Original Contributions

Mammalian Genome
Incorporating Mouse Genome

Mammalian Genome 10, 1119–1125 (1999).

Completion of the mouse aggrecan gene structure and identification of the defect in the *cmd-Bc* mouse as a near complete deletion of the murine aggrecan gene

Richard C. Krueger, Jr.,* Kiyoto Kurima, Nancy B. Schwartz

Departments of Pediatrics and Biochemistry and Molecular Biology, University of Chicago, MC 5058, 5825 S. Maryland Avenue, Chicago, Illinois 60637, USA

Received: 20 May 1999 / Accepted: 26 July 1999

Abstract. Mouse cartilage matrix deficiency (*cmd*), an autosomal recessive phenotype caused by absence of aggrecan, maps to Chromosome (Chr) 7 and is caused by a 7-bp deletion in exon 5 generating a premature stop codon (Watanabe et al. 1994). Another spontaneous mutation with the same locus and phenotype, *cmd-Bc*, has now been defined as the complete loss of exons 2 to 18, resulting in a significantly shortened mRNA (1.2 kb). The upstream breakpoint is in intron 1, 18.8 kb 3′ of exon 1; the downstream breakpoint lies 10.5 kb past the final aggrecan exon 18. The deletion is flanked by sequences homologous to topoisomerase I and II cleavage sites and a 7-bp direct repeat, suggesting the defect resulted from a nonhomologous recombination event. Additionally, the size of the first intron and the intron-exon structure between exons 12 and 14 were determined, establishing the length of the murine aggrecan gene as 68.6 kb. This report completes the structural analysis of the murine aggrecan gene, defines a second null mutation, and reinforces the importance of aggrecan in development.

Introduction

The large aggregating chondroitin sulfate proteoglycan, aggrecan, is one of the predominant structural molecules of cartilage. Aggrecan forms macromolecular complexes with hyaluronic acid and link protein in the extracellular matrix (ECM), providing cartilage with its characteristic resiliency. The other major macromolecule of cartilage, type II collagen, provides the tissue with its characteristic strength. Defects in these major components, as well as other ECM molecules, lead to associated skeletal abnormalities (Li and Olsen 1997; Schwartz and Domowicz 1998).

A nearly complete gene structure for murine aggrecan consisting of 18 exons has been reported (Watanabe et al. 1995). The murine aggrecan protein core is 2132 amino acids long (Mr 259,131), comprising three globular domains (G1, G2, and G3), and two glycosaminoglycan attachment domains, for chondroitin sulfate and keratan sulfate chains, respectively. The previously reported size for the mouse aggrecan gene did not include the EGF-like domain (part of G3) encoded by exon 13. Furthermore, the length of the entire gene was not known since the size of the first intron had not been reported, nor had the intron/exon structure of the murine gene from the 3′end of exon 12 to the 5′ end of exon 14 been published. This structure is presumably important to aggrecan function because exon 13, encoding an EGF-like domain, is known to be alternatively spliced (Baldwin et al. 1989; Walcz et al. 1994; Valhmu et al. 1995; Fulop et al. 1996).

Cartilage matrix deficiency (*cmd*) is a well-characterized genetic disorder of mouse aggrecan (Rittenhouse et al. 1978; Kimata et al. 1981; Brown and Harne 1982; Kimata 1982; Brennan et al. 1983; Brown et al. 1991; Li and Olsen 1997). The mouse *cmd* mutation was spontaneous on an agouti background. Mice homozygous for the mutation exhibit cleft palate, short limbs, snout, and tail, and die at birth, presumably owing to respiratory failure related to pulmonary hypoplasia (Houghton et al. 1989). Heterozygous *cmd* mice have been reported to have no phenotypic abnormality detectable at birth, but later exhibit dwarfism and age-associated spinal degeneration, and have a shortened life span compared with homozygous normal mice (Watanabe et al. 1997). Although only the aggrecan gene is affected, there is disrupted expression of other matrix genes in the growth plate, including link protein, syndecan 3, and several species of collagens (Wai et al. 1998). Thus these mouse models will be useful for investigating the influence of matrix on chondrocyte differentiation as well as the regulatory mechanisms involved in extracellular matrix development (Schwartz and Domowicz 1998; Schwartz et al. 1999).

The *cmd* defect has been reported to be the result of a 7-bp deletion in aggrecan exon 5, resulting in a premature stop codon, a truncated aggrecan core protein, and no mature aggrecan product in the matrix (Watanabe et al. 1994). This mutation is not unlike the aggrecan defect reported for the nanomelic chick, in which there is a single nucleotide change that results in a premature stop codon (Li et al. 1993). Nanomelia and *cmd* have similar lethal phenotypes. Another spontaneous recessive mutation at the *cmd* locus, *cmd-Bc*, occurring on a BALB/c GaBc background (Bell et al. 1986) causes short limbs and snout, enlarged abdomen, protruding tongue, and cleft palate in newborns. Homozygotes can be recognized as early as E15 by reduced limb length and abnormal shape of limbs (Fig. 1). The *cmd* and *cmd-Bc* mutations do not complement in a double heterozygote (Bell et al. 1986), but the defect in the *cmd-Bc* mutation has not been defined. This report identifies the defect in *cmd-Bc* mice and also defines the length of intron 1 and the intron/exon structure surrounding exon 13, thereby completing the structural information for the murine aggrecan gene.

The nucleotide sequences reported in this paper have been submitted to the GenBank™/EMBL Data Bank with accession numbers AF144254 (murine aggrecan intron one sequence), AF144255 (*cmd-Bc* defect), and AF144256 (wild-type sequence containing the 3′ breakpoint). The Mouse Genome Database lists the *cmd* mouse mutants as Agc (*cmd*) and Agc (*cmd-Bc*).

Correspondence to: N.B. Schwartz.

