Molecular cloning of the CD3 η subunit identifies a CD3 ζ -related product in thymus-derived cells

(T-cell antigen receptor/signal transduction/alternative splicing/thymic development)

YONG-JIU JIN*[†], LINDA K. CLAYTON*[†], FRANK D. HOWARD*[‡], SHIGEO KOYASU*[†], MONICA SIEH*,
RUTH STEINBRICH*, GEORGE E. TARR*[†], AND ELLIS L. REINHERZ*[‡]

*Laboratory of Immunobiology, Dana–Farber Cancer Institute and Departments of [†]Pathology and [‡]Medicine, Harvard Medical School, Boston, MA 02115

Communicated by Baruj Benacerraf, February 8, 1990

ABSTRACT The CD3 η subunit of the T-cell antigen receptor forms a heterodimeric structure with the CD3 ζ subunit in thymus-derived lymphoid cells and is apparently involved in signal transduction through the receptor. Here we report the primary structure of murine CD3 η as deduced from protein microsequencing and cDNA cloning. The mature protein is divided into three domains: a 9-amino acid extracellular segment, a 21-amino acid transmembrane segment including a negatively charged residue characteristic of CD3 subunits, and a 155-amino acid cytoplasmic tail. The NH₂-terminal sequences of CD3 η and CD3 ζ are identical through amino acid 122 of each mature protein but then diverge in the remainder of their respective COOH-terminal regions, consistent with alternatively spliced products of a common gene. The cytoplasmic domain of CD3 η is 42 amino acids larger than that of CD3 ζ but lacks one of six potential tyrosine phosphorylation sites as well as a putative nucleotide binding site previously identified in CD3 ζ . These structural features presumably account for the difference between CD3 η and CD3 ζ function and are consistent with the notion that CD3 η may be an important component of a T-cell receptor isoform(s) during thymic development.

The T-cell receptor (TCR) is a molecular complex consisting of multiple subunits that mediates the recognition of antigen in the context of a particular major histocompatibility complex (MHC) product (reviewed in refs. 1–3). The antigen/MHC product-binding moiety, termed Ti, is a disulfide-linked heterodimer consisting of one α and one β subunit on the vast majority of peripheral T lymphocytes. Both subunits are immunoglobulin-like, being composed of variable and constant domains, the former encoding the unique specificity of a given T-cell clone. Ti, in turn, is noncovalently associated with a set of invariant subunits, collectively termed CD3, involved in signal transduction. CD3 γ , δ , and ϵ subunits are encoded by related genes that are closely linked. In addition, a fourth CD3 subunit present in all TCRs is the 16-kDa CD3 ζ subunit, which is structurally (4) and genetically (5) distinct from the other invariant chains and exists predominantly as a 32-kDa homodimer within the TCR complex. A role for CD3 ζ in assembly and expression of the TCR complex as well as signal transduction upon antigen triggering has been documented (6).

CD3 ζ can be complexed in a heterodimeric form with a 22-kDa subunit termed CD3 η (7). Despite the fact that the copy number of TCR-associated CD3 ζ – η heterodimers is 1 order of magnitude less than that of CD3 ζ – ζ homodimers in antigen-specific T-cell hybridomas, recent evidence suggests that the CD3 ζ – η structure is important in coupling the TCR to inositolphospholipid hydrolysis (8). CD3 ζ – η may also be linked to antigen-induced cell death in these hybridomas (9).

There is virtually no definitive structural information concerning CD3 η except that it is not a posttranslationally modified form of CD3 ζ resulting from phosphorylation, glycosylation, or sulfation (10). Although amino acid analysis by metabolic labeling has shown clear-cut differences between CD3 η and CD3 ζ , certain anti-CD3 ζ peptide antisera react with CD3 η , implying a structural relatedness (10). The present study was aimed at elucidating the complete primary structure of murine CD3 η .[§]

MATERIALS AND METHODS

Materials. Restriction endonucleases were from New England Biolabs. SP6 and T7 RNA polymerases, RNasin, rNTPs, RQ1 DNase, and wheat germ extract were from Promega. Guanosine(5')triphospho(5')guanosine [G(5')ppp(5')G] was from Pharmacia. Translation-grade [³⁵S]methionine (>1000 Ci/mmol; 1 Ci = 37 GBq) was from NEN/DuPont. Protein molecular weight markers were from Bio-Rad. Premixed acrylamide/*N,N'*-methylenebisacrylamide (Protogel) was from National Diagnostics (Manville, NJ). A full-length human CD3 ζ cDNA insert in pGEM-3Z (referred to here as pGEM3Z- ζ) was provided by A. M. Weissman (National Cancer Institute, Bethesda, MD). The T-cell hybridoma 3DO54.8 was provided by P. Marrack (National Jewish Hospital, Denver, CO). M12.1.4 cells were from L. Glimcher (Harvard School of Public Health, Boston, MA).

