

Physical linkage of three CD3 genes on human chromosome 11

A. Tunnacliffe, L. Buluwela and T.H. Rabbitts

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Communicated by T.H. Rabbitts

T-cell antigen receptors are associated on T cell surfaces with a complex of proteins called CD3 (formerly T3). Human CD3 consists of at least four proteins, γ , δ , ϵ and ζ , and all but the latter have been cloned as cDNA. Using standard cloning techniques, together with field inversion gel electrophoresis, we have demonstrated the physical linkage of three CD3 genes. The genes for CD3 γ and CD3 δ are situated close together, about 1.6 kb apart, organized in a head-to-head orientation. The gene encoding CD3 γ has been sequenced, and is split into seven exons spread over 9 kb of DNA. Like CD3 δ , the CD3 γ gene has an unusual promoter which lacks a TATA-box and potential Sp1 binding sites. The CD3 γ –CD3 δ gene pair is within 300 kb of the CD3 ϵ gene, and therefore these genes form a tightly linked cluster in chromosome 11 band q23. The clustering of the CD3 genes may be significant in terms of their simultaneous activation during T-cell development.

Key words: CD3 γ gene/nucleotide sequence/T-cell antigen receptor/field inversion gel electrophoresis/promoter/gene cluster

Introduction

T-cell antigen receptors (TCRs) are responsible for antigen recognition in the vertebrate cellular immune response. The most common form of human TCR consists of two disulphide-linked glycoproteins, TCR α and TCR β , and a complex of proteins called CD3 (Meuer *et al.*, 1984; Reinherz, 1987). The α – β heterodimer binds antigen in association with major histocompatibility complex (MHC) proteins on host cell surfaces, but the role of CD3 is less clear. One possibility is that it acts as a signal transduction system, triggering T-cell activation subsequent to antigen binding (Tsoukas *et al.*, 1984). CD3 comprises at least four proteins (Borst *et al.*, 1983; Kannelopoulos *et al.*, 1983; Weissman *et al.*, 1986), γ , δ , ϵ and ζ , three of which have been cloned as cDNA (Krissansen *et al.*, 1986; van den Elsen *et al.*, 1984; Rabbitts *et al.*, 1985; Gold *et al.*, 1986). cDNA clones for CD3 δ and CD3 ϵ have been used to map their respective genes to band q23 of human chromosome 11 (Gold *et al.*, 1987). The coordinate activation of these genes, and probably those encoding the other CD3 proteins, is an early event during T cell ontogeny in the thymus (Furley *et al.*, 1986; Campana *et al.*, 1987; van Dongen *et al.*, 1987). Information on how this regulation is achieved may be derived from studying the structure and relative location of the genes involved. In this paper, we show that the CD3 γ and CD3 δ genes are situated about 1.6 kb apart, in a head-to-head orientation. Furthermore, the CD3 γ –CD3 δ gene pair is less than 300 kb from the CD3 ϵ gene, and therefore these genes are tightly linked on the chromosome.

Results

CD3 γ and CD3 δ genes are closely linked in a head-to-head orientation

Since it has been shown (Krissansen *et al.*, 1986) that cDNA sequences for CD3 γ and CD3 δ exhibit considerable homology, it seemed possible that their respective genes might be closely linked. We had previously isolated the human CD3 δ gene on two overlapping recombinant phages, $\lambda\delta 3A$ and $\lambda\delta 13$ (Tunnacliffe *et al.*, 1986), and therefore tested for CD3 γ gene sequences in these clones, using a CD3 γ cDNA clone pJ6T3 γ -2 (Krissansen *et al.*, 1986) as probe. Although no signal was seen with $\lambda\delta 3A$, the CD3 γ probe hybridized to regions of the $\lambda\delta 13$ insert. Preliminary sequencing of the hybridizing DNA showed that the CD3 γ gene lies upstream of CD3 δ , and is transcribed in the opposite direction, the two genes being situated head-to-head on the chromosome. A map of the CD3 γ and CD3 δ genes is given in Figure 1.

It was clear from initial experiments that the whole CD3 γ gene was not contained within $\lambda\delta 13$. Therefore, a phage λ library made from B-cell DNA was screened with pJ6T3 γ -2 and four hybridizing clones isolated (Figure 1). One of these, $\lambda\gamma 7$, contained all the genomic sequence corresponding to that of the cDNA clones previously isolated (Krissansen *et al.*, 1986), although extending only 11 bp beyond the 3' end of the cDNA sequence (see below). [Some T cells express a long form of CD3 γ mRNA (3.5 kb in contrast to the shorter 0.8 kb form). The extra length of RNA is thought to consist of a longer 3'-untranslated sequence (Krissansen *et al.*, 1986). Thus, while this sequence has not been cloned as cDNA and is not contained in our genomic clones, all CD3 γ coding sequences are within the limits of $\lambda\gamma 7$.]

