Characterization and expression of the murine CD3-ε gene
(T-cell receptors for antigen/CD3 gene family/T-cell-specific regulatory elements)
HANS CLEVERS*, SABRINA DUNLAP*, HARUO SAITO†, KATIA GEORGOPoulos*, THOMAS WILEMAN*, and COX TERHORST*
*Laboratory of Molecular Immunology and †Division of Tumor Immunology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115
Communicated by Frank Lilly, July 25, 1988

ABSTRACT The receptor for antigen on the surface of T lymphocytes consists of a variable disulfide-bridged heterodimer (TCR-α/β or -γ/δ) associated with invariant CD3 proteins (CD3-γ, -δ, -ε, and -ζ). The genes coding for the CD3 proteins are expressed in the earliest recognizable thymocytes, preceding the rearrangement and expression of the TCR genes. The isolation, characterization, and in vitro expression of the murine CD3-ε gene, as reported here, represent obligatory steps toward our understanding of the complex rules that govern T-cell-specific gene expression. The CD3-ε gene was transcribed from a non-TATA promoter and consisted of eight exons, two of which were unusually small (18 and 15 base pairs). The transmembrane exon was found to be homologous to the transmembrane exons of the CD3-γ and CD3-δ genes. In transient-transfection experiments, a genomic fragment comprising 4 kilobases of upstream sequence and extending into the second exon sufficient to drive the expression of a reporter gene in murine T cells.

Antigen receptors on the surface of human and murine T lymphocytes have been discovered through two lines of experimental research. (i) Monoclonal antibodies specific for functionally competent in vitro cultured T-cell clones were prepared. These clone-specific, or clonotypic, antibodies were then shown to affect the function of their respective T-cell clones and were used to characterize these antigen receptors biochemically. (ii) Genes were sought that specifically rearranged only in T cells. Subsequently, these genes were found to code for the structures recognized by the clonotypic antibodies (1–3).

Several lines of evidence support the concept that T-cell receptor (TCR) heterodimers are associated with the CD3 proteins. (i) Coprecipitation of the CD3 proteins with anti-TCR reagents and vice versa indicated a close proximity of these cell surface structures (4). (ii) Incubation of T cells with monoclonal antibodies directed against either TCR or CD3 leads to “comodulation,” or simultaneous disappearance from the cell surface, of both structures (5). (iii) Several mutants that do not express the TCR/CD3 complex on their surface were found to lack the TCR-α or TCR-β mRNA. Upon gene transfer of cDNA encoding the appropriate TCR chain, cell surface expression of the TCR/CD3 complex could be restored (6).

The murine TCR/CD3 complex consists of the TCR heterodimer and four CD3 proteins: two glycoproteins, CD3-γ (21 kDa) and CD3-δ (28 kDa), and two non-glycosylated proteins, CD3-ε (25 kDa) and CD3-ζ (17 kDa) (7, 8). Murine CD3-ε undergoes posttranslational modification, since the molecular mass of the protein backbone predicted from the cDNA sequence is 18 kDa (9). This was confirmed by in vitro translation of the cDNA clone pDL1 (9, 10). No conclusive experimental evidence has been provided for either a lipid anchor or oligosaccharide side chains.

Using antibodies generated with an octapeptide representing the N terminus of murine CD3-ε, we determined that the N-terminal 88 amino acids of murine CD3-ε are located on the outside of the cell (11). A 65-amino acid cytoplasmic domain containing a high level of basic amino acids follows a 26-amino acid transmembrane region. As is the case with the other known CD3 proteins, murine CD3-ε contains a negatively charged residue (aspartic acid) in its transmembrane region (9). This residue is probably involved in a salt bridge with one of the positively charged residues in the transmembrane segments of TCR-α, -β, -γ, or -ζ.

To begin to understand the genetic events that commit hemopoietic cells to the T-cell lineage, we decided to study the exclusive expression of the CD3-ε gene in T cells. Coordinate expression of the CD3-γ, -δ, and -ε genes is one of the earliest events in maturation of T cells within the thymus gland and precedes expression of TCR-α and -β in mice and humans (12–14). The intracellularly expressed CD3-γ, -δ, and -ε proteins form a subcomplex. Later in differentiation, TCR-α and -β bind to this subcomplex and subsequently the TCR/CD3 complex is transported to the cell surface (15). Thus, the CD3 genes set the stage for the eventual functional expression of TCRs on the surface of mature T lymphocytes.

