Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 603–607, January 1991
Immunology

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$^+$, TCR$\alpha\beta^+$ 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 line, 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.**

The CD7 T-cell differentiation antigen is a 40-kDa transmembrane glycoprotein found on most human thymocytes and on a major subset of peripheral blood T lymphocytes (1). The full-length CD7 cDNA, as originally isolated and sequenced by Aruffo and Seed (2), was 1.2 kilobases (kb) long with an open reading frame of 215 amino acids. A second cDNA clone of 1.7 kb, representing an intron-bearing precursor, was also characterized. With significant homology to the variable regions of the immunoglobulin κ chain and the T-cell receptor (TCR) γ chain at the N terminus, CD7 is a single-domain member of the immunoglobulin gene superfamily (2).

CD7 is present on T-cell precursors in fetal liver and thorax as early as 7 weeks of fetal gestation, prior to hematopoietic cell population of the thymic rudiment (3). CD7 is also found on T-cell precursors in thymus and bone marrow. Thus CD7 is an early marker of the human T-cell lineage, appearing before CD2, CD3, CD4, and CD8 (3). It is the most reliable marker for T-cell precursor leukemias and has been demonstrated on both primitive myeloid (4) and lymphoid leukemic cells (5). CD7 is present on subsets of both leukemic and normal hematopoietic cells that retain the ability to differentiate along multiple lineages (3, 5).

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 ref. 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 λFIX was screened with the CD7 cDNA clone, kindly provided by A. Aruffo and B. Seed (2). Two positively hybridizing λ 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^{hi+}, 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 1.5×10^8 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 $MgCl_2$/0.05% Triton

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

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 repelleting, 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 $MgCl_2$/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 *Eco*RI or *Bgl* II/*Hin*dIII, and analyzed on a Southern blot by using the full-length CD7 cDNA labeled with [^{32}P]dCTP as probe. Hybridization was performed overnight at 42°C, in 50% (vol/vol) formamide/3× SSC/0.2% Denhardt's solution/0.05% sodium pyrophosphate/1 mM EDTA, pH 8.0/10 mM Hepes, 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 *Hin*dIII fragment. A genomic library, prepared from a *Sau*3A partial digest of human DNA inserted into λFIX, was screened using the randomly labeled CD7 cDNA. One recombinant, designated λLSCD7-1, contained a 15-kb insert and was selected for characterization based on restriction mapping studies (Fig. 1*A*).

Structure and Sequence of the *CD7* Gene. By Southern blot analysis of λLSCD7-1 with the CD7 cDNA probe, we determined that all coding regions in the 15-kb recombinant

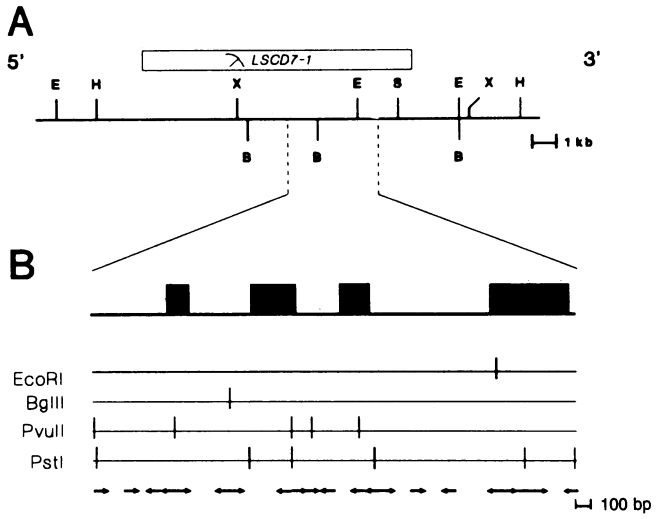

FIG. 1. (*A*) Restriction map showing λLSCD7 clones. B, *Bgl* II; E, *Eco*RI; H, *Hin*dIII; S, *Sal* I; X, *Xba* I. (*B*) Restriction map of *CD7* gene showing exon structure. Arrows represent sequencing primers.