* *Present address:* Cedars-Sinai Medical Center, Ahmanson Pediatric Center, Division of Neonatology, 8700 Beverly Blvd., Room 4310, Los Angeles, CA 90048, USA

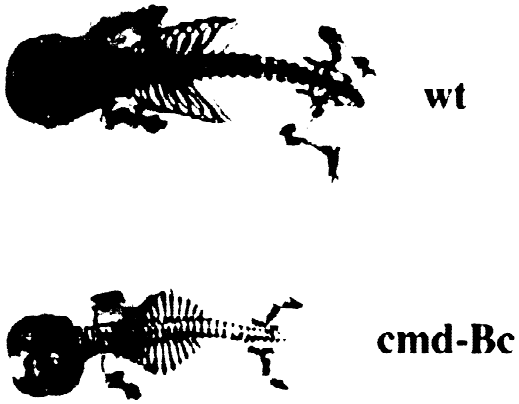


Fig. 1. Normal (wt) and cartilage matrix deficient-BC (cmd-Bc) mouse newborn skeleton. Calcified bone and cartilage were differentially stained with alzarin red and alcian blue, respectively.

Materials and methods

Genomic cloning and sequence analysis. A mouse C57BL/6 genomic BAC library (GenomeSystems, St. Louis, Mo.) was screened following the manufacturer's instructions with a 260-bp ^{32}P-labeled DNA fragment generated by PCR from exon 1 of the mouse aggrecan gene (primers #1 and #2-see Table 1 and Fig. 3A), with normal mouse liver genomic DNA as a template. A resulting clone, BAC-1, was digested with various restriction enzymes, and DNA fragments derived from intron 1 were subcloned into pBluescript® KS II (Stratagene, La Jolla, Calif.) for sequence analysis. Sequencing was done using the dRhodamine terminator cycle sequencing kit (Perkin-Elmer) and ABI model 377 DNA sequencer (Applied Biosystems, Foster City, CA).

Northern and Southern blot analysis. Epiphyseal cartilage total RNA, isolated from normal or *cmd-Bc* 16 to 17-day-old embryos, was used for Northern blot as described (Li *et al.* 1993). Hybridization was performed in QuikHyb® hybridization solution (Stratagene) according to the provided protocol with the aforementioned 260-bp ^{32}P-labeled PCR fragment from exon 1 of the mouse aggrecan gene as probe.

Genomic DNA was purified from the livers of known *cmd-Bc* mutant, heterozygous, and homozygous normal mice (Ausubel *et al.* 1997). Restriction fragment length polymorphisms (RFLPs) were detected by Southern blot analysis (Ausubel *et al.* 1997) of genomic DNA digested with *Nhe*I and *Kpn*I, with ^{32}P-labeled probes.

Synthesis and ^{32}P labeling of DNA probes. Probes for Southern and Northern blot analysis were synthesized in one of two ways. Probes derived from cDNA (exonic probes) were synthesized with primers based on the published sequence for aggrecan. cDNA was synthesized by RT-PCR using total RNA obtained from embryonic mouse cartilage and random hexamer priming, employing the SuperScript™ preamplification kit (GibcoBRL, Frederick, MD). Probes were synthesized that covered all 18 aggrecan exons.

Exonic cDNA probes were labeled by the random hexamer priming method with the RediPrime kit (Amersham, Arlington Heights, IL). Fragments derived from either intronic or intergenic regions were labeled with ^{32}P dCTP by PCR to produce high specific activity probes for Southern blot analysis. Briefly, a short PCR was performed under normal conditions (10 cycles of 94°C, 30 s; 60°C, 60 s; 72°C, 30 s), except that no cold dCTP was added to the reaction mix; instead, 5 μl of 3000 Ci/mmol ^{32}P dCTP (New England Nuclear, Boston, MA) was added; 2.5 μg of purified double-stranded DNA probe was used as template.

Inverse PCR of the cmd-Bc polymorphism. In order to determine the defect in the *cmd-Bc* gene, inverse PCR of a polymorphic fragment was performed (Ochman et al. 1990). Briefly, 20 μg of genomic DNA from the *cmd-Bc* mutant was digested overnight with *Nhe*I, then the fragments were circularized in a ligation reaction containing 1 μg of DNA per ml and T4 DNA ligase overnight at 4°C. The ligation reaction was extracted with phenol/chloroform and the DNA precipitated with ethanol. The DNA was dissolved in 10 μl of water and PCR was performed using primers #5 and #6. The amplified 3.8-kb band, as predicted by Southern blot analysis, was purified (QIAquick Gel Extraction Kit, QIAGEN, Santa Clarita, CA), and cloned into pCR® 2.1 using the TA Cloning® kit (Invitrogen, Carlsbad, CA) for sequence analysis.

Table 1. List of primers and their positions in the aggrecan gene.

No.	Gene location	Orientation	Sequence
1	Exon 1	Sense	5′-CATTCCCGCCCACCTACC
2	Exon 1	Antisense	5′-CCCTGCGAACCCCTTCTC
3	1230 bp 3′ of N1	Sense	5′-TGTCCTAGGTGGGATACGATGC
4	1523 bp 3′ of N1	Antisense	5′-AGACAAAGTTTAGTTCTTGCTCCCG
5	39 bp 3′ of N1	Antisense	5′-CTGATTTCACAGAGAGATGAAGTACGC
6	990 bp 3′ of N1	Sense	5′-CAGAGCTGTTGACCCCTGTACC
7	23 bp 5′ of N4	Antisense	5′-TTTGAGACAGAGTCCCTGCTGAC
8	391 bp 5′ of N4	Sense	5′-CAGGGACAGGCATAAGGGATG
9	Exon 18	Sense	5′-AGGCTACAGAAGCGGACGATG
10	365 bp 3′ of BP2	Antisense	5′-GTCCTGGCTTCTGCTGTGATATG
11	654 bp 3′ of BP2	Antisense	5′-GTGTGTGTGAAGGATGCTGAATAGAG

PCR of fragments longer than 10 kb. To span regions greater than 10 kb, PCR was performed with the XL PCR kit (Perkin-Elmer) with BAC-1 DNA as a template (Cheng et al. 1994). The reaction mixture was as follows per 100-μl reaction: 2 μg BAC-1 DNA (template) with 40 pmol of each primer in 1 × XL Buffer II containing 1.2 mM $Mg(OAc)_2$, 800 μM dNTP, and 4 units rTth DNA XL polymerase. Conditions were as follows: 94°C, 1 min (hot start, first cycle only); 16 cycles of 94°C for 30s, then 68°C, 10 min; 12 cycles 94°C 30 s, 68°C, 10 min + 15 s increment/cycle; then 72°C, 10 min. Amplifications routinely yielded single agarose gel bands, which were purified with the QIAEX®II Gel Extraction System (QIAGEN) and sequenced directly.

