chondrocytes was approximately 13% that of transfecting the fibroblasts.

**Cell Recovery and Assays**—Reagents for the luciferase and β-galactosidase assays were purchased from Promega. Because both luciferase
Fig. 1. Chick aggrecan genomic structure and cloning strategy. A, a schematic diagram of the genomic structure of the chick aggrecan gene. The names and sizes of the introns or exons are printed above the diagram. Exons are represented as open boxes, except the first untranslated exon, which is represented as a checkered box. Introns are represented as the lines connecting the boxes. The diagram is not to scale. B, the cloning strategy for the chick aggrecan promoter region. On the right side of the diagram is a chart that indicates the size, vector, and name used for each construct. As described under “Experimental Procedures,” cDNA from the untranslated and signal peptide exons was used to screen a chick genomic library. The 14-kb genomic fragment obtained was subcloned into the vector pGEM-4Z and is represented as a black rectangle with the checkered pattern indicating the region of overlap with the first untranslated exon. The 14-kb fragment was digested with BglII and BbsI, and the resultant 1.8-kb fragment was subcloned into the sequencing vector pGEM-4Z and the luciferase reporter vector pGL2-Basic, which does not contain a eukaryotic promoter region. Each orientation of the genomic inserts was confirmed by sequencing.

**Assays and β-galactosidase assays** were performed, Promega’s Reporter Lysis Buffer (RLB, E2071) was used to prevent the inhibition of β-galactosidase activity that occurs in buffers containing detergents such as Triton X-100. No deviations were made from the manufacturer’s protocol for preparation of extracts from tissue culture cells. The enzymatic activity of luciferase was measured with a luminometer (Analytical Luminescence Laboratory, Monolight 1500). The enzymatic activity for β-galactosidase was measured with a microplate reader (Dynatech) at 405 nm. Standard deviations were determined for the six assays performed on duplicate plates within one experiment.

**End-labeling of Probes for mRNA 5′ End Mapping**—The Z2 or Z3 oligonucleotides (Z2, 5′-AAATTCCTCTGGTTGATTTACCCAGTCTTTTCAGGCC-3′, nt 193–226; Z3, 5′-GAAAGAGACCATCATAAACCTCTGTACGCTCTC-3′, nt 68–101) for primer extension experiments or S1 analysis were end labeled using [γ-32P]ATP and T4 DNA kinase according to standard protocols (21). Three ethanol precipitations were performed to remove the residual [γ-32P]ATP from the labeled oligonucleotides.

**S1 Analysis of mRNA Using Single-stranded DNA Probes**—Established methods were used to perform S1 analysis (27). Single-stranded probes were made from the double-stranded 900(+) or D(+)-plasmids. Plasmids were alkali-denatured, and a 32P-5′-end-labeled oligonucleotide primer, Z2 or Z3, was annealed to the template, 900(+) or D(+), and extended with Klenow (Promega). Probes were cut to the appropriate 5′ length by digestion with restriction enzyme KpnI. The single-stranded probes were separated from the template DNA by alkali low-melting-point agarose electrophoresis, and radiolabeled bands were cut out and purified by phenol extraction and ethanol precipitation (21). Approximately 5000 cpm of probe was hybridized to 25 μg of total RNA from day-14 chick sternal chondrocytes. The hybridization occurred at 55°C for 12 h in an aqueous hybridization solution (21). The resultant RNA:DNA hybrid was digested with 200 units of S1 nuclease for 60 min. The products were electrophoresed in 6% polyacrylamide sequencing gels.

**Primer Extension**—Approximately 5000 cpm of labeled Z2 or Z3 probe was hybridized to 25 μg of RNA derived from day-1 chick sternal chondrocytes. Hybridization was done in S1 hybridization solution for 12 h at 30°C (21). Extended products were produced by treating the hybrid RNA-primer with 40 units of avian myeloblastosis virus reverse transcriptase (Promega). Products were extracted in phenol/chloroform, precipitated in ethanol, and electrophoresed on 6% polyacrylamide sequencing gels.

