Isolation and characterization of the genomic human CD7 gene: Structural similarity with the murine Thy-1 gene

(DNase I hypersensitive site/promoter/TATA-less/immunoglobulin gene superfamily)

LAURA E. SCHANBERG*, DONALD E. FLEENOR†, JOANNE KURTZBerg*, BARTON F. HAYNES‡, and RUSSEL E. KAUFMAN††‡‡

Departments of *Pediatrics, †Medicine, and ‡Biochemistry, Duke University Medical Center, Durham, NC 27710

Communicated by D. Bernard Amos, October 17, 1990

ABSTRACT The human CD7 molecule is a 40-kDa member of the immunoglobulin gene superfamily that is expressed on T-lymphoid and myeloid precursors in fetal liver and bone marrow. CD7 is also expressed on T lymphocytes in multiple stages of T-cell development, including a major subset of mature peripheral T cells. In this paper we report the isolation and characterization of the human CD7 gene and 5' flanking region. Sequence analysis revealed that the CD7 gene comprises four exons that span 3.5 kilobases. The 5' flanking region (506 base pairs) has a high G+C content and no "TATA" or "CCAAT" elements. DNase I sensitivity analysis of chromatin from the CD7* progenitor cell leukemia line, DU528, and the CD7*, CD4*, CD8*, TCRαβ* T-cell line, DU980 (where TCR is the T-cell receptor), revealed two distinct hypersensitive sites 5' of the CD7 gene. Hypersensitive site 1, present in the DU980 T-cell clone, was located 4.5 kilobases upstream of the presumed CD7 transcription initiation site. Only DNase I hypersensitive site 2, which mapped to the promoter region, was found in the DU528 line. Comparison of the organization of the CD7 gene with that of other members of the immunoglobulin gene superfamily revealed that the human CD7 gene most closely resembles the murine Thy-1 gene. Both CD7 and Thy-1 are encoded by small genes with four exons, contain TATA-less promoters, and have a similar functional organization. These structural similarities suggest that human CD7 and murine Thy-1 may be functional homologues.

Although the function of the CD7 molecule is not known, prior data have shown that CD7 is up-regulated on activated T cells and down-regulated by phorbol esters on peripheral blood T lymphocytes (for review, see ref. 6 and 7). CD7 monoclonal antibody partially blocks the allogeneic mixed-lymphocyte reaction, acts as a comitogen for T-cell proliferation, and induces a transmembrane calcium flux after antibody crosslinking (for review, see refs. 6 and 8). In this paper, we have described the cloning and characterization of the human CD7 gene. The coding information for CD7 is partitioned into four exons that span <3.5 kb. The 5' flanking region does not have a "TATA" or a "CCAAT" box and has characteristics similar to other regulated genes with TATA-less promoters (9–13). By using DNase I hypersensitivity analysis, we have identified two tissue-specific hypersensitive sites 5' of the coding region of the CD7 gene. Comparison of the structure of the CD7 gene with the structures of other members of the immunoglobulin gene superfamily revealed striking similarities of the CD7 gene with the murine and human Thy-1 genes.

MATERIALS AND METHODS

Isolation of Genomic Clone. A lung fibroblast human genomic DNA library constructed in AFIIX was screened with the CD7 cDNA clone, kindly provided by A. Aruffo and B. Seed (2). Two positively hybridizing A clones were isolated and characterized by restriction mapping and Southern blot analysis.

DNA Sequencing. Dideoxynucleotide sequencing was performed using the Sequenase system (United States Biochemical) on both double-stranded and single-stranded sequencing templates. The sequence was confirmed by sequencing complementary strands.

Cell Culture and Phenotyping. CD7+, CD4+, CD8+, sCD3+ DU528 cells (where hi is high expressor and s is surface) were cultured in standard medium as described (5, 14). The CD7+, CD4+, CD8+ DU980 cell line was isolated from a patient with a progenitor cell leukemia syndrome as described (14) and cultured in standard medium supplemented with human recombinant interleukin 2 (8 µg/ml) and T-cell conditioned medium. Cells were characterized by indirect immunofluorescence and analyzed using a Profile I flow cytometric analyzer (Coulter).

DNase I Analysis. DNase I analysis was performed using the method described by Enver et al. (15) with modifications. Approximately 1.5 × 10⁶ cells were pelleted, washed once with isotonic phosphate-buffered saline, resuspended in 12 ml of ice-cold lysis buffer [50% (vol/vol) glycerol/50 mM Tris·HCl, pH 7.9/100 mM KCl/5 mM MgCl₂/0.05% Triton X-100] and lysed with 50 µl of 0.1% (wt/vol) SDS and 10 µl of 1 M CaCl₂. The lysates were then extracted with phenol/chloroform and precipitated with 0.1 vol of ethanol. The RNA pellets were washed with 70% ethanol and resuspended in 10 µl of sterile water for gel electrophoresis

Abbreviation: TCR, T-cell receptor.