Protein Purification and Microsequencing. The murine ovalbumin-specific, I-A^d-restricted T–T hybridoma cell line 3DO54.8 (11) was grown in RPMI 1640 medium supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 50 units of penicillin G per ml, 50 μ g of streptomycin sulfate per ml, 50 μ M 2-mercaptoethanol, and 1 mM sodium pyruvate in spinner culture. For each two-dimensional gel electrophoresis, 2×10^9 murine 3DO54.8 cells were lysed for 2 hr in 20 ml of 1% digitonin lysis buffer (1% digitonin/150 mM NaCl/20 mM Tris, pH 7.4/1 mM phenylmethylsulfonyl fluoride/10 mM iodoacetamide with aprotinin at 10 μ g/ml), and a post-nuclear supernatant was prepared by centrifugation at 12,000 $\times g$ for 30 min. Immunoprecipitation was performed overnight at 4°C with 0.2 ml of Sepharose beads coupled with an anti-murine CD3 ϵ IgG monoclonal antibody (mAb), 145.2C11 (2C11) (3–5 mg/ml). Thereafter, the pelleted beads were washed once with 0.05% digitonin/150 mM NaCl/20 mM Tris, pH 7.4, twice with 150 mM NaCl/20 mM Tris, pH 7.4, and once with 20 mM Tris, pH 7.4. The protein was eluted from the beads by boiling for 5 min in 50 μ l of Laemmli nonreducing NaDodSO₄ sample buffer. Samples were subjected to electrophoresis in 12.5% polyacrylamide, using the

Abbreviations: aa, amino acid(s); mAb, monoclonal antibody; MHC, major histocompatibility complex; TCR, T-cell antigen receptor; PCR, polymerase chain reaction; PVDF, poly(vinylidene difluoride).
[§]The sequence reported in this paper has been deposited in the GenBank data base (accession no. M33158).

buffer system of Laemmli. The gel was then extracted and equilibrated with reducing Laemmli NaDodSO₄ sample buffer for 20 min before electrophoresis in a 12.5% polyacrylamide gel. After this two-dimensional diagonal gel electrophoresis, the protein sample was blotted to poly(vinylidene difluoride) (PVDF) membrane and visualized by Coomassie blue staining.

Spots corresponding to individual proteins were cut out of 40 PVDF blots and loaded into an Applied Biosystems model 470A gas-phase sequencer (12) with a 120A phenylthiohydantoin (PTH) analyzer on line. Each sample was sequenced with the 03RPTH program. After eight cycles, it was apparent that both CD3 ζ and CD3 η were NH₂-terminally blocked, so the 03RPTH run was interrupted after trifluoroacetic acid cleavage and chlorobutane extraction. Subsequently, CNBr cleavage was performed gas-phase in a sealed dish with solid CNBr and 70% trifluoroacetic acid in separate vessels. After 5–12 hr at room temperature in the dark, residual CNBr/trifluoroacetic acid vapors were blown away with argon in the assembled sequencer cartridge and sequencing was resumed with the 03RPTH run. Although the Nelson analytical software (Paramus, NJ) was used for data reduction, the low signal level encountered required visual assessment of chromatograms.

cDNA Library Construction and Screening. A mouse thymus cDNA library in the λ ZAP vector was purchased from Stratagene. The library was made by oligo(dT) priming of thymus RNA isolated from 6- to 8-week-old B6/CBA F1J females. The library was screened with a probe isolated by using the polymerase chain reaction (PCR) to amplify a portion of the murine CD3 ζ coding region from total thymus RNA. The oligonucleotides used for this reaction were constructed from published sequences (4, 5). The sense oligonucleotide for the PCR was 5'-GAGCTCAATCTAGGGC-GAAG-3' and is encoded in exon 4. The antisense oligonucleotide was 5'-AGCGAGGGGCCAGGGTCTGC-3' and is encoded in exon 8. RNA from BALB/c murine thymus was reverse-transcribed using an oligo(dT) primer and reverse transcriptase and subsequently amplified for 40 cycles (Technique machine; denaturation for 1 min at 94°C, hybridization for 1 min at 55°C, extension for 0.8 min at 72°C, final extension for 10 min at 72°C) and the resultant 274-base-pair (bp) product was isolated by electrophoresis in a low-melting-point agarose gel and labeled by the random-hexamer priming method (13).