Results

That the *cmd-Bc* mutation was significantly different from that reported for *cmd* was suggested by the following preliminary results. Using an exon 1 probe, genomic Southern blot analysis revealed no polymorphism between wild-type and homozygous *cmd-Bc* mutant mice. However, there was no hybridization of any probe derived from exons 2 through 18 to *cmd-Bc* genomic DNA; all exons were tested (results not shown). Further, PCR was successful only for exon 1, and unsuccessful for exons 2–18, when either cDNA obtained from *cmd-Bc* mouse cartilage RNA or *cmd-Bc* liver genomic DNA was used as template.

Northern blot analysis of total RNA obtained from normal and mutant littermates using the exon 1 probe revealed that the mutant mRNA was 1.2 kb in length, compared with 8.5 kb for the wild-type mRNA (Fig. 2). Although the size of the *cmd* mouse aggrecan mRNA was not previously reported (Watanabe et al. 1994), it would be assumed that the mRNA size would be largely unaffected by a 7-bp deletion in the gene. There also appeared to be less aggrecan mRNA in the *cmd-Bc* mutant than in normal cartilage, as judged by the intensity of the band. This is commensurate with data from *cmd* mouse and nanomelic chick (Li et al. 1993), where Northern blot analysis indicated a reduction in aggrecan mRNA levels in both mutants compared with wild-type controls. The data from Northern and Southern blot analyses strongly suggested that the *cmd-Bc* aggrecan gene was deleted after exon 1, that the 5′ breakpoint was within the first intron, and the 3′ breakpoint was downstream of the 3′ end of the gene.

In order to determine the location of the upstream breakpoint, a C57BL/6 BAC library array was screened with an exon 1 probe (raised with primers #1 and #2), and four BAC clones were obtained. All four clones were positive for exon 1 by PCR with primers #1 and #2. One of the four clones, BAC-1, contained exon 1 and exon 2, as well as exon 18 (that is, the entire gene) as ascertained by PCR and Southern blot analysis. By restriction mapping and Southern blot analysis, all of intron 1 was identified and subcloned (Fig. 3A). Its length was determined to be 28.8 kb, consistent with that reported for the rat aggrecan gene (Doege et al.

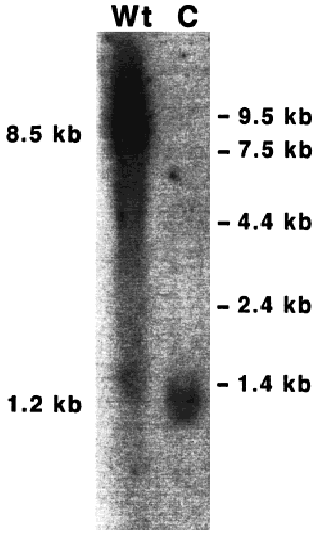


Fig. 2. Northern blot analysis of cartilage aggrecan expression. Autoradiograph of a Northern blot comparing total cartilage RNA (15 μg) derived from either wild type (Wt) or *cmd-Bc* (C) embryonic mice. P^{32}-labeled probe was derived from exon 1. The RNA size ladder is noted in the right margin. The size of the cartilage RNA bands as estimated by the ladder (r = 0.999) are noted in the left margin.

1994), but significantly longer than that reported for the human aggrecan gene (Valhmu et al. 1995) or chick aggrecan gene (Li and Schwartz 1995).

A 10.8-kb subclone obtained by *Nhe*I digestion of BAC-1, named *Nhe*I-10.8, was shown to span a portion of intron 1 from 16 to 26.8 kb 3′ exon 1 (Fig. 3). The 5′ breakpoint for the *cmd-Bc* deletion is contained in this region of the intron, as determined with probe A, amplified with primers 1230 bp (sense, #3) and 1523 bp (antisense, #4) 3′ of *Nhe*I site N1 (Fig. 3); this approach detected an RFLP between *cmd-Bc* and wild-type DNA, digested by either *Nhe*I or *Kpn*I (Fig. 3B). Heterozygous animals contained both the larger normal and smaller mutant restriction fragments.

Inverse PCR was used to amplify the region spanning the deletion in the *cmd-Bc* mutant 4.8-kb *Nhe*I DNA fragment. After genomic DNA from the *cmd-Bc* mutant was digested with *Nhe*I, diluted and ligated, PCR was performed with primers 39 bp (antisense, #5) and 990 bp (sense, #6) 3′ of the upstream *Nhe*I cutting site (N1, Fig. 3A). Unlike standard PCR, these primers are oriented so that primer extension proceeds divergently on a circular template, allowing amplification of flanking sequences without need of distal primers. As predicted by RFLP analysis, a 3.8-kb band, named CMD-*Nhe*I 3.8 (Fig. 3A), was amplified, cloned, and sequenced. Nucleotides 1–1472 were >98% identical to nucleotides 990–2470 obtained from the *Nhe*-10.8 clone (Fig. 4A), as would be predicted from the choice of the primers used in the inverse PCR (Fig. 3A). After that 3′ point the sequences diverged (Fig. 4A), suggesting that the location of the upstream breakpoint BP1 was 2470 bp 3′ of the N1 cutting site, that is, in intron 1 approximately 18.5 kb 3′ of exon 1 (Fig. 3A). In addition, a *Kpn*I cutting site lies 241 bp 3′ of the *cmd-Bc* breakpoint (K3), defining the 3′ end of the 4.7-kb *Kpn*I fragment identified by RFLP analysis (Fig. 3B, probe A).