Characterization of the CD3 γ gene

Nucleotide sequencing showed that the CD3 γ gene consists of at least seven exons (Figure 2), spread over 9 kb, whose positions are shown in Figure 1. The first exon, 132 bp long, includes the 5'-untranslated region and most of the signal peptide which is split at the first position of codon –4 (numbering according

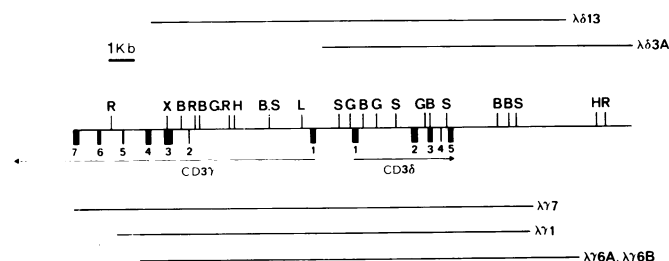


Fig. 1. Restriction map of the region of chromosome 11q23 carrying the genes from CD3 γ and CD3 δ . The extent of inserts of recombinant phage $\lambda\delta 3A$, $\lambda\delta 13$ (Tunnacliffe *et al.*, 1986), $\lambda\gamma 7$, $\lambda\gamma 1$, $\lambda\gamma 6A$ and $\lambda\gamma 6B$ (this paper) covering the region is shown. Restriction enzymes used were: B, *Bam*HI; G, *Bgl*III; H, *Hind*III; R, *Eco*RI; S, *Sac*I; L, *Sal*I; X, *Xho*I. Exons for each gene are numbered and shown as black boxes, with directions of transcription indicated by an arrow.

```

<Bg12>
AGATCTAGCAGCTTCTTGTCACTGGGAGGTTGGGCTTTAGAGACCCAGCCAGAGATTTGAATCCTGGGTCCAATACTGCCTACCTGTGGGGCTGGGCCAGCCATAAAA 110
TTTTTCAGAGTCTTATTCATTAGTACCATTATTAGGATTCACAAACAGGATTTTGCATGGTGCCTCAGGCATCATATGTGCTCATTAAAGGGGTAGTTATTAAATAATAA 220
TAATTGACTGACAGGCAATTTGAGCCTCCCGGTGAGACAAATGGACCTTTTCCCTGTGGCTACGAGGATCTGAAACTCTTCACGCTCGCAGTTAGACTGTCACT 330
TACCTGGGACAGAGTCACTGCTGTCTGTCTACCTGCTGTATCTTGTGCTGGCAGACATAACGGGAGCTCTGCACATTTTTGTGGCTCACTGACTGACTGGCTGAGGGAG 440
ATAGGGCTGAGATCTGGACATTCAGTCCGGCTCTGGCCCCGAAAATGTGCTGGCTGTCTCGGAATTTGCCACCTATTGCCCTCCAGGGCCCTTTTCATGAT 550

( Alu 5'
CTCAAAAGAATAGTGAACACAGGTGCCGTGTCTCACGCTGTAATCCCAACACTTTGGGAGGCTGAGGCAAGGTGGATCACAAAGTTCAGGAGTTCGAGACCAGCCTGACCA 660
ACAAGGTGAACCCCGTCTCTACTAAAATACAAAATTAGCCGGGCTGTGCTGGCACGCACCTGTAATCTCAGTACTCAGGAGGCTGAGGCGAGGAAATCGCTTGAACC 770

Alu 3' >
CCGGAGGTGGAGGTTGCAGTGAGCTGAGATAGCCCACTGCCTCCAGCCTGGGCAACAAAGTGAAGTCTGTCTCAAAAAGAAAAGAAAAAAGTGAAAAAAATT 880
CCTGAATGAAGGCTGGACTGAGGTGGCTTTCCATTTGGAGGTCACGCCCAAGCATCTGAGAGTCCCTCCTAAATTCATTACCTACAGCAACAACACTAGCAACAAGT 990
AACAACTGGCTACGATCCTAACAACTAATGACAGGGACATTTATCTCCCTCATGAAGAAACGGTCCAGGACCATCTCCACCCAGCATCCATTGCGGTTCCTGTGCAA 1100
GATGAGTCTCTGAGTGGAAATCCAGCACTCTCTCCCTCTCTCCACCACCTTACCCTCCTTAACGGAAAAACAAGGCATCTGCACCTGCACCCCTGCTGAGGCC 1210

Exon 1 :+ + + + +
CCTGTGCTCACACTTGCAGCAGAGGGTGGAGGCTCTGGGTTCTTGCCTTCTCTCAAGGCCCCAGCCCCAACAGTGAAGTGGGAGGCACTGACTGCTGCACAGGCT 1320
GGCTGGCTGGCTGGCTGCTAAGGGCTGCTCCACGCTTTTGGCCGGAGGACAGAGACTGACATGGAACAGGGGAAGGGCTGGCTGTCTCATCTGCTATCCTGCTATCTTCTT 1430
< T61 >
CAAGGTAAGGGCTACTAGGGTCTGGAAGCCTGGGAAGGGCTCAAGGGAAGAGCCCATCACTAGTGAGACAGGAATATTGGTATCCCTAACCTTCAGCCTACCTCTGC 1540
TGTCACCTTAGAGTTCAGAAAGAGGGCAAAATGGAGGCTCTTAACGTCTCTGCTAGAGAGAAACAGTGTCCATGGAGGAGAAGGAATCCTTGTCTCTGAAAAATGCAA 1650
ACAGAGTACTTAATG6CTGAAGAGAGGACCCCTGTTACC6CATCTTAGATTTGAATGACGCCCAAAAGGGCATAGGCCAAGAACTAAAAGGAAAAAGTATATGTTCC 1760

<Ss11>
TACTTCAGAGCTGGGGGCTAGCAGTGCAC<--- * 4.1 kb --->ATCTGAACAACCTCCATGCCAGCTAATACTCTATCTCTTCTTCTGTCTTACAGTACTTTG 1870

-1 +1
A R S I K Exon 3 :G N H L V K V Y
GCCCAGTCAATCAAAGGTAAGGAGAAATGGCTTCTTCTAT<---- * 380 bp --->ATAGAACCAGGCTTTTCTCATTTCAGGAAACCACTTGGTTAAGGGTGTAT 1980
D Y Q E D G S V L L T C D A E A K N I T W F K D G K M I G F L T E D K K K
GACTATCAAGAAGATGGTTCGGTACTTCTGACTTGTGATGCAGAAGCCAAAATATCACATGGTTTAAAGATGGGAAGATGATCGGCTTCTAACTGAAGATAAAAAA 2090
W N L G S N A K D P R G M Y Q C K G S Q N K S K P L Q V Y Y R
ATGGAATCTGGGAAGTAAATGCCAAGGACCTCGAGGGATGTATCAGTGAAGGATCAGAGCAAGATCAAAACCACTCCAAGTGTATTACAGAAGTATGTAATCCCTT 2200
TG6TCTGTTTGTGAAAT<---- 467 bp --->ATTGCAGACAGGCAAGGAGAAACGAACCAAGGAAAAACAACCTTTCGCAACCTGAAAGGTTTGTCTCTCTT 2310

Exon 4 :M C Q N C I E L N A A T I S G F L F A E I V S I F V L A V G V Y F I
TCCCTACAGTGTGTCAGAACTGCATTGAACATAATGACGCCACCATATCTGGCTTCTCTTTGCTGAAATCGTCAGCATTTTCGCTCTGCTGTTGGGGTCTACTTCATT 2420
A G Q D G V R Q S R
GCTGGACAGGATGAGGTTCCGACGTCGAGAGGTAAGAAAGTGTCTTAGATGAGAGATGGGACCACCTGAGACCCTCAGCTTTCTCTACCAT<--- * 1 kb -- 2530

Exon 5 :A S D K Q T L L P N D Q L Y Q
--->ATTCTGTTCTTTTTTGTGCAAGCTTCAGACAAGCAGACTCTGTTGCCCAATGACCAGCTCTACAGGTAAGGGATGAAGAATAAAGAGACAT<--- 816 bp 2640

Exon 6 : P L K D R E D D Q Y S H L Q G N
----->ATTAATTAATAGAGGATGGAAAAAATGACTTATGACTGTGCTGCTTCCAGCCCTCAAGGATCGAGAAGATGACAGTACAGCCACCTCAAGGAAACC 2750
Q L R R N *
AGTTGAGGAGGAATGAACCTCAGGACTCAGAGTAGGTTGGGTTCTTCAATGCCAATTCTAAT<---- 631 bp --->ATTATTTGTTTTTTTTGAACAAATGCAA 2860

Exon 7 :
TTTTTCTTTTTTCAAGTCCAGGTTCTCTCTCTATTCACTCCAGAAATCAAGCAATGCATTTTGGAAAGCTCCTAGCAGAGAGACTTTCAGCCCTAAATCTAGACTCA 2970
AGGTTCCAGAGATGACAAATGGAGAGAAAGGCCATCAGAGCAAAATTTGGGGTTTCTCAAAATAAAAATAAAAACAATACTGTGTTTCAGAAAGGCCACCTAT 3080
? ]
TGGGAAAATTGTAAGAAAGAAATGAAAAGATC