**RESULTS**

**Structural Analysis of the 5′ Portion of the Chick Aggrecan Gene**—To guide functional studies, the complete 1.8-kb Ag-1 sequence was determined and found to comprise 1875 bp (Fig. 3). Examination of the sequence revealed the lack of a classical TATA box or CCAAT box. When the Ag-1 fragment was analyzed for transcription factor binding sequences, it was found that at least 202 potential sites were present, including putative AP2 and Sp1 binding sites. The relative positions of some of these eukaryotic transcription factor-associated sequences are indicated in Fig. 3. The numbering of the sequence is relative to the most upstream transcription start site (as detailed below). The Ag-1 sequence was also compared with known promoter sequences in the eukaryotic promoter data base (EPD) using the National Center for Biotechnology Information BLAST server (25), and no extensive identity with other promoter sequences was found. However, tracts of multiple dA and dT residues, analogous to those found in Ag-1 in the ranges 250 to 280 and −144 to −78, respectively, were seen to occur in many other described promoter regions. These dA and dT tracts, in particular the dT16 from −87 to −78 and the dA21 from 250 to 270, constitute an inverse repeat or palindrome with the potential to give rise to a pair of large stem-and-loop structures or a cruciform structure (28). Hence, additional analyses were performed on the Ag-1 sequence with the aim of detecting other, less obvious, palindromic structures.

The Ag-1 sequence from positions −300 to 340 was analyzed by comparison to its own reverse complement sequence with the Wisconsin Package program COMPARE. The dot plot reveals a widely spaced pair of inverted repeats centered around −100 and 250, corresponding to the dT and dA tracts, separated by over 300 bp. However, no other potential secondary structures of comparable scale are seen in this sequence with the window/stringency parameters used in this analysis; a few less prominent repeat pairs occur in the downstream third of the sequence. Interestingly, the putative Sp1, AP2, and TFII sites in addition to other potential factor-specific sequences, as well as all three of the mapped start sites, lie in the putative loop portion of this potential structure. Such secondary structures, in addition to potential transcription factor binding sites, may be involved in mechanisms by which the aggrecan message is developmentally regulated.

**Determination of Transcription Starting Sites**—Two methods, S1 analysis and primer extension, were used to locate
the sites where transcription of the aggrecan mRNA is initiated. Because the 5′-untranslated cDNA sequence previously reported by this laboratory (11, 12) overlaps with the 3′ end of the Ag-1 genomic isolate by 145 nucleotides, transcription initiation occurs still farther upstream in Ag-1. Templates used to generate single-stranded DNA probes for S1 analysis included the 900(+) and D(+) plasmid constructs, as represented in Fig. 4C. S1 analysis with the downstream primer Z2 yielded three major protected fragments: 226 bp, 187 bp, and a 69/70-bp doublet, corresponding to start sites at positions 1, 40, and 157–158 (Fig. 4A, lanes 1 and 2). Position 1 in Fig. 3 is defined as the farthest 5′ transcription starting site. These locations were obtained with probes generated from both the 900(+) and the D(+) constructs. The two upstream transcription start sites at positions 1 and 40 were confirmed with the downstream primer Z3-generated probes, again using the 900(+) and D(+) constructs as DNA templates (Fig. 4B, lanes 4 and 5). Z3-generated probes from the 900(+) and D(+) constructs gave protected fragments of 101 and 62 bp, respectively, confirming the position 1 and 40 transcription starting sites. The Z3 primer lies upstream of the 157/158 transcription starting site.

Functional Analysis of the Aggrecan Promoter Sequence—Transient transfections of day-14 chick embryo sternal chondrocytes with the construct Ag-1(+) (the forward orientation of the 1.8-kb insert in the promoter/enhancer-free pGL2-Basic reporter vector) revealed a plasmid dose-dependent level of luciferase expression (Fig. 5A), i.e., increasing concentrations of transfected construct produced increases in luciferase activity, establishing that the 1.8-kb region contains elements capable of promoter function. In subsequent experiments, constructs Ag-1(+) and Ag-1(−), in addition to the negative control pGL2-Basic vector with no insert, were transiently transfected into both 14 day-old chick sternal chondrocytes and, to examine tissue specificity, into 10 day-old chick embryo fibroblasts. Transfections with either the negative control pGL2-Basic vector with no insert or the Ag-1(−) construct resulted in much lower luciferase expression, with activity equivalent to background in both transfected chondrocytes and fibroblasts.