A second probe (probe B) was amplified with primers 391 bp (sense, #8) and 23 bp (antisense, #7) upstream of the 3′ end of CMD-*Nhe*I 3.8 (N4, Fig. 3A). Probe B could be amplified by PCR with *cmd-Bc* mutant DNA, normal DNA, or BAC-1 DNA as template. This suggested that the 3′ end of the BAC-1 clone insert was downstream of the 3′ breakpoint , BP2. With this probe, another RFLP could be detected distinguishing *cmd-Bc* and wild-type DNA digested by *Nhe*I (Fig. 3B). The 4.8-kb fragment from *cmd-Bc* DNA is the same one seen with probe A. Heterozygous animals again exhibited both mutant and normal restriction fragments. No RFLP was detected with *Kpn*I, consistent with the observed *Kpn*I cutting site 240 bp 3′ of BP2, but 5′ of probe B.

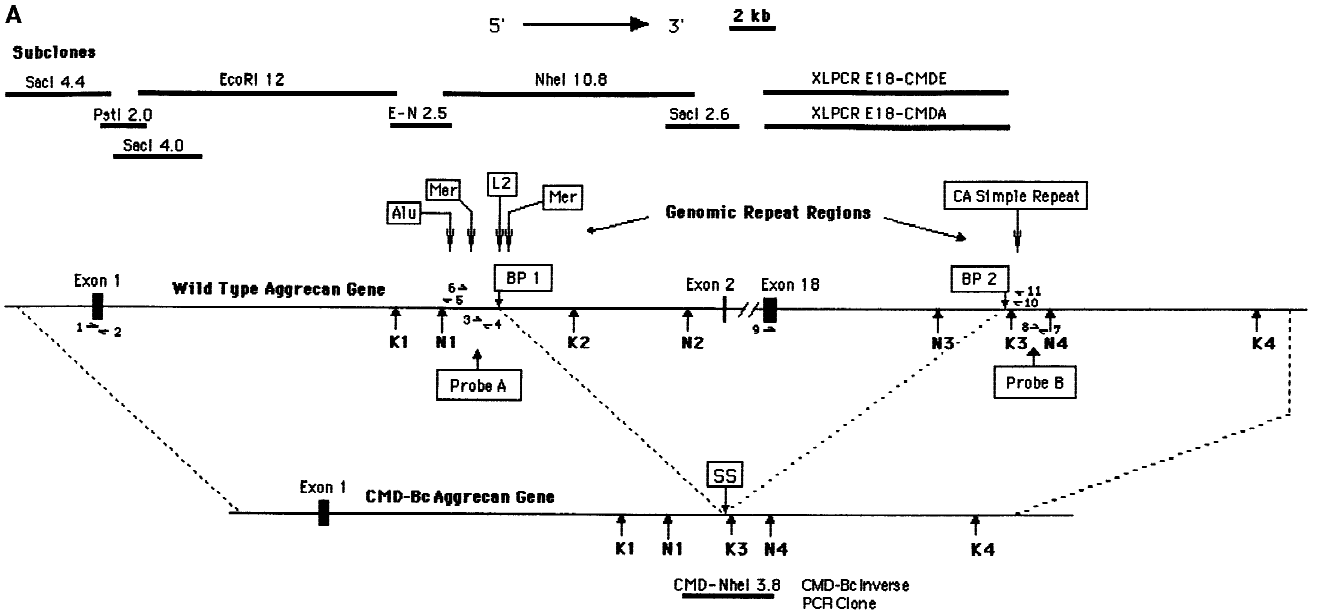
With BAC-1 DNA as a template, a sense primer from exon 18 (#9) and either of two antisense primers 365 bp or 654 bp 3′ of BP2 (#10 and #11, respectively; Table 1, Fig. 3A), a 10.5-kb DNA fragment could be amplified with rTth DNA XL polymerase; thus, BP2 is about 10.5 kb 3′ of the aggrecan gene. This fragment was gel purified and sequenced directly. As predicted, the 5′ sequence of this fragment was identical to the published sequence for the 3′ end of the aggrecan gene. The 3′ end of the fragment was 98% identical to the sequence from the *cmd-Bc* aggrecan gene up to the predicted breakpoint, after which point the sequences diverge (Fig. 4A). As was true for exons 2–18, primers immediately upstream of the 3′ breakpoint in the aggrecan gene, BP2, could PCR-amplify products from homozygous normal genomic DNA and BAC-1 DNA templates, but not from *cmd-Bc* mutant DNA templates.

Sequences homologous to the consensus cleavage site sequences for topoisomerase I (Jaxel et al. 1988; Shen and Shen 1990) and topoisomerase II (Spitzner and Muller 1988; Spitzner et al. 1990; Fig. 4A) flank both the 3′ and 5′ breakpoints. There was also a 7-bp direct repeat (5′-TGACAGA) flanking the splice junction; such repeats have been implicated in a deletion in the β-globin family of genes (Efstratiadis et al. 1980; Henthorn et al. 1990). Inverted repeats, which could possibly form intrastrand hairpins with ≥5 Watson-Crick basepairs, were detected at the splice junction (Fig. 4B), consistent with the model of Glickman and Ripley (1984). In addition, the region around BP1 contains one Alu-like repeat motif, two Mer-like motifs, and an L2-like motif. In the wild-type DNA, there is a simple CA repeat motif shortly downstream of BP2 (Fig. 3).

Using the BAC-1 DNA as a template and primers from exons 12, 13, and 14, the intron/exon structure around exon 13 (114 bp) was determined by PCR and sequence analysis. Introns 12 and 13 are 4.7 kb and 2.8 kb in length, respectively. The acceptor/donor sequences at the intron boundaries are presented in Table 2. These data complete the structural analysis for the murine aggrecan gene initiated by Watanabe and colleagues (1995). The cDNA sequence is 6659 bp long, including 114 bp for exon 13, encoding a protein core of 2170 amino acids [M(r) 263,231]. With the lengths of intron 1 (28.8 kb), intron 12 (4.7 kb), and intron 13 (2.8 kb) now known, the length of the entire murine aggrecan gene is calculated to be 68.8 kb, not unlike the size in other mammalian species (Schwartz et al. 1999), but considerably longer than the avian aggrecan gene (Li and Schwartz 1995).