localized to a 3-kb DNA segment. This region was mapped and sequenced, revealing a gene comprised of four exons and three introns (Fig. 1*B*). 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 3*A*). 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

```
                                                    -25
                                                    MetAlaGlyProProArgLeuLeuLeuLeuProLeuL
gtcctgTAGACCCAGAGAGGCTCAGCTGCACTCGCCCGGCTGGGAGAGCTGGGTGTGGGGAACATGGCCGGGCCTCCGAGGCTCCTGCTGCTGCCCCTGC    600
                                        +1
euLeuAlaLeuAlaArgGlyLeuProGlyAlaLeuAlaAlaGln
TTCTGGCGCTGGCTCGCGGCCTGCCTGGGGCCCTGGCTGCCCAAGgtaagagcttcccaggctctccatggccacagctccggagctctccctgccccat    700
gagctcagagcccccagtctgagccacagcacagccccccaggaagcgggtggggtgctgagcggcctccagtgtctgaggactcatttaagagaaggaaa    800
aagggtggacccggtggggagtggccggggctgtccaggcagggccgctgctttgggaggaagaagcccacagtctcggaacacgaggacagcacctccc    900
ccaacaccacagccggtgcccagatctgctccatgccccgtaaggcaccgtgtctttggcgacatgtcagccctgggctgtctcagggccccaccatccc   1000
                                                    +3
                                                    GluValGlnGlnSerProHisCysT
caccactgtcccctgcagggaggacattctctgtccttctggccagactgatggtgacagcccaggtcctccccagAGGTGCAGCAGTCTCCCCACTGCA   1100
                    20                                                  40
hrThrValProValGlyAlaSerValAsnIleThrCysSerThrSerGlyGlyLeuArgGlyIleTyrLeuArgGlnLeuGlyProGlnProGlnAspIl
CGACTGTCCCCGTGGGAGCCTCCGTCAACATCACCTGCTCCACCAGCGGGGGCCTGCGTGGGATCTACCTGAGGCAGCTCGGGCCACAGCCCCAAGACAT   1200
                                        60
eIleTyrTyrGluAspGlyValValProThrThrAspArgArgPheArgGlyArgIleAspPheSerGlySerGlnAspAsnLeuThrIleThrMetHis
CATTTACTACGAGGACGGGGTGGTGCCCACTACGGACAGACGGTTCCGGGGCCGCATCGACTTCTCAGGGTCCCAGGACAACCTGACTATCACCATGCAC   1300
    80                                                  100
ArgLeuGlnLeuSerAspThrGlyThrTyrThrCysGlnAlaIleThrGluValAsnValTyrGlySerGlyThrLeuValLeuValThr
CGCCTGCAGCTGTCGGACACTGGCACCTACACCTGCCAGGCCATCACGGAGGTCAATGTCTACGGCTCCGGCACCCTGGTCCTGGTGACAGgtagggaat   1400
gtgcccatcccagaccccctcccaaccccagctgctggccaggctctgctcccccagcccttgtcgtgggaccctcctcctacatgtgcctgaactgt   1500
tccagctcccagcccactgcccccagcagcctcctagatagctgcccctcctcccctccacagcctttccctgccccgaatcccaaaccccgggggctct   1600
                                                    120
                                        GluGluGlnSerGlnGlyTrpHisArgCysSerAspAlaProProArgAlaSe
aacaggttctccaccgggagaatccccttccttcttttttccttctcagAGGAACAGTCCCAAGGATGGCACAGATGCTCGGACGCCCCACCAAGGGCCTC   1700
                                        140
rAlaLeuProAlaProProThrGlySerAlaLeuProAspProGlnThrAlaSerAlaLeuProAspProProAlaAlaSerAlaLeuProAlaAlaLeu
TGCCCTCCCTGCCCCACCGACAGGCTCCGCCCTCCCTGACCCGCAGACAGCCTCTGCCCTCCCTGACCCGCCAGCAGCCTCTGCCCTCCCTGCGGCCCTG   1800
   160
AlaValIleSerPheLeuLeuGlyLeuGlyLeuGlyValAlaCysValLeuAlaArgThrGln
GCGGTGATCTCCTTCCTCCTCGGGCTGGGCCTGGGGGTGGCGTGTGTGCTGGCGAGGACACACAGgtcagtgtgagccccagctgccacctgcaccccaaag   1900
attgttccctctcctgagagcagcgtgggggcgccaaagccccagaggaaaaccctggtacccgccgcctcctctgcaggctctggaactctcagggtg   2000
gaagtggcctcggcggggagggaggggaccagcaccaaagcgacctgaggggagggagcagcccaggcgctgggcccgcgtgtgcgcgtgtgccggtg   2100
tgcaggtgtgagtgtctgaggggggatctctgaggtctgtgcgtgcaggttggagtgtgcgggcagaggagggagcgcccaggcgtgggcccgcgtgtgc   2200
acgtgtgcaggtgtgagcctgtgtgtggggatctctgaggactgtgcgtgcatgttggagtgtgccggcgtgctgtgtggccgtgggaaggggagggggc   2300
ctcagacagacacagagtctaggcgtggccactcccctcgtccctgctgtttgtctggggctggagtgggggtcccgcttggcttgggcttggtgggggg   2400
gtccaagccacttccactctccttcttcagaggcaccccagggtccgacccagcgcctgcctggggaaggctctgctgggaaccctggggctcaggagacc   2500
tgggggccagcccagagggcccgtgggggtctcgtgttaaaattagaaagctgaccggagtgggtggggtccaggacaccacgctggccactggttgga   2600
                    180                                                 200
                    IleLysLysLeuCysSerTrpArgAspLysAsnSerAlaAlaCysValValTyrGluAspMetSerHisSerArgCysA
gacggcggttctgtctttcagATAAAGAAACTGTGCTCGTGGCGGGATAAGAATTCGGCGGCATGTGTGGTGTACGAGGACATGTCGCACAGCCGCTGCA   2700
                    215
snThrLeuSerSerProAsnGlnTyrGln
ACACGCTGTCCTCCCCCAACCAGTACCAGTGACCCAGTGGGCCCCTGCACGTCCCGCCTGTGGTCCCCCCAGCACCTTCCCTGCCCCACCATGCCCCCCA   2800
CCCTGCCACACCCCTCACCCTGCTGTCCTCCCACGGCTGCAGCAGAGTTTGAAGGGCCCAGCCGTGCCCAGCTCCAAGCAGACACACAGGCAGTGGCCAG   2900
GCCCCACGGTGCTTCTCAGTGGACAATGATGCCTCCTCCGGGAAGCCTTCCCTGCCCAGCCCACGCCGCCACCGGGAGGAAGCCTGACTGTCCTTTGGCT   3000
GCATCTCCCGACCATGGCCAAGGAGGGCTTTTCTGTGGGATGGGCCTGGCACGCGGCCCTCTCCTGTCAGTGCCGGCCCACCCACCAGCAGGCCCCCAAC   3100
CCCCAGGCAGCCCGGCAGAGGACGGGAGGAGACCAGTCCCCCACCCAGCCGTACCAGAAATAAAGGCTTCTGTGCTTCCttttctgacttctgcatttat   3200
tcctcagccagcaggagccagggagggagatggctgacacttcggaccctccccaagcctccgcctggggctctgcag
```