A series of constructs that progressively deleted the Ag-1(+) sequence was used to relate the locations of potential transcription factor binding sites and secondary structure to promoter

![Schematic representation of the pGL2-Basic vector and the primers used to synthesize the deletion constructs. A, the 1.8 genomic fragment Ag-1. The checkered pattern represents the 145-bp identity with the aggrecan cDNA. Above the diagram are the restriction enzymes used to excise this genomic fragment from the clone G8. Below are shown the names and relative positions of the primers used to make the deletion constructs. To the right side of the diagram is a chart that indicates the coordinates that the primer pairs span, the size of the resultant PCR product, and the names of the constructs. B, below the diagram is a list of all of the primers used to make deletion constructs with the Ag-1 sequence as a template. The boldface represents sequences not found in Ag-1 but added to engineer restriction enzyme cutting sites. Note that prZ0 resides in the vector pGL2-Basic and not in the Ag-1 sequence. C, a schematic representation of the pGL2-Basic vector and the relative positions of the restriction sites used in these experiments. The arrow represents the direction of transcription of the vector. The (+) and (−) orientations are defined as the positionings of the insert with respect to luciferase gene in the same or reversed ways as it occurs with respect to the aggrecan coding sequence.](image-url)
function and tissue specificity. The constructs and transfection results are summarized in Fig. 6. The initial deletion removed approximately 500 bp from the upstream end of the Ag-1(+) construct, as well as a tract of 21 dA residues from the downstream end. The resulting construct, 1300(--), produced a modest increase in luciferase activity in chondrocytes versus that

![Image](image_url)
promoted by the construct Ag-1(+). Transfected fibroblasts showed little difference in luciferase activity from Ag-1(+) to 1300(+); the latter was slightly lower. Deletion of another 500 bp from the 5’ end (including a CIIS2 site) generated the construct 900(+); this deletion had a dramatic effect, because both chondrocyte and fibroblast luciferase yields nearly tripled when compared with assays of the original Ag-1(+) construct (to 140- and 30-fold, respectively). Although chondrocyte activity remained substantially higher than that in fibroblasts, there was a greater proportional increase in luciferase activity in fibroblasts, 260% when compared with the 1300(+) construct in fibroblasts versus a 160% increase in chondrocytes. This increase may be due to loss of tissue specificity or to coincidental but independent effects of silencers in both cell types.

FIG. 4. S1 analysis and primer extension. Conditions are described under “Experimental Procedures.” A, the results of S1 analysis, sequencing, and primer extension from the oligonucleotide Z2. Lane 1, S1 protection bands resulting from the D(+) -derived probe spanning nucleotides –69 to +226; lane 2, products resulting from the probe derived from the 900(+) construct, spanning the region –638 to +226; lane 3, results of a primer extension experiment using 32P-end-labeled oligonucleotide Z2. Lane 4, products from the probe spanning the region –638 to +101, derived from 900(+); lane 5, results of a primer extension experiment using 32P-end-labeled oligonucleotide Z3. Arrows, the location of the major bands. The bands at position 157–158 consistently appear as a doublet in both S1 analysis and primer extension experiments. Only bands that were generated in both types of experiments were marked; other bands are potentially artifactual because they cannot be duplicated in the complementary experiment. C and D schematically show the design and results of the S1 protection and primer extension experiments, respectively. Open boxes, the radiolabeled oligonucleotide Z2; slashed boxes, the radiolabeled oligonucleotide Z3. Bricks, RNA; p above the RNA, the determined transcription start sites.
site) produced the 500(+) construct. Promoter activity in chondrocytes returned to approximately 50-fold, similar to that assayed for the constructs Ag-1(+) and 1300(+); yet in fibroblasts, luciferase activity of the 500(+) construct was only slightly lower than that seen for the 900(+) construct (Fig. 6). This finding suggests that the upstream half of 900(+) may contain enhancer elements that are used in chondrocytes.