Filters were hybridized with probe at 42°C in 50% (vol/vol) formamide/5 $\times$ SSC/5 $\times$ Denhardt's solution/50 mM sodium phosphate, pH 6.5/0.1% NaDodSO₄ containing denatured salmon sperm DNA at 10 μ g/ml. (SSC is 0.15 M NaCl/0.015 M sodium citrate, pH 7.0; Denhardt's solution is 0.02% Ficoll/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin.) Washes were performed at room temperature in 2 $\times$ SSC/0.1% NaDodSO₄ for 30 min and at 65°C in 1 $\times$ SSC/0.1% NaDodSO₄ for 60–90 min. Positive phage clones were plaque-purified and small-scale preparations of each phage DNA were made. DNAs were digested with *Eco*RI, electrophoresed in agarose, and blotted to Zeta-Probe membranes (Bio-Rad) according to the manufacturer's recommendations. The membranes were separately hybridized, in 6 $\times$ SSC/5 $\times$ Denhardt's solution/0.5% NaDodSO₄ containing denatured salmon sperm DNA at 10 μ g/ml, to CD3 ζ -derived oligonucleotide 5'-AACCCCCAGGAAGGCGTATACAAT-3' (bp 415–438) at 64°C and to CD3 ζ -derived oligonucleotide 5'-CAGACCCTGGCCCCCTCGC-3' (bp 577–594) at 54°C (numbering according to ref. 4). Filters were washed at the hybridization temperature for 20 min in 6 $\times$ SSC/0.5% NaDodSO₄. All oligonucleotides were synthesized on an Applied Biosystems model 381A DNA synthesizer.

Synthetic RNA Production and *in Vitro* Translation. Plasmid pBS17 was linearized with *Xba* I, and capped, sense RNA

was synthesized *in vitro* by T7 RNA polymerase in the presence of G(5')ppp(5')G, using Riboprobe Gemini system reagents (Promega) according to the manufacturer's instructions. pGEM3Z- ζ was linearized with *Eco*RI and capped sense RNA was synthesized with SP6 RNA polymerase. The integrity of the RNA was checked by formaldehyde/Mops/agarose gel electrophoresis. Approximately 1 μ g of each RNA was translated *in vitro* using wheat germ extract according to the manufacturer's protocol. After 2 hr at 25°C, 10 μ l of the translation mixture was added to 40 μ l of 1 $\times$ reducing NaDodSO₄ sample buffer (10% glycerol/1% NaDodSO₄/1% 2-mercaptoethanol/0.001% bromophenol blue/62.5 mM Tris, pH 6.8) and heated in a boiling water bath for 3 min. [³⁵S]Methionine-labeled samples were then analyzed by NaDodSO₄/12.5% PAGE. Protein bands were visualized by autoradiography using Kodak X-Omat AR film for 12–14 hr at –70°C. Relative molecular weights of the resulting proteins were calculated using the SIZER program from the Ig Suite software package (IntelliGenetics) for comparison with known standards.

RESULTS AND DISCUSSION

Purification and Microsequencing of CD3 ζ and - η Proteins.

Anti-CD3 ϵ mAb (2C11) immunoprecipitates of surface-iodinated 3DO54.8 cells lysed in digitonin contain all the known components of the TCR complex. As shown in the autoradiograph of the two-dimensional nonreducing/reducing diagonal gel in Fig. 1, the Ti $\alpha\beta$ heterodimer at 40–60 kDa as well as CD3 γ , - δ , and - ϵ at 20, 25, and 26 kDa are readily detected. In addition, two other dimers are identified: a CD3 ζ - ζ homodimer (16-kDa subunit) and a CD3 ζ - η heterodimer (16-kDa and 22-kDa subunits). Consistent with previous reports, the frequency of incorporation of CD3 ζ into the homodimer is at least 1 order of magnitude greater than its incorporation into the ζ - η heterodimer (7). A third, 16-kDa off-diagonal spot may represent CD3 ζ disulfide-bonded to a 10-kDa subunit that is not retained in the gel. Note that the unlabeled spot at 26 kDa below the diagonal line represents a CD3 ϵ disulfide-linked homodimer of unknown significance (unpublished data).

To structurally characterize CD3 η , anti-CD3 ϵ mAb was used to batch-purify TCR subunits and the subunits were resolved into individual components by two-dimensional

FIG. 1. Two-dimensional diagonal NaDodSO₄/PAGE analysis of digitonin lysates of ¹²⁵I-labeled murine 3DO54.8 T cells immunoprecipitated by anti-CD3 ϵ mAb. Cells (2 $\times$ 10⁷) were labeled with 1 mCi of ¹²⁵I (Amersham) by lactoperoxidase-catalyzed surface iodination. Molecular mass markers are as follows: 97 kDa, phosphorylase b; 68 kDa, bovine serum albumin; 45 kDa, ovalbumin; 31 kDa, carbonic anhydrase; 22 kDa, soybean trypsin inhibitor; and 14 kDa, lysozyme.