FIG. 2. 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.

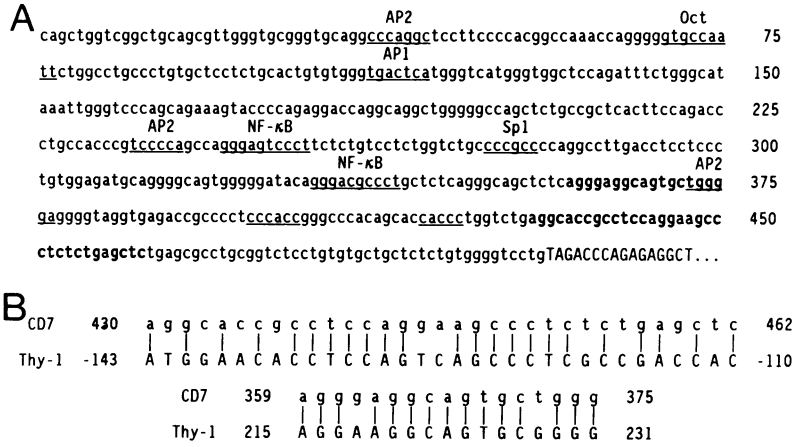

FIG. 3. (*A*) DNA sequence of the 5′ flanking region of the *CD7* gene. The 5′ end of the cDNA is capitalized. Regulatory element motifs are underlined. Regions of similarity with the murine *Thy-1* promoter are in boldface type. (*B*) Demonstration of two regions of similarity within the *CD7* gene 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.