Discussion

Unlike the two previous reports of naturally occurring null mutations in the aggregan gene, nanomelia in the chick (Li et al. 1993) and *cmd* in mice (Watanabe et al. 1994), this is the first report of an aggrecan defect caused by a nearly complete gene deletion. Deletions are a common type of mutation underlying many abnormal phenotypes. The characterization of a deletion consisting of the juxtaposition of noncontiguous sequences resulting from the breaking and rejoining of the DNA strands provides an excellent opportunity to study the mechanisms that are involved in the generation of mutations in the mammalian genome (Ueki et al. 1998). It has been postulated that topoisomerases I and II may be involved in recombinational events, since these enzymes form covalent linkages in both strands of the DNA helix and can cause transient single- and double-strand breaks, respectively. The presence of nucleotide sequences similar to the consensus cleavage site sequences for topoisomerase I (Jaxel et al. 1988; Shen and Shen 1990) and topoisomerase II (Spitzner and Muller 1988; Spitzner et al. 1990) and repetitive elements all located at or near the 3′ and 5′ breakpoints (Fig. 4A) lead us to speculate that this gene deletion may be the result of an erroneous recombination event. Topoisomerase II has been implicated in multiple deletions in the human T-lymphocyte HPRT gene (Rainville et al. 1995), and in translocation breakpoints associated with secondary leukemias, especially following treatment with topoisomerase inhibitors (Aplan et al.



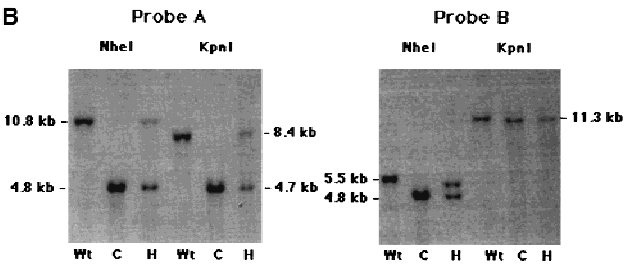


Fig. 3. Organization of intron 1 and 3′ end of the aggrecan gene and its relation to the *cmd-Bc* defect. **(A)** Map of the aggrecan gene from Wt and *cmd-Bc* mice. The location of the subclones and primers are noted (see Table 1). The region between the 5′ breakpoint (BP 1) and the 3′ breakpoint (BP 2) is deleted from the *cmd-Bc* mutant gene, and the regions outside of the breakpoints are spliced together (SS). N1–4 and K1–4 refer to *Nhe*I and *Kpn*I restriction sites, respectively. Probe A (synthesized using primers #3 and #4) and Probe B (synthesized using primers #7 and #8) were hybridized to Southern blots shown in Fig. 3B. Intron 1 is calculated to be 28.8 kb in length. BP1 is 2476 bp 3′ of *Nhe*I restriction site N1 or 18.8 kb 3′ of exon one. BP2 is 245 bp 5′ of *Kpn*I restriction site K3 or 10.5 kb 3′ of exon 18. In the region around BP1 there is one Alu-like repeat motif, two Mer-like motifs, and an L2-like motif. There is a simple CA repeat motif downstream of BP2. **(B)** Autoradio-gram of Southern blot restriction length polymorphism analysis of the *cmd-Bc* defect. The left and right panels are autoradiograms of identical blots hybridized with ^{32}P-labeled probes A and B, respectively. Wt, C, and H lanes have genomic DNA from wild type, *cmd-Bc*, and heteroxygous mice. DNAs were digested with the *Nhe*1 or *Kpn*1 as indicated. The size of the bands, as estimated by the DNA ladder ($r = 0.994$), are noted in the margins.

1996; Broeker et al. 1996; Kulkarni et al. 1999). A model for topoisomerase I-mediated deletions has also been proposed (Henningfeld and Hecht 1995). Deletions in the type IV collagen genes COL4A5 and COL4A6, defects which are associated with Alport's syndrome, also have topoisomerase I and II binding sites at the breakpoints (Ueki et al. 1998).

Both hairpin or branch helix formations (Fig. 4B) and short direct repeats (Fig. 4A) have been reported to be involved in deletions in eukaryotic genes by producing slipped mispairing during DNA replication (Efstratiadis et al. 1980; Glickman and Ripley 1984). In the β-globin gene causing hereditary persistence of fetal hemoglobin, HPFH-2, a short direct repeat was observed at the junctions of a >100 kb deletion (Henthorn et al. 1990). Not commented on in that paper were the presence of topoisomerase II binding sites (on the complementary strand) at both the 3′ and 5′ breakpoints in that gene. The aforementioned COL4A5 and COL4A6 genes also had short direct repeats present at the breakpoints (Ueki et al. 1998).

Nicked or gapped breakpoints caused by topoisomerase I and II could contribute to a single-strand loop excision during slipped mispairing. Such a model has been proposed by Baguley and Ferguson (1998). Although the frequency with which topoisomerase II consensus sequences occur in the murine genome is unclear, it is clear that these sites do not occur randomly in the genome. These sequences have been noted to be concentrated at matrix attachment sites, that is, at the bases of supercoiled looped DNA domains of approximately 50–100 kb (Cockerill and Garrard 1986; Blasquez et al. 1989; Sperry et al. 1989; Broeker et al. 1996); these have been clearly implicated as "hot spots" for illegitimate recombination events. The proximity of the breakpoint reported in this paper to the topoisomerase consensus sequence is consistent with previous reports (Sperry et al. 1989). Finally, it has been suggested that Alu elements may play a role in illegitimate recombination between the BCR and the ABL genes in chronic myeloid leukemia (Chen et al. 1989; Kulkarni et al. 1999). However, the high frequency of such sequences within the genome makes it unlikely that this is the sole mechanism of the deletion occuring in the *cmd-Bc* mouse.