diagonal gel electrophoresis. After blotting to PVDF and Coomassie staining, CD3 η and - ζ spots were cut out and loaded onto a gas-phase protein sequenator. Both NH₂ termini were found to be blocked. Consequently, CNBr fragmentation was performed. Table 1 lists the residues observed (italic type) in individual successive cycles relative to the expected residues derived from the CD3 ζ cDNA sequence where cleavage should occur after methionine residues at positions 75, 98, and 137 of the mature protein. Although the observed signals were all at the level of 1 pmol, residues in appropriate positions of each of the three corresponding predicted fragments were detected in the empirically derived CD3 ζ CNBr fragments. Parallel sequence analysis of an equivalent amount of CD3 η identified residues corresponding to known positions of the CD3 ζ CNBr fragments 1 and 2. This result unequivocally demonstrates a structural relatedness of CD3 ζ and - η polypeptides and is consistent with earlier results obtained using anti-peptide heteroantisera, which documented crossreactivity of anti-CD3 ζ peptide sera with CD3 η (10). Perhaps more importantly, in contrast to results obtained with CD3 ζ , Table 1 shows that the CD3 η sequence lacked any residues corresponding to CD3 ζ fragment 3.

Isolation of CD3 η cDNA. From the above results, we concluded that CD3 η and - ζ are related proteins but that CD3 η structure is distinct from CD3 ζ within its COOH-terminal region. To clone CD3 η , we first utilized a 274-bp CD3 ζ cDNA, derived by PCR as described in *Materials and Methods*, to screen a murine thymocyte cDNA library and identified 27 positive clones. After plaque purification, inserts from each of these clones were hybridized to a CD3 ζ oligonucleotide corresponding to bp 415–438 [amino acids (aa) 84–91] and an oligonucleotide corresponding to bp 577–594 (aa 138–143). Based on the above protein sequence data, we hypothesized that the former oligonucleotide would hybridize to both CD3 ζ and - η clones, whereas the latter oligonucleotide should hybridize exclusively to CD3 ζ clones. Inserts from 26 of 27 clones hybridized with both oligonucleotides whereas one clone, λ ZAP17, hybridized with oligonucleotide 415–438 but not 577–594. This potential CD3 η clone and the largest CD3 ζ inserts were subcloned from the λ ZAP vector by using the *in vivo* excision protocol (Stratagene). Plasmid DNA, termed pBS17, was prepared and sequenced on both strands by the method of Sanger *et al.* (14) with sequence-specific oligonucleotides as primers. As shown in Fig. 2, the complete nucleotide sequence of the pBS17 insert is 1362 bp long and contains a poly(A) sequence at its 3' end. An open reading frame of 618 bases (positions 139–756) begins with an ATG methionine codon and is flanked by 138 nucleotides of 5' untranslated sequence and 591 nucleotides of 3' untranslated sequence. A polyadeny-

lation signal (AATAAA) is located at bp 1329–1334 and followed by a poly(A) tail beginning at bp 1348.

The experimentally derived residues of the CNBr fragment of CD3 η are encoded in the pBS17 insert at nucleotide positions 430–477 and 502–531. All 15 residues identified by protein sequencing match the cDNA sequence precisely. Furthermore, as expected, each of these sequences is preceded by an ATG codon for methionine. The first 21 aa encoded by the open reading frame represent hydrophobic residues and probably constitute a leader peptide. Based on the method of von Heijne (15), the signal-peptide cleavage site is located NH₂-terminal to the glutamine residue encoded by nucleotides 202–204. This results in a mature protein of 185 aa with a predicted molecular weight of 21,001 and theoretical pI of 9.75. This molecular weight is in excellent agreement with results of NaDodSO₄/PAGE analysis of CD3 η . Within the mature protein, there is a single hydrophobic stretch (residues +10 to +30) most likely corresponding to a membrane-anchoring segment. This sequence is unusual in the presence of aspartic acid at +15. However, a single negatively charged amino acid has been noted in the transmembrane domain of each of the other CD3 subunits (2) and presumably forms a critical salt bridge with a positively charged residue in the Ti α or β subunit. Thus, the protein encoded by pBS17 is a type I integral membrane protein consisting of a 9-aa extracellular segment, a 21-aa transmembrane segment, and a 155-aa cytoplasmic segment. By analogy with CD3 ζ - ζ homodimers, the cysteine residue at +11 most likely forms the intrachain disulfide bond between CD3 η and CD3 ζ . The role of the cysteine residue at +164 is unknown. No N-linked glycosylation sites are present in the predicted CD3 η sequence.