```

Fig. 2. Sequence of the CD3 γ gene promoter and exons. Nucleotide sequence is shown, together with the predicted amino acid sequences (in one-letter code) of the six coding exons. The boundaries of sequences matching the CD3 γ cDNA sequence (Krissansen *et al.*, 1986), are indicated by square brackets, and the two differences between genomic and cDNA sequences are indicated by question marks: the G in the 5'-untranslated region is not seen in the cDNA sequence, whereas the G residue in the 3'-non-coding region is an A in cDNA. Donor and acceptor splice sites are underlined and delineate exon-intron boundaries. All conform to the GT-AG rule (Breathnach and Chambon, 1981). The position of the sequence in exon 1 which is complementary to the 22-base oligonucleotide TG1 used for S1 mapping is shown. Plus signs highlight transcription start sites. The numbers -1 and +1 above the second exon-encoded protein sequence indicate the C-terminal end of the signal peptide, and the first residue of the mature CD3 γ protein, respectively. The position of an Alu-type repeat element, 5' of the CD3 γ gene, is shown. The multiple poly(A)-addition signals which give rise to the 0.8 kb form of CD3 γ mRNA are overlined. The last four nucleotides, GATC, form the cloning site of $\lambda\gamma 7$. The BgIII site at the 5' end of the 1.8 kb sequence is identical to that at the end of the CD3 δ sequence previously published (Tunnacliffe *et al.*, 1986), and therefore the whole CD3 γ -CD3 δ intergenic sequence is known.

to Krissansen *et al.*, 1986). There follows a long first intron of ~4.5 kb before the second exon, of only 24 bp, which encodes the remainder of the signal peptide and the N-terminal four amino acids of the mature CD3 γ protein. The third exon (207 bp in length) encodes the majority of the extracellular domain of CD3 γ , while the proposed hydrophobic domain is coded for by the fourth exon (122 bp long), which also carries short sequences at each end for part of the extracellular and intracellular protein domains. The fifth (445 bp) and sixth (846 bp) exons encode the rest of the intracellular region. The seventh exon is entirely non-coding.

The sequence 5' of exon 1 is contiguous with that upstream of the CD3 δ gene (Tunnacliffe *et al.*, 1986), and contains an Alu-type repeat element.