Concentration of these consensus sequences within particular genes may indeed lead to the widely different frequencies of gene deletion mutations that have been observed for different genes. For example, relative frequencies of disease-related mutations secondary to deletions in the steroid sulfatase gene (10 exons, 146 kb), the LDL receptor (18 exons, 60 kb), and the dystrophin gene (>60 exons, 2.3 Mb) are 90%, 2–6%, and 70% respectively (Yen et al. 1990; Rainville et al. 1995). In addition, deletion sizes vary greatly and breakpoint clustering can occur. Although numerous skeletal disorders in humans due to defects in ECM molecules have been reported, such as osteogenesis imperfecta (Prockop et al. 1994) and chondrodysplasia (Warman et al. 1993), no human disease has yet been identified as being due to a mutation in the aggrecan gene (Watanabe et al. 1997). The fact that there have been two reported gene deletions in the murine aggrecan gene and no reports thus far of a similar deletion in human leads us to speculate that these differences may be related to differences in frequency of "hot spots" for illegitimate recombination between the human and mu-

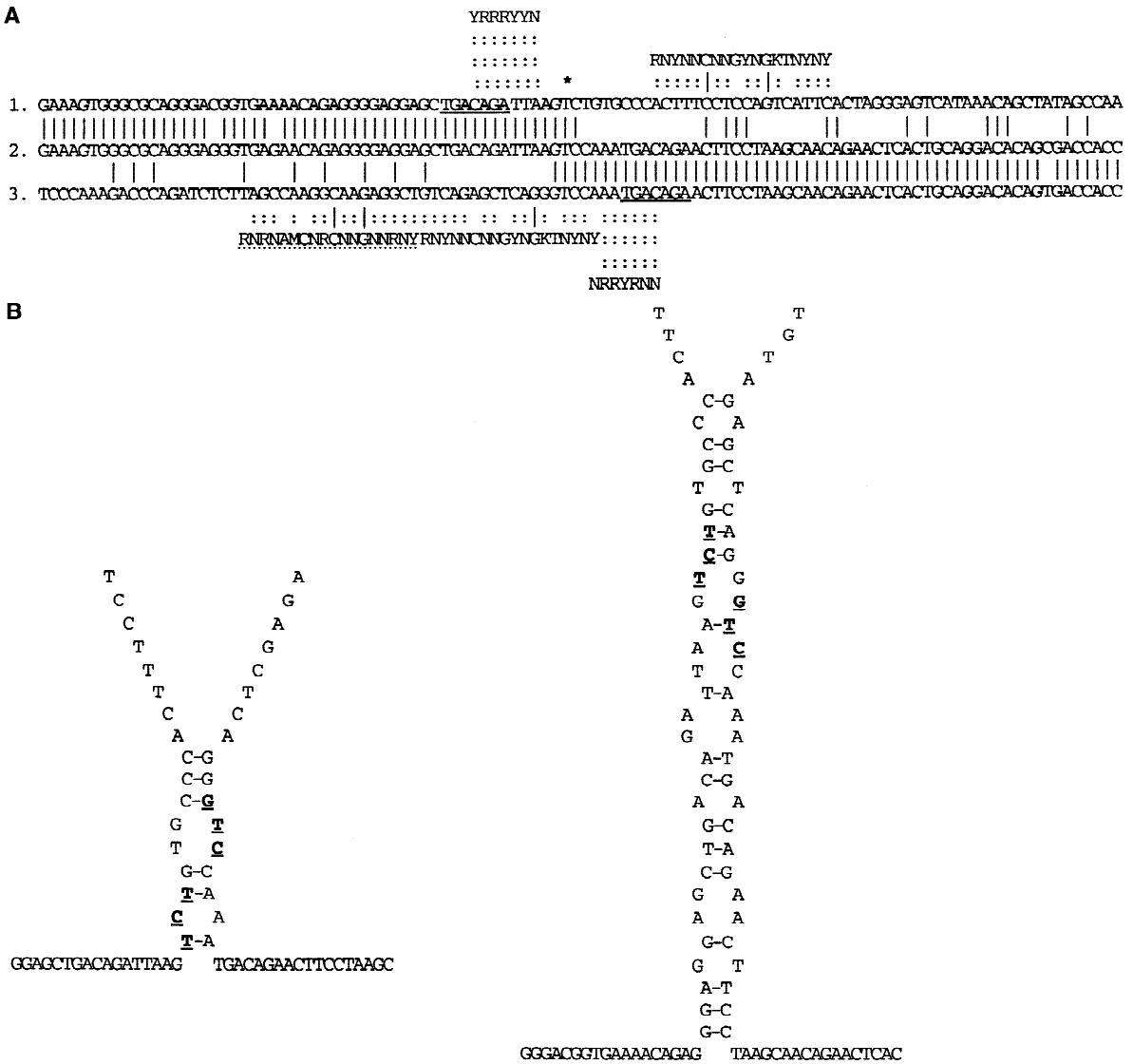


Fig. 4. (A) Overlap of wild-type sequences around the 5′ (BP1) and 3′ (BP2) breakpoints and the *cmd-Bc* deletion-bridging sequence. Sequence 1 is wild-type sequence around the 5′ breakpoint (BP1, Fig. 3A) obtained from BAC-1 subclone *Nhe*I 10.8. Sequence 2 is *cmd-Bc* sequence bridging the splice site (SS, Fig. 3A) obtained from inverse PCR subclone CMD-*Nhe*I 3.8. Sequence 3 is wild-type sequence containing the 3′ breakpoint (BP2, Fig. 3A) obtained from XL PCR products E18-CMDA and E18-CMDE. Splice site location is indicated by the *. Homologies to consensus cleavage sequences for topoisomerase I (5′-YRRRYYN or 5′-NRRYRNN) and topoisomerase II (5′-RNYNNCNNGYNGKTNYNY) are also noted. A topoisomerase II sequence was detected on the complementary strand of sequence 3 (dashed underline). Seven base-pair direct repeats (5′-TGACAGA) are noted by underlining. Sequences 1, 2, and 3 have been deposited in GenBank™/EMBL Data Bank. **(B)** Two possible hairpins formed by inverted repeats involving ≥5 Watson-Crick basepairs near the breakpoint. The possible breakpoint sites are indicated by underlined bold type.