Relatedness of CD3 η and CD3 ζ Sequences. Comparison of the predicted mature murine CD3 η and - ζ sequences by gene matrix analysis (Fig. 3) shows an identity of the NH₂-terminal 122 aa of the two proteins. In addition, the leader sequences are identical (data not shown). However, the COOH-terminal sequences of CD3 ζ and CD3 η are entirely divergent. Thus, the two proteins are identical in the extracellular and transmembrane segments but diverge in the cytoplasmic tails, resulting in CD3 ζ and - η cytoplasmic domains of 113 and 155 aa, respectively. Comparison of the amino acid residues in the predicted CD3 ζ and - η sequences agrees well with empirically derived amino acid analysis using metabolically labeled cell lysates (10), particularly in the presence of two tryptophan residues in CD3 η that are lacking from CD3 ζ and an additional cysteine residue in the CD3 η sequence. Furthermore, rabbit antiserum raised against a peptide corresponding to the predicted CD3 η COOH terminus specifically identified the CD3 η protein in Western blots of two-dimensional diagonal gels of anti-CD3 η immunoprecipitates,

Table 1. Microsequencing analysis of CNBr fragments of murine CD3 ζ and CD3 η proteins derived from preparative PVDF blots of two-dimensional gels

Protein	Fragment (positions)	Sequence*
CD3 ζ	1 (76–92)	¹ G ⁵ G K Q ¹⁰ R R R N ¹⁵ P Q E G V Y N A
	2 (99–111)	A E A Y S E I G T K G E R
	3 (138–143)	Q T L A P R
CD3 η	1	G G K Q Q R R R N P Q E G V Y N A
	2	A E A Y S E I G T K G E R
	3	Q T L A P R
	Unassigned residues	P V G A L L

*Cycle numbers are indicated above the sequences. Residues in italic type are those that were judged to be significantly above background. Certain residues such as lysine, arginine, serine, and threonine (K, R, S, and T) are not discernible at this low level of sequencing due to poor recovery. Unassigned residues were found in η only.

10 20 30 40 50 60 70
CGAGGAGG CACCCGGTAC ATGTTCTCTA GGAGAACAA GCCAGCAG ACTCCATCAG CGCCTCCTTT

80 90 100 110 120 130
TCTCTCATC CTCCAGGCA TAGCTGCTC TGCTCTGCC TCTGGTACC ATCCAGGGA AGCAGAAG

141 150 159 168 177 186
ATG AAG TGG AAA GTG TCT GTT CTC GCC TGG ATC CTC CAC GTG CGG TTC CCA GGA
MET Lys Trp Val Ser Val Leu Ala Cys Ile Leu His Val Arg Phe Pro Gly - 4

195 204 213 222 231 240
GCA GAG GCA CAG AGC TTT GGT CTG CTG GAT CCC AAA CTC TGC TAC TTG CTA GAT
Ala Glu Ala Gln Ser Phe Gly Leu Leu Asp Pro Lys Leu Cys Tyr Leu Leu Asp 15

249 258 267 276 285 294
GGA ATC CTC TTC ATC TAC GTC ATC ATC ACA GCC CTG TAC CTG AGA GCA AAA
Gly Ile Leu Phe Ile Tyr Gly Val Ile Ile Thr Ala Leu Tyr Leu Arg Ala Lys 33

303 312 321 330 339 348
TTC AGC AGG ACT GCA GAG ACT GCT GCC AAC CTG CAG CCC AAC CAG CTC TAC
Phe Ser Arg Ser Ala Glu Thr Ala Ala Asn Leu Gln Asp Pro Asn Gln Leu Tyr 51

357 366 375 384 393 402
AAT GAG CTC AAT CTA GGG CGA AGA GAG GAA TAT GTC TTG GAG AAG AAG CGG
Asn Glu Leu Asn Leu Gly Arg Arg Glu Tyr Asp Val Leu Glu Lys Lys Arg 69

411 420 429 438 447 456
GCT CGG GAT CCA GAG ATG GGA GGC AAA CAG CAG AGG AGG AAC CCC CAG GAA
Ala Arg Asp Pro Glu MET Gly Lys Gln Gln Arg Arg Arg Asn Pro Gln Glu 87

465 474 483 492 501 510
GGC GTA TAC AAT GCA CTG CAG AAA GAC AAG ATG GCA GAA GGC TAC AGT GAG ATC
Gly Val Tyr Asn Ala Leu Gln Lys Asp Lys MET Ala Glu Ala Tyr Ser Glu Ile 105

519 528 537 546 555 564
GGC ACA AAA GGC GAG AGG CGG AGA GGC AAG GGC CAC GAT GGC CTT TAC CAG GAC
Gly Thr Lys Gly Glu Arg Arg Arg Gly Lys Gly His Asp Gly Leu Tyr Gln Asp 123

573 582 591 600 609 618
AGC CAC TTC CAA GCA GTG CAG TTC GGG AAC AGA AGA GAG GAA GGT TCA GAA
Ser His Phe Gln Ala Val Gln Phe Gly Asn Arg Arg Glu Arg Glu Gly Ser Glu 141