A newly generated construct A(+), 590 bp, was made that was similar to the 500(+) construct, except that the insert contained the 3′ stretch of poly(dA) regions and 36 bp in the 5′ direction to include the putative IgHC.21 site (Fig. 3). These changes produced a modest increase in luciferase activity in chondrocytes only. Measured luciferase activity in fibroblasts modestly decreased when compared with the luciferase activity measured from fibroblasts transfected with the 500(+) construct. The deletion construct B(+), 547 bp, which does not contain the IgHC.21 site, lost approximately 40% of the activity of the A(+) construct in chondrocytes; the activity in fibroblasts was reduced by 70%, resulting in luciferase activity as low as that seen for many of the (−) constructs. A further deletion construct, D(+), 376 bp, which included only the three transcription start sites and the putative Sp1 and AP2 binding sites, produced a significant amount of luciferase activity in chondrocytes (nearly 60-fold), and in transfected fibroblasts luciferase activity was equivalent to the 1.8-kb Ag-1(+) construct. The D(+) construct deleted the poly(dT) region but included the poly(dA) region. The 308-bp construct, E(+), included the three major start sites at positions 1, 40, and 157/158 but did not include the consensus sequences Sp1-CS4, GGMT-IIF, and AP2-CS4. Deletion of these potential nuclear factor binding sites expressed a 75% loss of activity in chondrocytes while not substantially altering luciferase activity in transfected fibroblasts. Construct E(+) had comparable luciferase activity in both chondrocytes and fibroblasts of approximately 15-fold when compared with the no-insert control vector. The 140-bp construct F(+) did not include any of the determined starting sites and produced modest luciferase activity in transfected chondrocytes and baseline luciferase activity in transfected fibroblasts. In all but one instance, the reverse orientation constructs of all of these genomic fragments yielded minimal luciferase activity in both transfected chondrocytes and fibroblasts. That exception, the activities seen for the 500(−) construct, suggests that some low-level promoter activities may result from largely accidental sequence assemblages. In sum, the data suggest the following functional roles for portions of the aggrecan gene: 1) general repression upstream of the pr900 site, especially between −638 and −1038 (pr1300); 2) strong chondrocyte-specific enhancement in the pr900-pr500 interval (−638 to −247); 3) a positive element, possibly IgHC.21, occurs in the small prA-prB interval (−283 to −240); 4) the prB-prD segment (−240 to −69) has a negative role, strongest in fibroblasts; and 5) the small (−69 to −1) pd-prE interval, bearing Sp1 and AP2 elements, is stimulatory in chondrocytes. It is also apparent that constructs lacking either the dT or dA tracts (e.g., 900(+) and D(+)) are quite active; therefore, interaction between these repeats is not required for promoter function in this system.

**DISCUSSION**

We have found that a 1.8-kb genomic fragment from the 5′ end of the chick aggrecan gene is able to drive expression of the pGL2-Basic luciferase reporter gene in a tissue-specific manner. Determining the sequence of this construct revealed more than 202 potential transcription factor binding sites. This structural information allowed us to proceed with a functional analysis of the effects of potentially active cis elements that may confer tissue and developmental specificity on expression of the aggrecan gene by using a series of nested deletion constructs. These sequences ranged from the full 1.8 kb (Ag-1(+) or) to a minimal 140-bp construct (F(+)).

Of the numerous potential cis elements found in the Ag-1 sequence, several are of particular interest with respect to control of aggrecan expression. Positions −873 and −721 in the Ag-1 sequence are the 5′ ends of two copies of the sequence CACCTCC (CIIS2), which has been suggested to be a silencer.
motif in the COL2A1 promoter (29). This particular sequence has been shown to inhibit transcription of the type II collagen promoter in fibroblasts while not significantly changing expression in chondrocytes (29). Indeed, this seems to be consistent with our results because deletion of these two motifs from the 1300(+) to 900(+) constructs reduced the cell type specificity of luciferase expression while the overall promoter activities increased. This motif is also present in the promoter region of COL4A2; however, tissue-specific regulation in fibroblasts versus chondrocytes remains to be investigated in this system (30).

The chick aggrecan 5′ flanking region contains a second silencer consensus sequence, (CIIS1) ACCCTCTCT (29) at position 127, which is also found in COL2A1. The CIIS1 sequence occurs in an interspersed rat repetitive sequence (31) and in another repetitive sequence found in the avian genome named the CR1 element (32, 33). Further negative regulatory functions have been shown in the chick lysozyme gene (34), rat insulin gene (31), mouse IgH gene (35), human β-interferon gene (36), and the human ε-globin gene (37). In the Ag-1 sequence, this motif is located within 200 bp downstream of the putative Sp1 site. A "push and pull" mechanism has been proposed for transcriptional regulation in two systems, the low density lipoprotein receptor gene and the COL2A1 gene (29, 38). This model proposes that the sterol-dependent binding of a protein to a consensus sequence could inhibit the positive activation of a nearby Sp1 binding site (38); such a silenced element acting in a "push and pull" mechanism could likewise be responsible for the temporal and tissue-specific regulation of the aggrecan gene.