Earlier we (16, 17) described the human and murine CD3-δ and CD3-γ genes. These genes are closely linked both in mice and in humans separated by ~1.4 kilobases (kb) and organized in a head-to-head configuration (16–18). Exon/intron boundaries are located at homologous positions in the two genes with the exception that the CD3-γ gene contains an additional 24-base-pair (bp) coding exon (exon 2) and a separate 3’ untranslated exon. Based on this similarity of gene organization, on the physical linkage, and on the high sequence homology of the CD3-γ and CD3-δ genes, it is highly likely that these two genes arose through a duplication event.

In this study we characterize the exon/intron organization and the transcription start site of the murine CD3-ε gene,‡

MATERIALS AND METHODS

Characterization of the CD3-ε Gene. Standard techniques were used for isolation, subcloning, and sequencing of the genomic DNA. The murine cDNA clone pDL1 was used as a probe to screen a AEMBL4 phage library containing 15- to 20-kb partial Sau3A1 fragments of BALB/c mouse fibroblast DNA. Ten positive bacteriophages were screened with two RNA probes: a 92-base probe complementary to the 5’ sequence of pDL1 and a 670-base probe complementary to the 3’ sequence of pDL1. The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J04110).

†The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J04110).

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.
the 3’ sequence of pDL1. These probes were generated from pDL1 in the pGEM-4 vector by using SP6 and T7 RNA polymerases according to the manufacturer’s instructions (Promega Biotec, Madison, WI). The gene was found to be contained within the inserts of two of these bacteriophages, φH3 and φH4. The plasmids pE1, pE2, and pE3 were constructed by ligating HindIII fragments obtained from the bacteriophage φH3 into the pGEM-4 vector; pE4 was constructed by ligating a HindIII fragment obtained from bacteriophage φH4 and pGEM-4. Exons 1, 2, and 5–8 were mapped by restriction enzyme analysis and probing with pDL1. The small exons 3 and 4 were found by using oligonucleotide probes (TGGCACTTTGGCAGAGG and ATGCCGAGAACCATTG, respectively; see Fig. 2). Restriction fragments containing exon sequences were subcloned from the plasmid constructs in pGEM-4 for sequence analysis. Double-stranded sequencing was performed by the dioxyde chain-termination method in pGEM-4, essentially according to the manufacturer’s instructions.

**Primer Extension Analysis.** The insert from pDL1 was digested with Xmn I (restriction site at bp 94) and Smal I (polynuker site 5’ of the insert) and dephosphorylated with calf intestinal phosphatase. The small 5’ fragment was isolated and subsequently digested with Nla IV (site at bp 11). The 83-bp fragment was then end-labeled at its negative strand with phosphate T4 polynucleotide kinase (see Fig. 3a). Approximately 100,000 dpm of this probe was coprecipitated with 50 µg of total RNA from mouse L cells or from the mouse T-cell hybridoma DO11.10. The resulting pellet was digested and dissolved in 30 µl of hybridization buffer (80% formamide/1 mM EDTA/40 mM Mops, pH 8.2/400 mM NaCl. After 3 min of heating at 85°C, hybridization was allowed to occur for 16 hr at 42°C. After reprecipitation, the samples were taken up in reverse transcriptase reaction mixture [0.1 M Tris, pH 8.3/140 mM KCl/10 mM MgCl2/4 mM dithiothreitol/0.5 mM each dNTP, containing 10 units of reverse transcriptase (Promega Biotec)] and incubated for 10 min at room temperature and then for 20 min at 42°C. The extended products were then analyzed by electrophoresis alongside a dioxyde sequencing “ladder” in a denaturing 8% polyacrylamide gel.