†To whom reprint requests should be addressed.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. M37271).
X-100/100 mM dithiothreitol], and incubated on ice for 10 min. After repelling, the nuclear extract was resuspended in 1 ml of ice-cold buffer A (100 mM NaCl/50 mM Tris-HCl, pH 8.0/3 mM MgCl2/0.1 mM phenylmethylsulfonyl fluoride) and centrifuged at 2500 rpm in a HB4 rotor for 5 min at 0°C. After resuspension in 500 μl of buffer A, the nuclei were divided into several samples and exposed to various concentrations of DNase I (Worthington) for 30 min at 37°C. DNA was extracted, digested with EcoRI or Bgl II/HindIII, and analyzed on a Southern blot by using the full-length CD7 cDNA labeled with [32P]dCTP as probe. Hybridization was performed overnight at 42°C, in 50% (vol/vol) formamide/3× SSC/0.2% sodium pyrophosphate/1 mM EDTA, pH 8.0/10 mM Heps, pH 7.5/0.1% SDS containing boiled salmon sperm DNA (100 μg/ml) (1× SSC = 0.15 M NaCl/0.015 M sodium citrate, pH 7.0; 1× Denhardt’s solution = 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin).

RESULTS

Cloning of the CD7 Gene. Southern blot analysis of genomic DNA with the CD7 cDNA showed that the entire CD7 gene coding region was located within a single 20-kb HindIII fragment. A genomic library, prepared from a Sau3A partial digest of human DNA inserted into A[FI], was screened using the randomly labeled CD7 cDNA. One recombinant, designated ALSCLD7-1, contained a 15-kb insert and was selected for characterization based on restriction mapping studies (Fig. 1A).

Structure and Sequence of the CD7 Gene. By Southern blot analysis of ALSCLD7-1 with the CD7 cDNA probe, we determined that all coding regions in the 15-kb recombinant were localized to a 3-kb DNA segment. This region was mapped and sequenced, revealing a gene comprised of four exons and three introns (Fig. 1B). The 3280-base-pair (bp) sequence included the entire coding region of the CD7 gene and 506 bp of the 5′ flanking area (Figs. 2 and 3A). The exon–intron junctions were deduced by comparison of genomic sequences with the CD7 cDNA sequence. Exons 1 and 2 were interrupted by a 431-bp intron. We compared the sequence of the genomic intron 1 with the published sequence and identified an extra cytidine at position 569 of the 1.7-kb clone that bears

![Fig. 1](https://example.com/fig1.png)  
(A) Restriction map showing ALSCLD7 clones. B. Bgl II; E. EcoRI; H. HindIII S. Sst I; X. Xho I. (B) Restriction map of CD7 gene showing exon structure. Arrows represent sequencing primers.

![Fig. 2](https://example.com/fig2.png)  
DNA sequence of the CD7 gene. The amino acid sequence is shown above the DNA sequence. The cDNA is in capital letters. The transmembrane domain is underlined. Position 1 is the first amino acid of the mature CD7 protein.
Fig. 3. (A) DNA sequence of the 5' flanking region of the CD7 gene. The 5' end of the cDNA is capitalized. Regulatory element motifs are underlined. Regions of similarity with the murine Thy-1 promoter are in boldface type. (B) Demonstration of two regions of similarity within the CD7 gene promoter and the murine Thy-1 gene promoter. The numbering of the Thy-1 gene uses the transcription start site upstream of exon 1a as the reference point.

By inspection, several regulatory element motifs were identified in the CD7 gene. Three potential CACCC boxes (105, 95, and 74 bp upstream of the putative cap site) were identified. Consensus transcription factor NF-κB sequence motifs were found on the coding strand at nucleotide 246 (GGGGAGTCCTTT) and nucleotide 331 (GGGAACGGCT) (Fig. 3A). Each of these sequence motifs differed from a proposed consensus sequence for NF-κB, GGGANNYYCC (17), by only one pyrimidine in position 10 of the motif. In addition, there was a Sp1 motif (17) with the sequence CCGGCC on the coding strand at nucleotide 275 (Fig. 3A) and an 8-10-nucleotide match with a proposed consensus sequence for the conserved octamer element ATGCCAAAT (17) at nucleotide 69 (Fig. 3A). One possible transcription factor AP-1 and several potential transcription factor AP-2 binding sequences were identified (17), as shown in Fig. 3A.