627 636 645 654 663 672
CTC ACA AGG ACC CTT GGG TTA AGA GCC CGC CCC AAA GGT GAA AGC ACC CAG CAG
Leu Thr Arg Thr Leu Gly Leu Arg Ala Arg Pro Lys Gly Glu Ser Thr Gln Gln 159

681 690 699 708 717 726
AGT AGC CAA TCC TGT GGC AGC GTC TTC AGC ATC CCC ACT CTG TGG AGT CCA TGG
Ser Ser Gln Ser Cys Ala Ser Val Phe Ser Ile Pro Thr Leu Trp Ser Pro Trp 177

735 744 753 762 771 780 796
GCA CCC AGT AGC AGC TCC CAG CTC TAAGGCCCTG TGCTCAGCTC TGGTGATGAC CCGTGGCTGT
Pro Pro Ser Ser Ser Gln Leu 185

806 816 826 836 846 856 866
GTCCATGAG TTGTGGTGAG GATGGGACTT TTGAAAAATC TGATGTCCCA ATTTCTTTCA TGCACTCT

876 886 896 906 916 926 936
ACTCAGAACT GAGCAAGGGC CAAAACTCTT GGGCATGAT CTGAATGAAT CTCTCAACTT AGAAAAACT

946 956 966 976 986 996 1006
TGCTGCCCC TCTGAGCTGG CCAAGTGCTC CCACCTACCC TTTGGCATGC CTCGAAGTGC CAGGACGCCA

1016 1026 1036 1046 1056 1066 1076
CAGACTGCAT TGGGGCCGAA GACTTCCCTT TTCCTTGGTT TTTTGTGTTT TTTTGTGTTT ATACAATCAT

1086 1096 1106 1116 1126 1136 1146
TAAGAAATCT TTGGTTTGG CTGGAATGG AAAAAACAAA CAAATCAAG AAACCAACCT CCCTGGCTT

1156 1166 1176 1186 1196 1206 1216
ATAGACGAG TATTATGACC TGACCTGGCT GAGCTTTCCC CCCCTCAAC TTTGGGGGTG GAAATTCGAA

1226 1236 1246 1256 1266 1276 1286
GTTAAGAACT ACATTCAAG AAAACGTTGA AAGGCCGGA GAAGCAGCTT CCAGAAAGCC CGTCGGATAT

1296 1306 1316 1326 1336 1346 1356
AAGATTGCA AATAATAATA ACTATTATTA TAATATATTA ATAATAAACT TAAATATTTG GAAAAAATAA...

FIG. 2. cDNA and predicted protein sequence of clone pBS17 insert. The positions of the predicted signal peptidase cleavage site ($\downarrow$), transmembrane region (overbar), CNBr fragmentation sites ($\downarrow$), and polyadenylation signal (double underline) are shown. Note that the last 6 adenine residues of the insert's poly(A) tail are not shown. The CD3 η nucleotide sequence between the solid arrowheads (bp 35–568) is identical with that of CD3 ζ (5).

proving the authenticity of the pBS17 clone (data not shown). Computer search of GenBank 61, EMBL 20, PIR 21.0, and Swiss-Prot 11 data bases utilizing the CD3 η -specific DNA and protein sequences failed to uncover any significant matches with known structures.

Eta QSFGLDPKLCYLLDGILFIYGVIIITALYLRKFSRSAETAANLQDPNQL 50
Zeta QSFGLDPKLCYLLDGILFIYGVIIITALYLRKFSRSAETAANLQDPNQL 50

Eta YNELNLGRREEYDVLEKKRARDPEMGKKQRRRNPEQGVYNALQDKMAE 100
Zeta YNELNLGRREEYDVLEKKRARDPEMGKKQRRRNPEQGVYNALQDKMAE 100

Eta AYSEIGTKGERRRGKHDGLYQDSHFQAVQFQGNRRERSELTRTLGLRA 150
Zeta AYSEIGTKGERRRGKHDGLYQGLSTATKDTYD----- 133

Eta RPKGESTQSSQSCASVFSIPTLWSPWPSSSSQL 185
Zeta -----ALHMQTLAPR 143

FIG. 3. Protein sequence alignment of mature murine CD3 ζ and - η by genetic code matrix analysis. Open gap cost and unit gap cost were each set at 100. Identity value = 85.3%. Identity of amino acids in the two sequences is marked. Beyond aa 122 the sequences are divergent.