Although the CD3 δ gene is split into only five exons, its exon-intron boundaries occur in positions homologous to those of CD3 γ . Since the CD3 γ cDNA sequence can be aligned with that of CD3 δ , it is perhaps not surprising that their respective genes have very similar exon structures. The major differences are: (i) the presence of sequences coded in exon 2 of the CD3 γ gene, which are absent from CD3 δ and which account for some of the

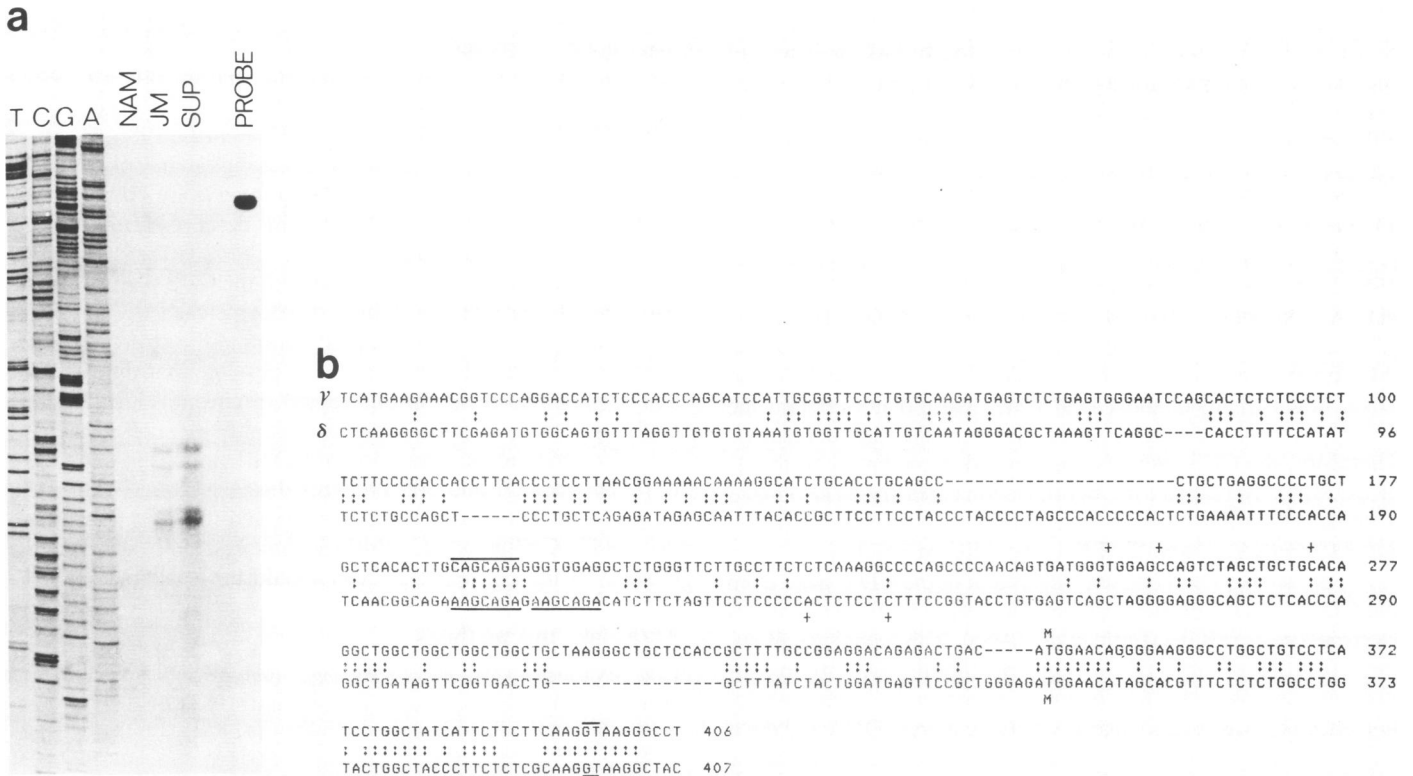


Fig. 3. Determination of CD3 γ transcription initiation positions by S1 mapping, and comparison of CD3 γ and CD3 δ promoter regions. **(a)** A single-stranded, continuously-labelled probe, complementary to, and extending beyond, the coding strand of CD3 γ exon 1 (sequence positions 1197 to 1427) was made as described in Materials and Methods. The probe was annealed with 20 μ g total cytoplasmic RNA from human T-cell lines JM and SUP-T1 and, as a control, the B-cell line Namalwa (NAM in the figure). A sequencing ladder was used to locate transcription start points. **(b)** An alignment of first exon and putative promoter sequences of the CD3 γ and CD3 δ genes is shown, with CD3 γ sequences on the top line. Nucleotide identities are indicated by marks between the lines. Initiator methionines (M), first exon donor splice sites (GT) and major transcription initiation sites (+) are labelled. The heptamers AAGCAGA (CD3 γ) and CAGCAGA (CD3 δ) (see text) are indicated.

extra length of the mature CD3 γ polypeptide over CD3 δ ; and (ii) a subdivision of the CD3 γ 3'-non-coding region into at least two exons (Figure 2).