an intron (2). The first intron had stop codons in all three reading frames despite the addition of the extra base. All predicted intron–exon borders had appropriate consensus donor or acceptor splice site sequences. The DNA sequence of the exons was identical to the reported sequence of the CD7 cDNA.

Characterization and Sequence of the 5′ Flanking Region of the *CD7* Gene. By using the first nucleotide of the cDNA as the presumed transcription start site based on the data that both of the cDNA clones isolated by Aruffo and Seed (2) have nearly identical 5′ sequences (the shorter 1.2-kb clone has an extra guanine in position one of the cDNA), we sequenced the region 506 bp upstream that contained no consensus acceptor splice site. Analysis of the 5′ flanking sequence of the *CD7* gene (Fig. 3*A*) revealed no TATA box or CCAAT motifs. The entire *CD7* gene was 66% G+C-rich with several localized areas of higher G+C content in the 5′ flanking region, including a 20-bp region that was 90% G+C centered at nucleotide 406 (position −100 relative to the transcription start site). This is consistent with other TATA-less promoters such as those found in the adenosine deaminase gene, the 3-hydroxy-3-methylglutaryl coenzyme A reductase gene, and the hypoxanthine-guanine phosphoribosyltransferase gene (16).

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 (GGGAGTCCCTT) and nucleotide 331 (GGGACGCCT) (Fig. 3*A*). 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 CCCGCC on the coding strand at nucleotide 275 (Fig. 3*A*) and an 8- of 10-nucleotide match with a proposed consensus sequence for the conserved octamer element ATGCAAAT (17) at nucleotide 69 (Fig. 3*A*). One possible transcription factor AP-1 and several potential transcription factor AP-2 binding sequences were identified (17), as shown in Fig. 3*A*.

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^{hi+}, CD4$^-$, CD8$^-$, CD3$^-$ progenitor cell leukemia line DU528 (14), the CD7$^-$, CD4$^+$, CD8$^+$, TCR$\alpha\beta^+$ T-cell line DU980 (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. 4*A*). Studies done with the DU980 T-cell line showed that DNase I hypersensitive site 1 was located ≈5 kb upstream from the transcription initiation site (Fig. 4*B*). 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. 4*C*). Both DNase I hypersensitive sites have been confirmed by mapping with additional restriction enzymes, including *Eco*RI, *Hin*dIII, 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

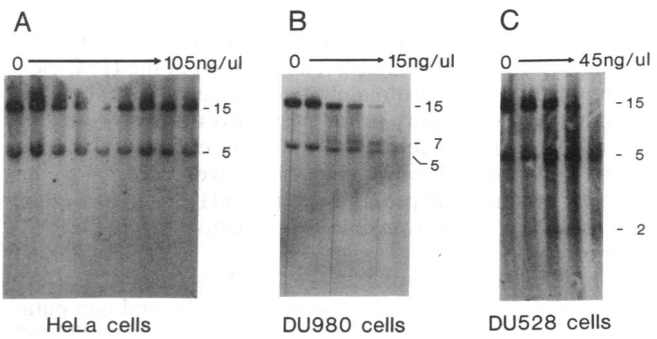

FIG. 4. DNase I hypersensitivity experiments. DNA was exposed to increasing concentrations of DNase I and digested with *Eco*RI. The full-length CD7 cDNA was used as probe.