Given the NH₂-terminal identity and COOH-terminal divergence of CD3 ζ and - η sequences, it is likely that CD3 η results from an alternatively spliced transcript of a gene common to both CD3 η and CD3 ζ . The region of CD3 η divergence from the published CD3 ζ DNA sequence follows that segment encoded by exon 7 at bp 568, which is 1 bp beyond the splice junction at bp 567. An identical junctional sequence was found in an independent clone identified by rescreening the library with a CD3 η -specific probe (bp 633–1362, pBS17; data not shown). It is therefore likely that CD3 ζ exon 8 has been spliced out from the pre-mRNA and replaced by a downstream exon encoding the unique CD3 η sequence that also has a guanine nucleotide at the 5' end of the exon. The CD3 η exon does not appear to be within the intron between CD3 ζ exons 7 and 8, because a PCR product derived from priming DNA with amplimers corresponding to CD3 ζ exon 7 (sense) and exon 8 (antisense) sequences did not hybridize with a CD3 η -specific cDNA clone (data not shown). There is also the possibility of 5' alternative splicing, given the divergence between CD3 ζ and - η nucleotide sequences noted in the 5' untranslated region of the CD3 η cDNA 5' of bp 35 (Fig. 2) and the published flanking sequence of the murine genomic CD3 ζ clone (5). For these reasons, we consider the alternative possibility that CD3 η and CD3 ζ arose from gene duplication more remote, although not yet formally excluded.

RNA blot analysis of 2-week-old BALB/c thymocytes and the T-T hybridoma 3DO54.8 with a CD3 η -specific probe (bp 633–1362, pBS17) readily identified a 1.7-kilobase CD3 η mRNA. This message was not detected in mouse liver or in B-cell lines A20 and M12 (data not shown). That CD3 ζ mRNA is also 1.7 kilobases in size readily explains the lack of detectable CD3 η mRNA with the related CD3 ζ probe.

In Vitro Translation of Synthetic Murine CD3 η RNA. Analysis of the CD3 η cDNA open reading frame predicts a molecular weight of 23,339 for pre-CD3 η . *In vitro* transcription of the pBS17 cDNA insert with T7 RNA polymerase yielded synthetic RNA with an apparent size of $\approx$ 1.5 kilobases in formaldehyde/Mops/1.2% agarose gel electrophoresis (data not shown). Translation of this RNA *in vitro* resulted in a major protein product of 23 kDa (Fig. 4, lane 2). The size of this polypeptide is in excellent agreement with the predicted molecular weight of pre-CD3 η . The minor bands at 24 and 16 kDa may have resulted from translation of endogenous RNA or premature termination of RNA translation (16). Synthetic RNA from the human CD3 ζ -encoding clone pGEM3Z-CD3 ζ produced a major protein product of 17.8 kDa (Fig. 4, lane 1), in close agreement with the predicted molecular weight of 18,608 for pre-CD3 ζ , while bovine papilloma virus RNA gave the expected pattern (lane 4). Translation in the absence of added RNA gave no detectable protein (lane 3).

FIG. 4. *In vitro* translation of CD3 η . Synthetic RNA ($\approx 1 \mu\text{g}$) was translated using wheat germ lysate and analyzed by reducing NaDodSO₄/PAGE as described. Lane 1, human CD3 ζ ; lane 2, murine CD3 η ; lane 3, no added RNA; lane 4, bovine papilloma virus RNA. Molecular mass standards are as in Fig. 1.

Implications. The important contribution of CD3 ζ to TCR-mediated activation has been demonstrated by analysis of CD3 ζ mutants derived from antigen-specific MHC-restricted T-T hybridomas (6). Loss of CD3 ζ expression leads to virtual abrogation of interleukin 2 production upon stimulation by antigen and MHC product. The role of CD3 η is less certain. However, analysis of cellular variants expressing different levels of CD3 η has led to the notion that CD3 η expression may be important and perhaps necessary for inositolphospholipid turnover (8). Furthermore, cell-cycle arrest of T-cell hybridomas upon TCR crosslinking is linked to the level of CD3 η expression (9).

The cloning of CD3 η as reported here provides a definitive structural basis for biologically distinct differences between CD3 η and CD3 ζ function. CD3 ζ possesses six cytoplasmic tyrosine residues whereas CD3 η has only five (Fig. 5). Given that CD3 ζ , but not CD3 η , has been reported to be phosphorylated on tyrosine residues after TCR triggering (17), the tyrosine at position 132 may have a regulatory role. Others have construed that the sequence Gly-Xaa-Gly-Xaa₂-Gly-Xaa₃-Gly-Xaa₃-Ala-Xaa-Lys within the CD3 ζ sequence (aa 114–129) may represent a nucleotide binding site (4).

FIG. 5. Model of putative CD3 ζ - η heterodimers and CD3 ζ - ζ and CD3 η - η homodimers. The single negatively charged (–) aspartic residue (at aa +15) in the transmembrane region is shown. Tyrosine residues (●) are indicated as is the putative nucleotide binding site in the cytoplasmic domain (□). CD3 ζ - and CD3 η -specific protein stretches are highlighted by solid and hatched lines, respectively.