The unusual proximity of the 5' ends of the CD3 γ and CD3 δ genes suggested that their co-expression might derive from common sequences in the short space between them. To assess potentially functional sequence motifs in the intergenic region, and to determine a precise distance between the two genes, the transcription initiation sites of the CD3 γ gene were determined by S1 mapping (Figure 3a). As with CD3 δ (van den Elsen *et al.*, 1986; Tunnacliffe *et al.*, 1986), multiple initiation positions were seen. Three major CD3 γ start sites are shown in Figure 2; similar results were obtained by primer extension analysis (data not shown). The complete sequence between the CD3 γ and CD3 δ genes (Figure 2) shows that the genes are exactly 1625 bp apart. The sequences immediately adjacent to the cap sites are aligned in Figure 3b. Like CD3 δ , no TATA-box (Breathnach and Chambon, 1981) or potential Sp1-binding sites (CCGCCC or its complement) (Gidoni *et al.*, 1984) are found immediately upstream of the CD3 γ transcription initiation positions. In the case of CD3 δ , two major start sites were defined (Tunnacliffe *et al.*, 1986) and it was noted that each of these was exactly 28 bp downstream from a repeated heptamer sequence AAGCAGA (Figure 3b). It was therefore postulated that this sequence might function like a TATA-box, governing the position of transcription initiation. Although a sequence very similar to this (CAGCAGA) occurs near the initiation sites of the CD3 γ gene, it is 57 bp distant from the most upstream site. This disparity makes the heptamer an unlikely candidate for a TATA-box-like element in this gene, but does not preclude some other role. In general,

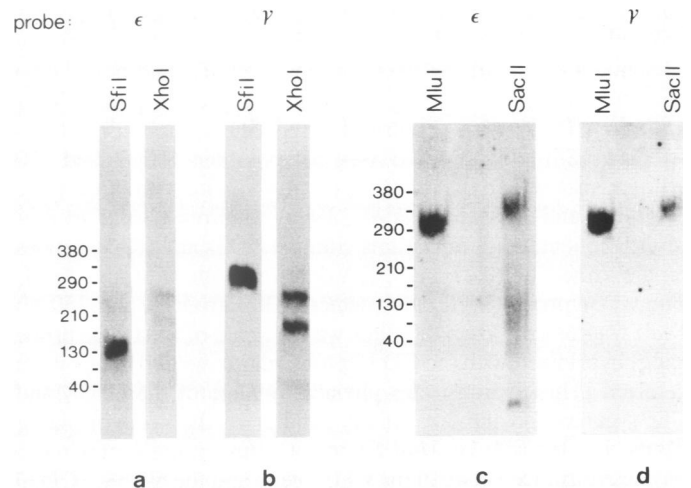


Fig. 4. FIGE analysis of large DNA fragments containing CD3 γ and CD3 ϵ genes. Panels (a) and (b) represent filters made from the same FIGE gel, on which duplicate samples of SfiI and XhoI digests of DNA from T-cell line JM have been fractionated. Panels (c) and (d) represent a second FIGE gel where MluI and SacII digests had been run. Filters in (a) and (c) were hybridized with a CD3 ϵ probe, while those in (b) and (d) were hybridized with a CD3 γ probe. In (d), the same filter as in (c) has first been stripped of CD3 ϵ before rehybridization with CD3 γ . Positions of λ oligomers (units of 42 kb, beginning with the monomer, rounded to the nearest 10 kb) are shown at the side of each gel.

however, the alignment of sequences upstream of the CD3 γ and CD3 δ genes defines few significant homologies. It is clear, therefore, that functional studies are mandatory before further definition of these promoters is possible.

CD3 ϵ gene is linked to the CD3 γ –CD3 δ gene pair

Since the CD3 γ and CD3 δ genes are closely linked, it was of interest to investigate linkage of other CD3 genes. The CD3 ϵ gene was a candidate for linkage to the CD3 γ –CD3 δ gene pair, although it shows only a low level of sequence homology with CD3 δ (Gold *et al.*, 1986), since both the CD3 δ and CD3 ϵ genes have recently been mapped to band q23 of human chromosome 11 (Gold *et al.*, 1987). We have isolated a CD3 ϵ cDNA clone, pJM4C11, from a human T-cell cDNA library (Sims *et al.*, 1984), which corresponds to positions 420 to 1145 of the CD3 ϵ cDNA clone pDJ4, described in Gold *et al.* (1986). However, pJM4C11 did not hybridize to the genomic clones described in this paper. Moreover, a cosmid containing a further 20 kb of DNA downstream of the CD3 δ gene also did not contain the CD3 ϵ gene (A.T., P.Little and T.H.R., unpublished results). Therefore, the CD3 ϵ gene is not situated immediately to the 3' of CD3 δ . Conventional filter hybridization experiments also make it unlikely that the CD3 ϵ gene is closer than 15–20 kb to the 3' of CD3 γ (A.T. and T.H.R., unpublished results).