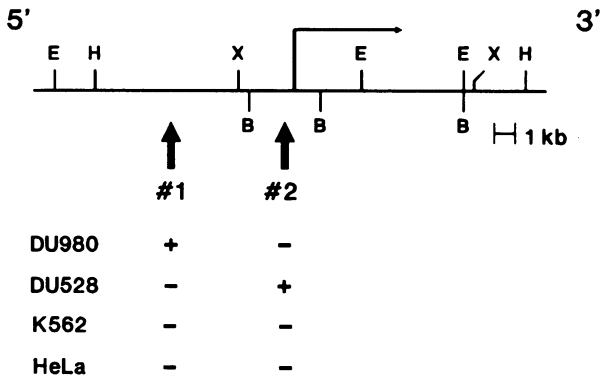

FIG. 5. Summary of DNase I hypersensitivity experiments. Genomic restriction map of *CD7* shows location of hypersensitive sites 1 and 2. B, *Bgl* II; E, *Eco*RI; H, *Hin*dIII; X, *Xba* I. +, Presence of DNase 1 hypersensitive site; −, absence of DNase 1 hypersensitive site.

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 phosphotidylinositol-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 deoxyribonucleotidyltransferase gene. Another transcription initiator element (HIP1) 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 HIP1 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 from 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 (δ, ε, and γ chains), neural cell adhesion molecule, carcinoembryonic 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 one of the two alternative first 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. 3*B*).

The two promoters of murine Thy-1 show little similarity in DNA sequence but share a conserved nonamer, CTCCCTGCT, at 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; GCCCTCTCTG at position −55 from the presumed start site and CTCCTGT at position −29. 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

intron. Possible AP-1 and AP-2 motifs were identified in the first intron of *CD7*.

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.