**Quantitative S1 Nuclease Analysis.** Plasmid Constructs. The construction of pμE4 has been described in detail (11). In short, it contains DNA from four different sources (see Fig. 4a). (i) The plasmid backbone is a Sal I-Xho I piece taken from pμ (19) and contains pBR322 sequences [the ampicillin-resistance gene (AmpR) and the replication origin (ori)], a truncated form of the murine histone H4 gene serving as an internal control for transfection efficiency, and the polyoma virus early region (PyT), which allows plasmid replication in mouse cells. (ii) A 4.5-kb fragment (EcoRI/partial BsmI 1) from pE1 extends into exon 2 to a point 14 bp upstream of the ATG codon of the CD3-ε gene. (iii) The human CD3-ε cDNA clone pDJ4 (20) serves as reporter gene. (iv) Simian virus 40 sequence [small tumor antigen (t) intron and polyadenylation signal] replaces the untranslated tail of DJ4, thus stabilizing the hybrid message. pμE10 was derived from pμE4 by depleting the mouse genomic CD3-ε sequence.

pProbe3 was generated by subcloning an Nsi I (site at bp −211 relative to the transcription start site)–Pst I fragment from pμE4 in pGEM-4, and subsequently deleting the first intron by replacing a Ban I (site in exon 1)–Hga I (site in exon 2) fragment by a corresponding fragment taken from the cDNA clone pDL1 (see Fig. 4a).

**S1 nuclease protection.** DO11.10 cells (107) were resuspended in 15 ml of transfection buffer (RPMI 1640/50 mM Tris, pH 7.4, containing DEAE-dextran (250 µg/ml) and plasmid (10 µg) and incubated at 37°C for 30 min. The cells were then washed twice and cultured for 48 hr, after which total RNA was prepared. The Nsi I–Pst I fragment from pProbe3 was dephosphorylated and end-labeled with T4 kinase (Fig. 4a). The probe (50,000 dpm) was hybridized with 20 µg of RNA as described for the primer extension analysis, in a total volume of 40 µl. After 16 hr the hybrids were digested with 50 units of S1 nuclease at 37°C for 60 min. The protected fragments were then visualized in a 6% polyacrylamide sequencing gel. As a control for transfection efficiency, S1 analysis was performed as described by Groschedl and Baltimore (19).

**Immunofluorescence Staining.** Transfected cells (106) were resuspended in 100 µl of the anti-human CD3-ε antibody UCHT-1 (1:100 dilution in phosphate-buffered saline containing 1% bovine serum albumin) and incubated for 30 min on ice. The cells were then washed, incubated with fluorescein-conjugated goat anti-mouse immunoglobulin for 30 min on ice, washed, and analyzed on an EPICS V cell sorter (Coulter).

**RESULTS**

**Isolation and Partial Nucleotide Sequence of the Murine CD3-ε Gene.** A genomic library, made with 15- to 20-kb fragments from partial Sau3A1 digestion of BALB/c mouse fibroblast DNA, in the bacteriophage λEMBL4 was screened with a murine CD3-ε cDNA clone, pDL1 (9). Ten positive bacteriophages were isolated from a total of 106 λEMBL4 bacteriophages. From two of these bacteriophages (φH3 and φH4) with overlapping inserts, four plasmid constructs, pE1, pE2, pE3, and pE4, were made that together contained the complete gene (Fig. 1). Exon sequences were mapped by restriction enzyme analysis, by probing with pDL1 and appropriate oligonucleotides, and by DNA sequencing.

The CD3-ε gene was located on an 11-kb DNA segment (Fig. 1). Eight exons were identified by comparison with the cDNA sequence (Fig. 2). Consensus 5’ and 3’ splice recognition sequences were present at all exon/intron boundaries. Thus, we distinguished a 64-bp 5’ untranslated first exon and a 102-bp second exon encoding most of the leader peptide. With the use of appropriate oligonucleotides, exon 3 (18 bp), encoding the remainder of the leader peptide, and exon 4 (15 bp), encoding the N terminus of the mature protein, were found within a 6-kb segment. Although the exact boundary of leader peptide and mature protein has not been determined by amino acid sequencing, we assume that, based on homology with the human CD3-ε protein, the aspartic residue is the N-terminal amino acid (Fig. 2) (9, 20). Exon 5 (216 bp) encodes the extracellular portion of CD3-ε; exon 6 (168 bp) contains the sequence for the putative transmembrane segment. The intracellular domain is encoded by exons 7 and 8 (47 and 844 bp, respectively).

**Localization of the Transcription Start Site of the CD3-ε Gene.** Primer extension analysis of mRNA was used to locate the transcription start site, 40 bp upstream of the exon 1 sequence present in the murine cDNA clone pDL1 (Fig. 3). This was confirmed by S1 nuclease analysis of RNA of murine T cells transiently transfected with pμE4 (see below). Analysis of the genomic nucleotide sequence revealed that the CD3-ε gene, like the CD3-γ/δ gene pair (16–18), is transcribed from a non-TATA-box promoter; a CAAT box could not be found in any of the CD3 genes.