An estimate of the maximum separation of the CD3 ϵ gene from the CD3 γ –CD3 δ cluster was obtained using field inversion gel electrophoresis (FIGE) (Carle *et al.*, 1986). This technique allows the separation of large DNA fragments (up to 1000 kb) on conventional agarose gels. A selection of infrequently cutting restriction enzymes was used to digest very large human DNA and these digests were fractionated on FIGE, and transferred to cellulose nitrate filters which were hybridized with cDNA probes for CD3 ϵ and CD3 γ (Figure 4). The intention was to demonstrate that both probes hybridize to an identical restriction fragment, and therefore that the CD3 ϵ and CD3 γ genes are physically linked. Figures 4a and 4b show the hybridization pattern obtained using CD3 ϵ and CD3 γ probes with human T-cell DNA cleaved with *Sfi*I and *Xho*I. The CD3 ϵ probe detects one *Sfi*I band of about 130 kb but no signal was seen with *Xho*I digests (Figure 4a). This is probably because the CD3 ϵ gene is flanked by *Xho*I sites giving a fragment too small to detect with these FIGE conditions. Two faint bands are just visible, which may represent partial *Xho*I digestion. In contrast (Figure 4b), the CD3 γ probe detects an ~300 kb *Sfi*I band and two *Xho*I bands (about 160 kb and 250 kb), consistent with the presence of an *Xho*I site within this gene (Figure 1 and Krissansen *et al.*, 1986). Figures 4c and 4d represent sequential hybridizations with the CD3 ϵ and CD3 γ probes on the same filter carrying DNA digested with *Mlu*I and *Sac*II. The CD3 ϵ probe (Figure 4c) detected 300 kb *Mlu*I and 350 kb *Sac*II fragments. When the filter was stripped of the CD3 ϵ probe and rehybridized with the CD3 γ probe, the same two bands were detected. This demonstrates physical linkage of the CD3 γ and CD3 ϵ genes and the size of the smaller of the co-hybridizing bands, i.e. the 300 kb *Mlu*I fragment, thus giving a maximum physical distance between the CD3 ϵ gene and the CD3 γ –CD3 δ gene pair. Furthermore, digestion with *Not*I and *Nru*I yielded bands of 350 kb and 600 kb respectively with either ϵ or γ probes (unpublished), confirming the physical linkage of these genes.

Discussion

Our data show that three human CD3 genes are clustered within a 300 kb region of DNA, occurring in band q23 of chromosome 11. The head-to-head arrangement of the CD3 γ and CD3 δ genes is unusual in that most clusters of related genes are organized head-to-tail, e.g. the mouse MHC class I gene cluster (Flavell *et al.*, 1986). Where divergently transcribed genes are found, they are usually not highly homologous, exceptions being the *yp1*–*yp2* and *sgs7*–*sgs8* gene pairs of *Drosophila* (Hung and

Wensink, 1983; Garfinkel *et al.*, 1983). An assumption often made about divergently transcribed eukaryotic gene pairs is that they share common regulatory elements and in some cases this has been demonstrated (Johnston and Davis, 1984; Osley *et al.*, 1986; Farnham and Schimke, 1986; Kay *et al.*, 1986; Baum *et al.*, 1987). The CD3 γ –CD3 δ intergenic region may also contain bi-directionally active elements, although none are obvious from analysis of sequence alone. Clearly, this deserves further examination by functional studies.

The clustering of the CD3 ϵ gene with the CD3 γ –CD3 δ gene pair suggests that their co-ordinate developmental activation might most readily be achieved by an alteration in the structure of the chromatin domain they occupy. It is possible that the gene for CD3 ζ may also reside in this region of chromosome 11, thus enabling it to be switched on concomitantly by such a mechanism. We estimate the separation of CD3 ϵ from CD3 γ and CD3 δ to be between 20 and 300 kb. This distance is well within the range of current chromosome walking and jumping (Poustka *et al.*, 1987; Collins *et al.*, 1987) techniques, and therefore the cloning of the entire CD3 locus should be feasible.

Materials and methods

Cells, RNA preparations and oligonucleotide synthesis

Human T-cell lines JM (Schneider *et al.*, 1977) and SUP-T1 (Hecht *et al.*, 1984) and the human B-cell line Namalwa (Reedman and Klein, 1973) were grown in RPMI 1640 medium with 10% fetal calf serum at 37°C. Approximately 2×10^8 cells were harvested for cytoplasmic RNA preparation according to Favalaro *et al.* (1980). Oligonucleotide TG1 was made on an Advanced Biosystems DNA synthesizer (model 380B) by T. Smith and purified on native gels according to Carter *et al.* (1985).

Isolation and sequencing of CD3 γ genomic clones

The isolation of overlapping phage clones $\lambda\delta 3A$ and $\lambda\delta 13$ has been described previously (Tunnacliffe *et al.*, 1986). Additional recombinant phage containing CD3 γ sequences were isolated from a genomic library of DNA from the B-cell line SH (Lefranc *et al.*, 1986) made in the phage vector $\lambda 2001$ (Karn *et al.*, 1984), using the CD3 γ cDNA clone pJ6T3 γ -2 (Krissansen *et al.*, 1986) as probe. Restriction maps and the boundaries of regions hybridizing to pJ6T3 γ -2 were determined for these phages and using these results one phage, $\lambda\gamma 7$, was chosen for sequence analysis. Sequencing was carried out largely by the shotgun dideoxy technique (Sanger *et al.*, 1980; Biggin *et al.*, 1983) where either whole $\lambda\gamma 7$ or linearized, self-ligated plasmid subclones were sonicated, fragments repaired, fractionated on low-melting temperature gels, recovered and cloned into M13 vectors. M13 phage containing exon sequences were identified by screening plaques with the insert of pJ6T3 γ -2. This 1.8 kb *Bgl*III–*Sall* fragment containing CD3 γ exon 1 was purified first and then random shotgun sequencing was used to determine the sequence of the whole region. In order to obtain double-stranded sequences or to locate the relative positions of exons, some of the sequence involved site-directed cloning in M13 vectors.