Identification and Localization of Two DNase I Hypersensitive Sites in CD7+ and CD7- Cell Lines. We examined cell lines with various patterns of CD7 expression for the presence of DNase I hypersensitive sites surrounding the CD7 gene. The cell lines studied included the CD7+ HeLa cells, CD4+, CD8-, CD3- progenitor cell leukemia line DU528 (14), the CD7+, CD4+, CD8+, TCRAβ+ T-cell line DU908 (5), the CD7+ erythroleukemia cell line K562, and the CD7- HeLa cell line. The nonlymphoid CD7- lines HeLa and K562 did not show any evidence of a DNase I hypersensitive site (Fig. 4A). Studies done with the DU908 T-cell line showed that DNase I hypersensitive site 1 was located ≈5 kb upstream from the transcription initiation site (Fig. 4B). Examination of the progenitor cell leukemia line DU528, revealed the presence of a second DNase I hypersensitive site, located within the immediate 5' flanking region (Fig. 4C). Both DNase I hypersensitive sites have been confirmed by mapping with additional restriction enzymes, including EcoRI, HindIII, and Xba I (Fig. 5).

Comparison of the CD7 Gene with Other Members of the Immunoglobulin Gene Superfamily. Comparison of the organization of the CD7 gene with that of other genes in the immunoglobulin gene superfamily (CD2, CD3 (γ, δ, and ε chains), CD4, CD8, immunoglobulin, TCR (α, β, γ, and δ chains), major histocompatibility complex class I and II, Thy-1, neural cell adhesion molecule, peripheral myelin protein zero, and carcinoembryonic antigen) (18) revealed that the organization of the CD7 gene shows several striking similarities to the human and murine Thy-1 genes. Whereas genes for CD2 (19), CD3 (γ, δ, and ε chains) (20, 21), CD4 (22), CD8 (23), neural cell adhesion molecule (24), and carcinoembryonic antigen (25) have TATA-less promoters like CD7 and Thy-1, only CD7 and Thy-1 are small genes that are organized into four exons (26–29). The mouse Thy-1 gene has an alternative first exon that splices into the second exon (9). Both exon 1a and 1b encode the 5' untranslated region in the mouse (9, 29), and exon 2 encodes the signal peptide. Exon 2 of murine Thy-1 corresponds to exon 1 in CD7, which encodes the leader peptide. Exon 3 in murine Thy-1 and exon

Fig. 4. DNase I hypersensitivity experiments. DNA was exposed to increasing concentrations of DNase I and digested with EcoRI. The full-length CD7 cDNA was used as probe.
2 in CD7 code for the mature protein. The fourth and last exon in murine Thy-1 and the third exon in CD7 encode the transmembrane domains of Thy-1 and CD7, respectively. The final exon in CD7 contains the sequence for an intracellular domain that is not present in Thy-1, a phosphatidylinositol-anchored molecule (30).

DISCUSSION

Our studies revealed that the human CD7 gene is contained within 3.0 kb of DNA and is composed of four exons and three introns. The promoter is G+C-rich and lacks TATA and CCAAT sequences, placing this gene in an emerging group of tissue-specific genes with TATA-less promoters. We have identified two DNase I hypersensitive sites upstream of the CD7 gene, each associated with a different lymphohematopoietic cell line. Finally, comparison of the CD7 gene with others in the immunoglobulin gene superfamily demonstrated organizational similarity between the CD7 gene and the Thy-1 gene.

Our analysis of the 5' flanking region of the CD7 gene revealed the absence of a typical TATA element. The promoters of genes that are regulated without the usual TATA or CCAAT sequences have been placed into two classes (31). The first has generally been associated with constitutively active housekeeping genes that have G+C-rich promoters with CpG islands, Sp1 binding sites, and often multiple start points (for review, see ref. 16). The second group includes promoters that also lack a TATA element but are not G+C-rich. Smale and Baltimore (31) have suggested that the second class of promoters is differentially or developmentally regulated; they describe a 17-bp motif, the initiator (Inr), that includes the transcription initiation start site and is sufficient for basal transcription of the lymphocyte-specific terminal deoxynucleotidyltransferase gene. Another transcription initiation element (HPI) has been identified in the G+C-rich promoter of the dihydrofolate reductase gene (32). The CD7 promoter shows identity at the presumed transcription start site with the first 5 nucleotides of the Inr element, including at the -3 and -2 positions, which were shown to be necessary to maintain promoter activity (31). The HPI motif was not found at the transcription initiation site of the CD7 promoter.

The CD7 gene promoter is G+C-rich with multiple CpG dinucleotides, AP-1 and AP-2 binding motifs, and one putative Sp1 binding site. Nevertheless, it does not readily fall into the class of constitutive promoters since CD7, like Thy-1, demonstrates both tissue-specific and developmentally regulated gene expression (10, 33).