1. Haynes, B. F., Eisenbarth, G. S. & Fauci, A. S. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5829–5833.
2. Aruffo, A. & Seed, B. (1987) *EMBO J.* **6,** 3313–3316.
3. Haynes, B. F., Denning, S. M., Singer, K. H. & Kurtzberg, J. (1989) *Immunol. Today* **10,** 87–91.
4. Coco, F. L., de Rossi, G., Pasqualetti, D., Lopez, M., Diverio, D., Latagliata, R., Fenu, S. & Mandelli, F. (1989) *Br. J. Haematol.* **73,** 480–485.
5. Kurtzberg, J., Waldmann, T. A., Davey, M. P., Bigner, S. H., Moore, J. O., Hershfield, M. S. & Haynes, B. F. (1989) *Blood* **73,** 381–390.
6. Ware, R. E., Scearce, R. M., Dietz, M. A., Starmer, C. F., Palker, T. J. & Haynes, B. F. (1989) *J. Immunol.* **143,** 3632–3640.
7. Jung, L. K. L. & Fu, S. M. (1988) *Eur. J. Immunol.* **18,** 711–715.
8. Ledbetter, J. A., June, C. H., Grosmaire, L. S. & Rabinovitch, P. S. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 1384–1388.
9. Ingraham, H. A. & Evans, G. A. (1986) *Mol. Cell. Biol.* **6,** 2923–2931.
10. Spanopoulou, E., Giguere, V. & Grosveld, F. (1988) *Mol. Cell. Biol.* **8,** 3847–3856.
11. Sehgal, A., Patil, M. & Chao, M. (1988) *Mol. Cell. Biol.* **8,** 3160–3167.
12. Tewari, D. S., Cook, D. M. & Taub, R. (1989) *J. Biol. Chem.* **264,** 16238–16245.
13. Meeker, T. C., Loeb, J., Ayres, M. & Sellers, W. (1990) *Mol. Cell. Biol.* **10,** 1680–1688.
14. Kurtzberg, J., Bigner, S. H. & Hershfield, M. S. (1985) *J. Exp. Med.* **162,** 1561–1578.
15. Enver, T., Brewer, A. C. & Patient, R. K. (1985) *Nature (London)* **318,** 680–683.
16. Dynan, W. S. (1986) *Trends Genet.* **6,** 196–197.
17. Wingender, E. (1988) *Nucleic Acids Res.* **16,** 1879–1902.
18. Williams, A. F. & Barclay, A. N. (1988) *Annu. Rev. Immunol.* **6,** 381–405.
19. Diamond, D. J., Clayton, L. K., Sayre, P. H. & Reinherz, E. L. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 1615–1619.
20. Saito, H., Koyama, T., Georgopoulos, K., Clevers, H., Haser, W. G., LeBien, T., Tonegawa, S. & Terhorst, C. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 9131–9134.
21. Clevers, H. C., Dunlap, S., Wileman, T. E. & Terhorst, C. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8156–8160.
22. Gorman, S. D., Tourvieille, B. & Parnes, J. R. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7644–7648.
23. Norment, A. M., Lonberg, N., Lacy, E. & Littman, D. R. (1989) *J. Immunol.* **142,** 3312–3319.
24. Chen, A., Reyes, A. & Akeson, R. (1990) *Mol. Cell. Biol.* **10,** 3314–3324.
25. Schrewe, H., Thompson, J., Bona, M., Hefta, L. J. F., Maruya, A., Hassauer, M., Shively, J. E., von Kleist, S. & Zimmerman, W. (1990) *Mol. Cell. Biol.* **10,** 2738–2748.
26. Giguere, V., Isobe, K. & Grosveld, F. (1985) *EMBO J.* **4,** 2017–2024.
27. Ingraham, H. A., Lawless, G. M. & Evans, G. A. (1986) *J. Immunol.* **136,** 1482–1489.
28. Seki, T., Spurr, N., Obata, F., Goyert, S., Goodfellow, P. & Silver, J. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6657–6661.
29. van Rijs, J., Giguere, V., Hurst, J., van Agthoven, T., van Kessel, A. G., Goyert, S. & Grosveld, F. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 5832–5835.
30. Low, M. G. & Kincade, R. W. (1985) *Nature (London)* **318,** 62–64.
31. Smale, S. T. & Baltimore, D. (1989) *Cell* **57,** 103–113.
32. Means, A. L. & Farnham, P. J. (1990) *Mol. Cell. Biol.* **10,** 653–661.
33. Vidal, M., Morris, R., Grosveld, F. & Spanopoulou, E. (1990) *EMBO J.* **9,** 833–840.
34. Hirsch, M. R., Gaugler, L., Deagostini-Bazin, H., Bally-Cuif, L. & Goridis, C. (1990) *Mol. Cell. Biol.* **10,** 1959–1968.
35. Yu, H., Porton, B., Shen, L. & Eckhardt, L. A. (1989) *Cell* **58,** 441–448.
36. Ehrlich, R., Maguirre, J. E. & Singer, D. S. (1988) *Mol. Cell. Biol.* **8,** 695–703.
37. Calame, K. & Eaton, S. (1988) *Adv. Immunol.* **43,** 235–275.
38. Koller, B. H., Geraghty, D., Orr, H. T., Shimizu, Y. & DeMars, R. (1987) *Immunol. Res.* **6,** 1–10.
39. Dedrick, R. L. & Jones, P. P. (1990) *Mol. Cell. Biol.* **10,** 593–604.
40. Lemke, G., Lamar, E. & Patterson, J. (1988) *Neuron* **1,** 73–83.
41. Sen, R. & Baltimore, D. (1986) *Cell* **46,** 705–716.
42. Meyer, K. B. & Neuberger, M. S. (1989) *EMBO J.* **8,** 1959–1964.
43. McDougall, S., Peterson, C. L. & Calame, K. (1988) *Science* **241,** 205–208.
44. Krimpenfort, P., de Jong, R., Uematsu, Y., Dembic, Z., Ryser, S., von Boehmer, H., Steinmetz, M. & Berns, A. (1988) *EMBO J.* **7,** 745–750.
45. Winoto, A. & Baltimore, D. (1989) *EMBO J.* **8,** 729–733.
46. Adkins, B., Mueller, C., Okada, C. Y., Reichert, R. A., Weissman, I. & Spangrude, G. (1987) *Annu. Rev. Immunol.* **5,** 325–365.
47. McKenzie, J. L. & Fabre, J. W. (1981) *J. Immunol.* **126,** 843–850.
48. Kroczek, R. K., Gunter, K. C., Germain, R. N. & Shevach, E. M. (1986) *Nature (London)* **322,** 181–184.