S1 nuclease analysis

A single-stranded, continuously labelled DNA probe, complementary to, and extending beyond, the coding strand of the CD3 γ exon 1 (sequence positions 1197 to 1427) was made by the 'prime cut' method of Biggin *et al.* (1984) using the 22 base oligonucleotide TG1 (5' AAGAATGATAGCCAGGATGAGG 3') to prime off single-stranded DNA of an M13 sequencing clone BS49. The labelled fragment was released by digestion with *Pst*I and purified by gel electrophoresis. Purified probe was used in DNA protection experiments on 20 μ g amounts of cytoplasmic RNA from the T-cell lines JM and SUP-T1 and B-cell line Namalwa (control) using the conditions described previously (Tunnacliffe *et al.*, 1986). DNA fragments protected from S1 nuclease digestion were analysed on sequencing type gels and the position of transcription start points determined by comparison with a standard sequence ladder made using the TG1 oligonucleotide as a primer on BS49 DNA in the presence of dideoxy sequencing mixes.

FIGE analysis of large DNA fragments

DNA was prepared from PBS-washed cells of the human T-cell line JM, largely as described (Gardiner *et al.*, 1986). About 2.5×10^8 cells were resuspended in 5 ml L buffer (0.1 M EDTA, 10 mM Tris pH 7.5, 20 mM NaCl) and mixed with an equal volume of 1% low gelling temperature agarose (FMC) at 42°C. This was then poured into a 9 cm Petri dish and allowed to set at 4°C. Then 1 cm square plugs were cut out and transferred to L buffer plus 1% sarkosyl and 1 mg/ml proteinase K (predigested 4 h, 37°C) at 50°C for 48 h with one

buffer change. Plugs were washed in 10:1 TE (10 mM Tris pH 8, 1 mM EDTA) and treated with PMSF (phenylmethylsulphonyl fluoride, 1 mM in 10:0.1 TE) for 2 h at room temperature (RT). The plugs were then washed 2–3 times in 10:0.1 TE for 2 h, RT and then overnight in TE at 4°C. Plugs were stored in fresh 10:0.1 TE at 4°C for up to 3 months. Each plug had a volume of 100 μ l and contained approximately 20 μ g of DNA. Restriction digests were carried out overnight at 37°C (or 50°C for *Sfi*I) in 100 mM NaCl (except *Sac*II, where no salt was added), 50 mM Tris pH 7.5, 10 mM MgCl₂, 1 mM DTT, 200 μ g/ml acetylated bovine serum albumin (BSA; Anglian Biotechnology) with 80–100 units of restriction enzyme in a total volume of 250 μ l. Digested plugs were then stored in TE at 4°C. For electrophoresis, half-plugs were loaded onto wells of a 1% agarose gel (buffered with 89 mM Tris pH 8, 89 mM boric acid, 2 mM EDTA). Markers used were oligomers of λ 2001, whose monomer is 42 kb long, made as described by Kenwick *et al.* (1987). FICE conditions were based on Carle *et al.* (1986): gels were run at a constant voltage of 200 V with a forward pulse time of 1 s increasing by 0.45 s at each of 100 intervals. The reverse pulse time was initially 0.3 s, increasing with 0.15 s increments. Total run time was 24 h. Circulation of buffer through a chilling system was used to maintain temperature at 15°C. After the run had finished, the gel was stained with ethidium bromide and DNA nicked by u.v.-irradiation prior to blotting onto nitrocellulose (Southern, 1975). Hybridization, with oligolabelled probes (Feinberg and Vogelstein, 1983) of specific activity $>10^9$ c.p.m./ μ g made from cDNA inserts of pJ6T3 γ -2 (CD3 γ) or pJM4C11 (CD3 ϵ), was as described (Sims *et al.*, 1984). After overnight incubation at 65°C, filters were washed in 5 \times SSC, 0.1% SDS at 65°C, and autoradiographed. Before reprobing, filters were washed at 90°C in 0.1% SDS, 0.1 \times SSC and autoradiographed to ensure that all probe had been removed.

Acknowledgements

We would like to thank G.Krissansen and M.Crumpton for providing clone pJ6T3 γ -2, A.Forster for excellent technical assistance, T.Smith for synthesizing oligonucleotide TG1 and R.Watson for help with the FICE system. A.T. and L.B. are the recipients of MRC Training Fellowships.