However, nucleotide binding sites have been shown to occur at the junction between a β -sheet and an α -helical segment and, as such, are preceded by and include hydrophobic residues (18, 19). This is not the case for CD3 ζ , where the putative site is preceded by the four charged residues Glu-Arg-Arg-Arg. Nevertheless, if a di- or trinucleotide binding site exists within the CD3 ζ sequence at positions 114–129, it is almost certain to be altered in CD3 η since that putative motif has been disrupted within the CD3 η sequence after the third glycine residue. How this difference will affect the functions of CD3 η and - ζ remains to be shown, but it could result in CD3 ζ and - η being differentially regulated by exogenous signals and/or themselves mediating different functional effects.

Fig. 5 also raises the possibility that three dimeric forms of CD3 ζ and/or - η may exist: CD3 ζ - ζ homodimers, CD3 ζ - η heterodimers, and CD3 η - η homodimers. These CD3 ζ - ζ , CD3 ζ - η , or CD3 η - η dimers could be differentially expressed during thymocyte development. That the η - η homodimer has not been previously observed may be a consequence of the low level of CD3 η expression and the lack of CD3 η -specific reagents. Analysis of murine CD3 η and - ζ RNA expression during gestation, as well as use of CD3 η -specific antibodies to examine protein levels in individual cells, will help resolve these issues. One must seriously entertain the possibility that either CD3 η - η or CD3 ζ - η dimers may be responsible for linking TCRs to different cellular pathways than does CD3 ζ - ζ . Perhaps programmed cell death of thymocytes resulting from TCR crosslinking is mediated by CD3 η . If this is the case, CD3 η will have an important role in selection of thymocytes.

We thank Dr. Luciano D'Adamio for helpful discussion and PCR expertise, Dr. Philippe Moingeon for assistance with cell culture and purification of 2C11 mAb, and Rebecca Hussey for excellent technical assistance. This work was supported in part by National Institutes of Health Grants AI21227 and AI19807. Y.-J.J. is supported by a fellowship from the Cancer Research Institute.

- Meuer, S. C., Acuto, O., Hercend, T., Schlossman, S. F. & Reinherz, E. L. (1984) *Annu. Rev. Immunol.* **2**, 23–50.
- Clevers, H., Alarcon, B., Wileman, T. & Terhorst, C. (1988) *Annu. Rev. Immunol.* **6**, 629–662.
- Davis, M. M. & Bjorkman, P. J. (1988) *Nature (London)* **334**, 395–402.
- Weissman, A. M., Baniyash, M., Hou, D., Samelson, L. E., Burgess, W. H. & Klausner, R. D. (1988) *Science* **239**, 1018–1021.
- Baniyash, M., Hsu, V. W., Seldin, M. F. & Klausner, R. D. (1989) *J. Biol. Chem.* **264**, 13252–13257.
- Sussman, J. J., Bonifacio, J. S., Lippincott-Schwartz, J., Weissman, A. M., Saito, T., Klausner, R. D. & Ashwell, J. D. (1988) *Cell* **52**, 85–95.
- Baniyash, M., Garcia-Morales, P., Bonifacio, J. S., Samelson, L. E. & Klausner, R. D. (1988) *J. Biol. Chem.* **263**, 9874–9878.
- Merčep, M., Bonifacio, J. S., Garcia-Morales, P., Samelson, L. E., Klausner, R. D. & Ashwell, J. D. (1988) *Science* **242**, 571–574.
- Merčep, M., Weissman, A. M., Frank, S. J., Klausner, R. D. & Ashwell, J. D. (1989) *Science* **246**, 1162–1165.
- Orloff, D. G., Frank, S. J., Robey, F. A., Weissman, A. M. & Klausner, R. D. (1989) *J. Biol. Chem.* **264**, 14812–14817.
- Haskins, K., Kubo, R., White, J., Pigeon, M., Kappler, J. & Marrack, P. (1983) *J. Exp. Med.* **157**, 1149–1169.
- Hewick, R. M., Hunkapiller, M. W., Hood, L. E. & Dryer, W. J. (1981) *J. Biol. Chem.* **256**, 7990–7997.
- Feinberg, A. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690.
- Hames, B. D. & Higgins, S. J., eds. (1986) *Transcription and Translation: A Practical Approach* (IRL, Oxford).
- Baniyash, M., Garcia-Morales, P., Luong, E., Samelson, L. E. & Klausner, R. D. (1988) *J. Biol. Chem.* **263**, 18225–18230.
- Wierenga, R. K. & Hol, W. G. J. (1983) *Nature (London)* **302**, 842–844.
- Moller, W. & Amons, R. (1985) *FEBS Lett.* **186**, 1–6.