Table 2. Exon and intron boundary sequences for murine aggrecan exons 12–14

Exon	Domain	Exon size	Acceptor sequence	Exon sequence	Donor sequence	Intron size (kb)
12	CS	3482 bp	ttcctccag	GGA CAG	gtatggagt	4.7
13	G3-EGF	114 bp	ttcttgcag	ACA TCG	gtacggccg	2.8
14	G3-Lectin	158 bp	gtgttacag	ACC ACA	gtgagtgtg	0.55

rine gene. However, it should be noted that a variable number of tandem repeat (VNTR) polymorphisms has been discovered recently in a region of the human aggrecan gene encoding chondroitin sulfate attachment sites, and these repeats have been associated with bilateral hand osteoarthritis (Doege et al. 1997; Horton et al. 1998).

Although the phenotypes of the cmd and cmd-Bc mutants have been described and are similar, they have not been directly compared. In particular, aberrant expression of other ECM molecules reported in the cmd mutant (Wai et al. 1998) or the dwarfism and spinal degeneration reported in the cmd heterozygote (Watanabe et al. 1997) have not been studied in the cmd-Bc mutant. It should now be possible to determine whether these phenotypes result only from the absence of mature aggrecan, or also from the accumulation of immature aggrecan precursor, which would occur in the cmd mutant. Definition of the defect will advance the *cmd-Bc*

mouse as a model for the study of aggrecan function in cartilage development and regulation of matrix assembly, and will promote related studies seeking a human homolog of this genetic disease.

Acknowledgments. We thank Judy Henry and Mary Lou Spach for technical help in animal husbandry and tissue preparation and Dr. Miriam Domowicz for preparation of the *cmd-Bc* and WT mice in Fig. 1. We also thank Edward W. Pirok III and James R. Mensch Jr. for advice during the course of this study. This paper is dedicated to the memory of Hao Li. This research was supported by grant AR-19622 and HD-09402 (N.B. Schwartz).

References

Aplan PD, Chervinsky DS, Stanulla M, Burhans WC (1996) Site-specific DNA cleavage within the MLL breakpoint cluster region induced by topoisomerase II inhibitors. Blood 87, 2649–2658

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG et al. (1997) *Short Protocols in Molecular Biology,* 3rd ed. (New York: John Wiley and Sons, Inc.)

Baguley, BC, Ferguson LR (1998) Mutagenic properties of topoisomerase-targeted drugs. Biochim Biophys Acta 1400, 213–222

Baldwin CT, Reginato AM, Prockop DJ (1989) A new epidermal growth factor-like domain in the human core protein for the large cartilage-specific proteoglycan. Evidence for alternative splicing of the domain. J Biol Chem 264, 15747–15750

Bell L, Juriloff M, Harris MJ (1986) A new mutation at the cmd locus in the mouse. J Hered 77, 205–206

Blasquez VC, Sperrry AO, Cockerill PN, Garrard WT (1989) Protein: DNA interactions at chromosomal loop attachment sites. Genome 31, 503–509

Brennan MJ, Oldberg A, Ruoslahti E, Brown K, Schwartz NB (1983) Immunological evidence for two distinct chondroitin sulfate proteoglycan core proteins: differential expression in cartilage matrix deficient mice. Dev Biol 98, 139–147

Broeker PL, Super HG, Thiramn MJ, Pomykala H, Yónebayashi Y et al. (1996) Distribution of 11q23 breakpoints within the MLL breakpoint cluster region in de novo acute leukemia and in treatment-related acute myeloid leukemia: correlation with scaffold attachment regions and topoisomerase II consensus binding sites. Blood 87, 1912–1922

Brown KS, Harne L (1982) Brachymorphic (bmm/bmm) cartilage matrix deficiency (cmd/cmd) and disproportionate micromelia (Dmm/Dmm); three inborn errors of cartilage biosynthesis in mice. Prog Clin Biol Res 94, 245–249

Brown KS, Yamada Y, Abramczuk J, Kimata K (1991) New genetic approaches to craniofacial growth and malformation in the mouse. J Craniofacial Genet Dev Biol 11, 357–365

Chen SJ, Chen Z, d'Auriol L, Le Coniat M, Grausz D et al. (1989) Ph1+bcr-acute leukemias: implication of Alu sequences in a chromosomal translocation occurring in the new cluster region within the BCR gene. Oncogene 4, 195–202

Cheng S, Fockler C, Barnes WM, Higuchi R (1994) Effective amplification of long targets from cloned inserts and human genomic DNA. Proc Nat Acad Sci USA 91, 5695–5699

Cockerill PN, Garrard WT (1986) Chromosomal loop anchorage of the Kappa immunoglobulin gene next to the enhancer in a region containing topoisomerase II sites. Cell 44, 273–282

Doege KI, Garrison K, Coulter SN, Yamada Y (1994) The structure of the rat aggrecan gene and preliminary characterization of its promoter. J Biol Chem 269, 29232–29240

Doege KJ, Coulter SN, Meek LM. Maslen K, Wood JG (1997) A human-specific polymorphism in the coding region of the aggrecan gene. Variable number of tandem repeats produce a range of core protein sizes in the general population. J Biol Chem 272, 13974–13979

Efstratiadis A, Posakony JW, Maniatis T, Lawn RM, O'Connell C et al. (1980) The structure and evolution of the human beta-globin gene family. Cell 21, 653–668

Fulop C, Cs-Szabo G, Glant TT (1996) Species-specific alternative splicing of the epidermal growth factor-like domain 1 of cartilage aggrecan. Biochem J 319, 935–940

Glickman BW, Ripley LS (1984) Structural intermediates of deletion mutagenesis: a role for palindromic DNA. Proc Natl Acad Sci USA 81, 512–516

Henningfeld KA, Hecht SM (1995) A model for topoisomerase I-mediated insertions and deletions with duplex DNA substrates containing branches, nicks, and gaps. Biochemistry 34, 6120–6129

Henthorn PS, Smithies O, Mager DL (1990) Molecular analysis of deletions in the human beta-globin gene cluster: deletion junctions and locations of breakpoints. Genomics 6, 226–237