References

- Baum, J.A., Geever, R. and Giles, N.H. (1987) *Mol. Cell. Biol.*, **7**, 1256–1266.
- Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965.
- Biggin, M., Farrell, P.J. and Barrell, B.G. (1984) *EMBO J.*, **3**, 1083–1090.
- Borst, J., Alexander, S., Elder, J. and Terhorst, C. (1983) *J. Biol. Chem.*, **258**, 5135–5141.
- Breathnach, R. and Chambon, P. (1981) *Annu. Rev. Biochem.*, **50**, 349–383.
- Campana, D., Thompson, J.S., Amlot, P., Brown, S. and Janossy, G. (1987) *J. Immunol.*, **138**, 648–655.
- Carle, G.F., Frank, M. and Olson, M.V. (1986) *Science*, **232**, 65–68.
- Carter, P., Bedouelle, H., Waye, M.M.Y. and Winter, G. (1985) *Site Directed Mutagenesis in M13*. Anglian Biotechnology, Colchester, UK.
- Collins, F.S., Drumm, M.L., Cole, J.L., Lockwood, W.K., van de Woude, G.F. and Iannuzzi, M.C. (1987) *Science*, **235**, 1046–1049.
- Farnham, P.J. and Schimke, R.T. (1986) *Mol. Cell. Biol.*, **6**, 365–371.
- Favaloro, J., Treisman, R. and Kamen, R. (1980) *Methods Enzymol.*, **65**, 718–749.
- Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.*, **132**, 6–13.
- Flavell, R.A., Allen, H., Burkly, L.C., Sherman, D.H., Waneck, G.L. and Widera, G. (1986) *Science*, **233**, 437–443.
- Furley, A.J., Mizutani, S., Weilbaecher, K., Dhaliwal, H.S., Ford, A.M., Chan, L.C., Molgaard, H.V., Toyonaga, B., Mak, T., van den Elsen, P., Gold, D.P., Terhorst, C. and Greaves, M.F. (1986) *Cell*, **46**, 75–87.
- Garfinkel, M.D., Pruitt, R.E. and Meyerowitz, E.M. (1983) *J. Mol. Biol.*, **168**, 765–789.
- Gardiner, K., Laas, W. and Patterson, D. (1986) *Somat. Cell Mol. Genet.*, **12**, 185–195.
- Gidoni, D., Dynan, W.S. and Tjian, R. (1984) *Nature*, **312**, 409–413.
- Gold, D.P., Puck, J.M., Pettey, C.L., Cho, M., Coligan, J., Woody, J.N. and Terhorst, C. (1986) *Nature*, **321**, 431–434.
- Gold, D.P., van Dongen, J.J.M., Morton, C.C., Bruns, G.A.P., van den Elsen, P., van Kessel, A.H.M.G. and Terhorst, C. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1664–1668.
- Hecht, F., Morgan, R., Kaiser-McCaw Hecht, B. and Smith, S.D. (1984) *Science*, **226**, 1445–1447.
- Hung, M.-C. and Wensink, P.C. (1983) *J. Mol. Biol.*, **164**, 481–492.
- Johnston, M. and Davis, R.W. (1984) *Mol. Cell. Biol.*, **4**, 1440–1448.
- Kanellopoulos, J.M., Wigglesworth, N.M., Owen, M.J. and Crumpton, M.J. (1983) *EMBO J.*, **2**, 1807–1814.
- Karn, J., Matthes, W.D., Gait, M.J. and Brenner, S. (1984) *Gene*, **32**, 217–224.
- Kay, R.J., Boissy, R.J., Russnak, R.H. and Candido, E.P.M. (1986) *Mol. Cell. Biol.*, **6**, 3134–3143.

- Kenwick, S., Patterson, M., Speer, A., Fischbeck, K. and Davies, K. (1987) *Cell*, **48**, 351–357.
- Krissansen, G.W., Owen, M.J., Verbi, W. and Crumpton, M.J. (1986) *EMBO J.*, **5**, 1799–1808.
- Lefranc, M.-P., Forster, A., Baer, R., Stinson, M.A. and Rabbitts, T.H. (1986) *Cell*, **45**, 237–246.
- Meuer, S.C., Acuto, D., Hercend, T., Schlossman, S.F. and Reinherz, E.L. (1984) *Annu. Rev. Immunol.*, **2**, 23–50.
- Osley, M.A., Gould, J., Kim, S., Kane, M. and Hereford, L. (1986) *Cell*, **45**, 537–544.
- Poustka, A., Pohl, T.M., Barlow, D.P., Frischauf, A.-M. and Lehrach, H. (1987) *Nature*, **325**, 353–355.
- Rabbitts, T.H., Lefranc, M.-P., Stinson, M.A., Sims, J.E., Schroder, J., Steinmetz, M., Spurr, N.L., Solomon, E. and Goodfellow, P.N. (1985) *EMBO J.*, **4**, 1461–1465.
- Reedman, B.M. and Klein, G. (1973) *Int. J. Cancer*, **11**, 499–520.
- Reinherz, E.L. (1987) *Nature*, **325**, 660–663.
- Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) *J. Mol. Biol.*, **143**, 161–178.
- Schneider, U., Schwenk, H.U. and Bornkamm, G. (1977) *Int. J. Cancer*, **19**, 621–626.
- Sims, J.E., Tunnacliffe, A., Smith, W.J. and Rabbitts, T.H. (1984) *Nature*, **312**, 541–545.
- Southern, E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
- Tsoukas, C.D., Valentine, M., Lotz, M., Vaughn, J.H. and Carson, D.A. (1984) *Immunol. Today*, **5**, 311–313.
- Tunnacliffe, A., Sims, J.E. and Rabbitts, T.H. (1986) *EMBO J.*, **5**, 1245–1252.
- van den Elsen, P., Shepley, B.-A., Borst, J., Coligan, J.E., Markham, A.F., Orkin, S. and Terhorst, C. (1984) *Nature*, **312**, 413–418.
- van den Elsen, P., Georgopoulos, K., Shepley, B.-A., Orkin, S. and Terhorst, C. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2944–2948.
- van Dongen, J.J.M., Quertermous, T., Bartram, C.R., Gold, D.P., Wolvers-Tettero, I.L.M., Comans-Bitter, W.M., Hooijkaas, H., Adriaansen, H.J., de Klein, A., Raghavachar, A., Ganser, A., Duby, A.D., Seidman, J.G., van den Elsen, P. and Terhorst, C. (1987) *J. Immunol.*, **138**, 1260–1269.
- Weissman, A.M., Samelson, L.E. and Klausner, R.D. (1986) *Nature*, **324**, 480–482.

Received on May 26, 1987