**In Vitro Expression of the Murine CD3-ε Gene.** To examine whether the 5’ flanking region of the CD3-ε gene was sufficient to drive expression of a reporter gene in murine T cells, a plasmid (pμE4) containing a hybrid gene was constructed (Fig. 4a). By using a replicon system with the polyoma virus origin of replication, a genomic fragment comprising 4 kb of upstream sequence and extending into exon 2 was attached to the human CD3-ε cDNA. This plasmid construct was transfected into a murine T–T hybridoma cell line (DO11.10). When mRNA from these cells was
Fig. 1. Partial restriction map of the mouse CD3-e gene. The mouse CD3-e gene spans 11 kb and contains eight exons. These exons are projected on the corresponding CD3-e mRNA sequence. 5' UT, 5' untranslated region; LP, leader peptide; EX, extracellular domain; TM, transmembrane segment; IN, intracellular domain; 3' UT, 3' untranslated region. Restriction enzyme sites: e. Acc I; e. Xba I; h, HindIII; k, Kpn I; l, Sca I; p, Pst I; s, Sac I; x, Xmn I.

analyzed after 48 hr for the presence of reporter gene message by quantitative S1 nuclease analysis, it was found that the transcription start site in the cloned CD3-e gene fragment (Fig. 4b) was identical to that determined by primer extension analysis (Fig. 3). Furthermore, expression of the protein product of the reporter gene, the human CD3-e chain, could easily be detected by cell surface staining of the transfected murine T cells (Fig. 4c).

DISCUSSION

The murine CD3-e gene was isolated and characterized as a first step in the elucidation of the regulatory mechanisms governing its unique expression in all cells of thymic origin. The murine CD3-e gene, consisting of eight exons, is somewhat different from the CD3-δ and CD3-γ genes in that two of its exons are unusually small (15 and 18 bp). However, from comparisons of the protein sequences of murine CD3-e with the other two CD3 polypeptide chains, it is evident that a CD3 family exists (9). Whereas CD3-γ and -δ are closely related, the sequence homologies between CD3-e and CD3-γ/δ center around the cysteine residues. The most striking homology between the CD3 genes was found at the 5' end of the transmembrane exon, exon 6 (Fig. 5). Alignment of the transmembrane exon of the murine CD3-e gene with the corresponding exons in the murine and human CD3-γ and CD3-δ genes demonstrated that the 5' boundaries were placed at identical base pairs in all these genes (Fig. 5). Since partial sequences of the human CD3-e gene were available (H.C., S.D., and C.T., unpublished data), the 5' boundary of its transmembrane exon was included.

The high conservation of the amino acid sequence N-terminal of the transmembrane segment implies that it plays an important role in the functioning of each CD3 molecule. A search of the National Biomedical Research Foundation Protein Identification Resource (Release 15: Dec. 28, 1987) for the consensus sequence Xaa-Cys-Glx-Xaa-Cys-Glu-Xaa-Asx revealed it to be unique to the CD3 proteins. This protein sequence might serve as a signal involved in the assembly of the TCR/CD3 complex, possibly through interaction with the recently recognized protein CD3-ω that is transiently associated with unassembled CD3 and TCR chains in the endoplasmic reticulum (15, 21).

Based on amino acid sequence comparisons and secondary structure predictions, the CD3 genes are thought to be related

Fig. 2. Partial nucleotide sequence of the mouse CD3-e gene. Exon sequences predicted from the CD3-e cDNA clone pDL1 are capitalized; individual exons are boxed. The leader peptide and the transmembrane segment of the predicted protein sequence are underlined. The polyadenylation signal (AATAAA) is boxed. An arrow indicates the transcriptional start site (see Fig. 3). Upstream sequences with strong homology to sequences in a conserved upstream stretch in human and mouse CD3-δ genes are indicated by broken underlining.
to the immunoglobulin gene superfamily (9). Interestingly, the highest homologies exist with the domains of the neural cell adhesion molecule (N-CAM) (9), which, like the human CD3-γ, -δ, and -ε genes, maps to band q23 on chromosome 11. Analysis by pulsed-field electrophoresis has shown that the human CD3-γδ pair is located within 300 kb of the human CD3-ε gene (ref. 18; H.C., S.D., and C.T., unpublished work). Since the murine CD3 genes map to chromosome 9.