We have attempted to identify DNase I hypersensitive sites in the CD7 gene because of their association with transcriptional regulatory elements. In this paper, we have described two strong hypersensitive sites upstream of the CD7 gene that are associated with the gene's expression. Hypersensitive site 1, the most upstream of these sites, is located -4.5 kb on the 5' side of the presumed transcription initiation site and was found in the DU980 T-cell line that is CD7+. The second DNase I hypersensitive site was found near the presumed cap site and is associated with expression of CD7 on the DU528 progenitor cell line. Like murine Thy-1, CD7 may utilize two promoters and have two transcription start sites associated with different stages of T-cell development (9). CD7 mRNA from a variety of phenotypically distinct T-cell lines has been examined and found to be of uniform size (1.2 kb) in all cell lines tested (2). These later data do not exclude the possibility of differential splicing from alternate first exons. Alternatively, the various hypersensitive sites may represent the use of at least two upstream regulatory elements providing different mechanisms for regulation of the CD7 gene. Hypersensitive site 1 may be involved in the repression of CD7, a regulatory strategy that has been described in several genes in the immunoglobulin gene superfamily (34-36). Finally, hypersensitive site 1 may be associated with another transcription unit.

Comparison of the CD7 gene with other genes in the immunoglobulin gene superfamily revealed that CD7 most closely resembles the human and murine Thy-1 gene structure and functional organization. The 5' flanking region of both CD7 and murine Thy-1 (the human Thy-1 promoter has not been described) lacks TATA and CCAAT sequences. Comparison with the promoters of other genes in the immunoglobulin superfamily revealed two groups. The TCR genes (for review, see ref. 37), major histocompatibility complex class I (for review, see ref. 38) and II (39) genes, and the immunoglobulin genes (for review, see ref. 38) all have promoters with TATA sequences for binding of the RNA polymerase II complex. Both CD7 and the murine Thy-1 gene fit into a less well-studied group of genes that includes genes for CD4, CD8, CD2, CD3 (β, ε, and γ chains), neural cell adhesion molecule, carinoembryonic antigen, and peripheral myelin protein zero (40). All of these tissue-specific genes have TATA-less promoters.

Further similarities are noted in the 5' flanking regions of CD7 and murine Thy-1. The murine Thy-1 gene contains two promoters, each upstream of the transcription start site and in one promoter there are two exons. Both of the promoters for the murine Thy-1 gene, like that for CD7, do not contain a TATA or CCAAT box and are G+C-rich. There is a 33-bp region ≈50 bp upstream of the presumed start site in CD7 that shows significant similarity with a region ≈125 bp on the 5' side of the upstream start site in Thy-1 (9). At about position −100, there are duplicated CACCC motifs (17) in both genes. There is also a conserved 17-base region at position −140 in the CD7 gene that shows 90% similarity to a region 60 bases upstream from the second start site in Thy-1 (Fig. 38).

The two promoters of murine Thy-1 show little similarity in DNA sequence but share a conserved nonamer, CTCCTCCTCTGTCATCCCTCTCC in position −47 or −49, which is similar to sequences in the near 5' flanking regions of other promoters without TATA motifs (9). Similar sequences are found in the 5' flanking region of the CD7 gene, GCCCTCCTGTCATCCCTCTCC in position −55 from the presumed start site and CTCTCGTCATCCCTCTCC in position −59. Experiments looking at the regulation of Thy-1 have shown tissue-specific expression and suggested downstream transcriptional control elements for lymphoid expression (33) similar to those found in the immunoglobulin and TCR genes (41-45). Many of the housekeeping genes are also regulated by elements found downstream of the promoter in the first
No mouse homologue to CD7 has been identified. In murine lymphoid tissue, murine Thy-1 has a similar tissue distribution to that of human CD7 in human lymphoid tissue, appearing on prothymocytes, thymocytes, and peripheral T cells (for review, see ref. 46). In contrast, human Thy-1 is expressed primarily on nonlymphoid cell types (47). CD7 in humans and Thy-1 in mice have both been used to define T-cell precursors in extrathymic tissues (3). Structurally, the Thy-1 and CD7 proteins have several similarities, including the location and number of cysteine residues (2). Functional data in vitro have demonstrated similarities in murine Thy-1 and human CD7. For instance, crosslinking either human CD7 or murine Thy-1 triggers a rise in T-lymphocyte intracellular calcium (8, 48). Thy-1 in mice (46) and CD7 in man (3) are expressed on T-cell precursors, function as T-cell activation costimulants, and are single-domain members of the immunoglobulin gene superfamily with similar gene organization and promoter structure. Thus these data suggest that human CD7 and murine Thy-1 may be functional homologues.

We thank Belinda Womack for her technical assistance and Roberta Reeves for help with the preparation of the manuscript. This research was supported in part by National Institutes of Health Grants CA28936 and DK38699. R. E. K. is a recipient of a Biomedical Projects Grant from the Arthritis Foundation. L. E. S. is an Arthritis Foundation Pediatric Fellow. R. E. K. and J. K. are Leukemia Society of America Scholars.