The Ag-1 sequence contains one putative NF-I site at position −1282. The NF-I proteins are transcriptional activators derived from a multigene protein family in the vertebrate phylum (39–42). Chick tissues contain NF-I products that are derived from four separate genes that have the potential of producing 12 isoforms (42). Recently, it has been shown that the silencer SI is very similar to the NF-I/CTF family, and an additional silencer, SII, is similar to an NF-I/CTF half-site (43). This suggests that NF-I-related proteins can mediate transcriptional repression in cells of mesenchymal origin (42). Our sequence does not contain the sequence motifs of SI or SII, but Szabo et al. (43) suggest that the NF-I family of regulator proteins can be modulated as silencers in addition to their previously accepted role as activators. The presence of a putative NF-I site raises the possibility of mesenchyme-specific regulation controlled by this element in addition to possible modulation by unreported silencers, thus creating a more dynamic system than one based solely on NF-I activation.

From footprinting analysis, Long and Linsenmayer (44) reported a novel transcription factor binding sequence, ACACAGA, acting in the regulation of COL10A1, and suggested that this factor may act as a silencer. The proximal promoter region of COL10A1 is responsible for regulating expression in hypertrophic chondrocytes (44). Our reported sequence contains four positions, −1140, −491, 151, and 214, where the

FIG. 6. Structure and differential promoter functions of the aggrecan 5′ flanking region. A, a schematic of the genomic structure of the chick S103L-reactive CSPG (aggrecan) gene. B, the set of deletion constructs. Inserts derived from the aggrecan promoter/enhancer region were ligated into the pGL2-Basic (Promega) reporter vector, which carries the luciferase gene. Both the forward and reverse orientations were constructed as indicated by (+) or (−). Subsequent deletion constructs were generated by PCR using the construct Ag-1(+) as a template (see "Experimental Procedures." C, relative luciferase promoter activity of the various deletion constructs. The activity of the reporter vector with no insert was defined as one and used to calculate the activities of the other constructs. Duplicate plates were transfected, and each plate was assayed for luciferase activity three times. All experimental details are as presented in the legend to Fig. 5.
The chick aggrecan promoter exhibits <40% sequence similarity to either the mouse promoter (19) or the 120-bp rat (20) promoter fragment, indicating that this promoter/enhancer region is not highly conserved across the taxa. Interestingly, the untranslated first exon in chick aggrecan contains less than 45% similarity compared with rat, mouse, or human sequences (19, 20, 46). Although the lack of identifiable similarity between the chick and mammalian aggrecan first exons might be attributable to the existence of fewer selection pressures on an untranslated sequence, this argument is not readily extended to promoter sequences. Also puzzling is that although the rat and mouse promoter sequences share 93% identity with each other, none of the described transcriptional start sites coincide with each other in these two similar promoter regions.

There are, however, similarities in TATA-binding motifs among promoters of cartilage-specific genes. As is the case for the mouse and rat aggrecan promoter regions, the chick 5’ flanking sequence lacks a classical TATA box and contains multiple transcriptional start sites (19, 20). Although a TATA-less promoter with multiple GC-rich regions is the hallmark of many housekeeping genes (47), many other genes that are temporally regulated have been shown to have promoters with similar structures (48, 49). It is interesting that the 5’ flanking sequence of the chick link protein gene also contains multiple transcriptional start sites and lacks a classical TATA box (50); rather, it has a TATA motif-like sequence TCTAA (51). The chick aggrecan sequence contains two TCTAA motifs, one that is 31 bp and another that is 94 bp upstream of the start sites at positions 40 and 157–158, respectively (Fig. 3). The TCTAA sequence is also present in the human and chick link protein promoter region (50, 52) and in the serine/glycine-rich proteoglycan (51). However, human link protein has only one transcription start site (52). Thus, it would be interesting to determine whether the human aggrecan sequence has only one transcription start site, which would provide further evidence for similarity in the evolution of the link protein and the aggrecan genes, as has been suggested (13).

Overall, this study has established the 5’ flanking sequence as having at least three major transcription start sites in addition to several putative cis elements and a potential secondary structure that may control expression of the aggrecan gene. We have demonstrated tissue-specific promoter activity with the 1.8-kb region and have systematically mapped subregions that produce activation or repression of downstream reporter genes in two cell types in culture. This study paves the way for more directed studies of the individual cis elements identified and their interaction with trans-acting factors so that we may better understand the mechanisms by which the aggrecan gene is regulated.