Fig. 4. Assay for transient expression of a reporter gene driven by the 5′ upstream region of the murine CD3-ε gene. (a) pμE4 was constructed to test the ability of the CD3-ε promoter to drive expression of a reporter gene in murine T cells. The backbone is taken from pμ (19); Amp (ampicillin-resistance gene) and ori (origin of replication) are from pBR322; H4 is a truncated histone H4 gene serving as an internal control for transection efficiency; PyT, the polyoma early region, allow for replication of the plasmid in murine cells. We inserted 4.5 kb of the 5′ flanking region of the murine CD3-ε gene, containing the putative promoter and extending into the second exon. The human cDNA clone DJ4 (hatched box) (20) with simian virus 40 small tumor antigen (t) intron and polyadenylylation signal (black boxes) was used as a reporter gene. Arrows indicate direction and extent of transcription. pμE10 was constructed from pμE4 by deleting the CD3-ε genomic fragment. pProbe 3 was constructed as a probe for S1 nuclease analysis of RNA transcribed from pμE4. It is derived from an Nsi I–Pst I fragment taken from pμE4, representing bp +211 to +846. The first intron has been removed (see Materials and Methods). Boxed sequences will be protected by RNA from pμE4. Endogenous CD3-ε will not protect the end-label of the probe, because of the presence of 10 bp of polylinker sequence between the Bsm I and Pst I sites in the probe. (Since pProbe is derived from pμE4, these 10 bp are present in pμE4). I and II, exons 1 and 2. (b) S1 nuclease analysis of total RNA from DO11.10 cells mock-transfected (−) or transfected with pμE10 (E10) or pμE4 (E4). pProbe was the probe. The transfection efficiencies of E10 and E4 were identical as determined by S1 analysis of the H4 internal control (data not shown). Transfection with pμE4 resulted in protection of a fragment of the probe (arrow). The start site predicted from the length of this fragment coincides exactly with the start site of the CD3-ε gene as determined by primer extension. Scale at left shows length in bases. (c) DO11.10 cells transfected with pμE4 (E4) or the promoterless pμE10 (E10) were surface-stained with the anti-human CD3-ε monoclonal antibody UCHT-1. Transfection with pμE10 resulted in background fluorescence, cell transfected with pμE4 expressed the human CD3-ε protein on their surface.
(22, 23), it is plausible that the murine CD3-δ gene could also be located at a similar distance from the murine CD3-γ/δ pair.

Given the evolutionary relation and the coordinate expression of the CD3 genes, it is conceivable that their genes contain common regulatory elements. Some of these regulatory elements would have to govern the T-cell-specific expression of CD3-γ, -δ, and -ε; others will be shared with other genes, and some elements may be unique for each of the CD3 genes.

Exons <20 bp long, as found in the murine CD3-ε gene, have rarely been reported: the chicken collagen gene contains an 11-bp exon (24), the human renin gene contains a 9-bp exon (25), and the smallest exon reported is a 6-bp exon in the cardiac troponin T gene (26). The fast skeletal muscle troponin T gene has a very interesting organization in this regard: it contains at least five small exons of length 12–18 bp that can be differentially spliced into the mature message. This alternative splicing is subject to tissue- and developmental stage-specific regulation (27). Such a mechanism requires the preservation of the reading frame; i.e., each exon involved in this process must contain a multiple of 3 bp. Each of the two small adjacent exons in the CD3-ε gene contains a multiple of 3 bp. Although preliminary studies have not provided evidence for the existence of alternatively spliced CD3-ε mRNA species, it is conceivable that they could be produced from this gene.

In conclusion, the isolation, characterization, and in vitro expression of the murine CD3-ε gene are the first steps toward a deeper understanding of T-cell-specific gene regulation. Since the CD3 genes play a sine qua non role in TCR cell surface expression, their expression in the earliest recognizable T cells represents a pivotal step in the transition from prothymocyte into thymocyte.


![Fig. 5. Nucleotide and amino acid homology comparison at the 5' boundary of the transmembrane exons of murine (M) and human (H) CD3-ε, CD3-δ, and CD3-γ genes. The transmembrane exons of the six genes are aligned at 5' boundaries. Sequences are shown through the first codon of the transmembrane segment.](image-url)