Horton WE, Jr, Lethbridge-Cejku M, Hochberg MC, Balakir R, Precht P et al. (1998) An association between an aggrecan polymorphic allele and bilateral hand osteoarthritis in elderly white men: data from the Baltimore Longitudinal Study of Aging (BLSA). Osteoarthritis Cartilage 6, 245–251

Houghton MJ, Carey JC, Seegmiller RE (1989) Pulmonary hypoplasia in mice homozygous for the cartilage matrix deficiency (cmd) gene: a model for human congenital disorder. Pediatr Pathol 9, 501–512

Jaxel C, Kohn KW, Pommier Y (1988) Topoisomerase I interaction with SV40 DNA in the presence and absence of camptothecin. Nucleic Acids Res 16, 11157–11170

Kimata K (1982) [Cartilage matrix deficiency: a new genetic disorder in the mouse and related subjects (author's transl)]. Seikagaku—J Jpn Biochem Soc 54, 162–168

Kimata K, Barrach HJ, Brown KS, Pennypacker JP (1981) Absence of proteoglycan core protein in cartilage from the cmd/cmd (cartilage matrix deficiency) mouse. J Biol Chem 256, 6961–6968

Kulkarni S, Reiter A, Smedley D, Goldman JM, Cross NCP (1999) The genomic structure of ZNF198 and location of breakpoints in the t(8;13) myeloproliferative syndrome. Genomics 55, 118–121

Li H, Schwartz NB (1995) Gene structure of chick chondroitin sulfate proteoglycan (aggrecan) core protein. J Mol Evol 41, 878–885

Li H, Schwartz NB, Vertel BM (1993) cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core protein and identification of a stop codon in the aggrecan gene associated with the chondrodystrophy, *nanomelia.* J Biol Chem 268, 23504–23511

Li Y, Olsen BR (1997) Murine models of human genetic skeletal disorders. Matrix Biol 16, 49–52

Ochman H, Medhora MM, Garza D, Hartl DL (1990) Amplification of flanking sequences by inverse PCR. In *PCR Protocols: A Guide to Methods and Applications,* M Innis, DH Gelfand, JJ Sninsky, T White, eds. (New York: Academic Press, Inc.) pp 219–227

Prockop DJ, Kuivaniemi H, Tromp G (1994) Molecular basis of osteogenesis imperfecta and related disorders of bone. Clin Plastic Surg 21, 407–13

Rainville IR, Albertini RJ, Nicklas JA (1995) Breakpoints and junctional regions of intragenic deletions in the HPRT gene in human T-Cells. Somatic Cell Mol Genet 21, 309–326

Rittenhouse E, Dunn LC, Cookingham J, Calo C, Spiegelman M et al. (1978) Cartilage matrix deficiency (cmd): a new autosomal recessive lethal mutation in the mouse. J Embryol Exp Morphol 43, 71–84

Schwartz NB, Domowicz M (1998) Proteoglycan gene mutations and impaired skeletal development. In Skel Morphol Growth (Chicago: AAOS Publications) pp 413–435

Schwartz NB, Pirok EWR, Mensch JR Jr, Domowicz MS (1999) Domain organization, genomic structure, evolution, and regulation of expression of the aggrecan gene family. Prog Nucleic Acid Res Mol Biol 62, 177–225

Shen CC, Shen CK (1990) Specificity and flexibility of the recognition of DNA helical structure by eukaryotic topoisomerase I. J Mol Biol 212, 67–78

Sperry AO, Blasquez VC, Garrard WT (1989) Dysfunction of chromosomal loop attachment sites: illegitimate recombination linked to matrix association regions and topoisomerase II. Proc Natl Acad Sci USA 86, 5497–5501

Spitzner JR, Muller MT (1988) A consensus sequence for cleavage by vertebrate DNA topoisomerase II. Nucleic Acids Res 16, 5533–5556

Spitzner JR, Chung IK, Muller MT (1990) Eurkaryotic topoisomerase II preferentially cleaves alternating purine-pyrimidine repeats. Nucleic Acids Res 18, 1–11

Ueki Y, Naito I, Oohashi T, Sugimoto M, Seki T et al. (1998) Topoisomerase I and II consensus sequences in a 17-kb deletion junction of the COL4A5 and COL4A6 genes and immunohistochemical analysis of esophageal leiomyomatosis associated with Alport syndrome. Am J Hum Genet 62, 253–261

Valhmu WB, Palmer GD, Rivers PA, Ebara S, Cheng JF et al. (1995) Structure of the human aggrecan gene: exon-intron organization and association with the protein domains. Biochem J 309, 535–542

Wai AW, Ng LJ, Watanabe H, Yamada Y, Tam PP et al. (1998) Disrupted expression of matrix genes in the growth plate of the mouse cartilage matrix deficiency (cmd) mutant. Dev Genet 22, 349–358

Walcz E, Deak F, Erhardt P, Coulter SN, Fulop C et al. (1994) Complete coding sequence, deduced primary structure, chromosomal localization, and structural analysis of murine aggrecan. Genomics 22, 364–371

Warman ML, Abbott M, Apte SS, Hefferon T, McIntosh I et al. (1993) A type X collagen mutation causes Schmid metaphyseal chondrodysplasia. Nat Genet 5, 79–82

Watanabe H, Kimata K, Line S, Strong D, Gao L et al. (1994) Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the aggrecan gene. Nat Genet 7, 154–158

Watanabe H, Gao L, Sugiyama S, Doege K, Kimata K et al. (1995) Mouse aggrecan, a large cartilage proteoglycan: protein sequence, gene structure and promoter sequence. Biochem J 308, 433–440

Watanabe H, Nakata K, Kimata K, Nakanishi I, Yamada Y (1997) Dwarfism and age-associated spinal degeneration of heterozygote cmd mice defective in aggrecan. Proc Natl Acad Sci USA 94, 6943–6947

Yen PH, Li XM, Tsai SP, Johnson C, Mohandas T et al. (1990) Frequent deletions of the human X chromosome distal short arm result from recombination between low copy repetitive elements. Cell 61, 603–610