Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 7644–7648, November 1987
Immunology

Structure of the mouse gene encoding CD4 and an unusual transcript in brain

(L3T4/T-cell differentiation antigens/helper T cells)

SCOTT D. GORMAN, BÉATRICE TOURVIEILLE, AND JANE R. PARNES

Department of Medicine, Division of Immunology, Stanford University Medical Center, S021, Stanford, CA 94305

Communicated by Harden M. McConnell, July 14, 1987

ABSTRACT **The T-cell differentiation antigen CD4 plays an important role in the function of T cells that recognize class II major histocompatibility complex proteins. Mouse CD4 (L3T4) has previously been shown to be evolutionarily related to immunoglobulin variable regions based on the predicted protein sequence from cDNA clones. The gene encoding L3T4 was found to be transcribed not only in a subset of T-lineage cells but also unexpectedly in brain, where a shorter transcript was found. In the present study the gene encoding L3T4 is shown to span 26 kilobases and to contain 10 exons. The structural organization is similar to that of other members of the immunoglobulin gene superfamily except for the striking presence of an intron in the middle of the sequence encoding the amino-terminal immunoglobulin-like homology unit. The structure of the shorter L3T4 transcript in mouse brain has been determined. This mRNA appears to be generated from a transcriptional start site within the coding sequence in exon VI. If translated, this transcript would encode a protein of 217 amino acids that lacks the usual L3T4 signal peptide and the amino-terminal 214 amino acids of the mature protein.**

CD4 is a cell-surface glycoprotein expressed on T lymphocytes that recognize class II major histocompatibility complex (MHC) proteins. It is traditionally used as a marker for the helper/inducer subset of peripheral T cells, although it is also present on cytotoxic cells that recognize class II MHC molecules. CD4 is thought to play a role in increasing the avidity of the interaction between T cells and antigen-presenting cells, perhaps by binding a nonpolymorphic determinant on class II molecules, and monoclonal antibodies specific for CD4 block the function of class II MHC-restricted T cells (1–8). However, studies in which anti-CD4 mAbs block T-cell function in the absence of class II MHC molecules have suggested an alternative or additional role for CD4 in transducing a negative signal to the T cell (9–11). There is currently no direct evidence that allows one to exclude or conclusively prove either of these models.

We have described (12) the isolation of cDNA clones encoding the mouse CD4 protein, L3T4, and demonstrated that this molecule is evolutionarily related to immunoglobulin variable (V) regions. We found that L3T4 is encoded by a single nonrearranging gene on mouse chromosome 6 (12, 13). This gene was shown to be transcribed into two polyadenylated mRNA species: a 3.7-kilobase (kb) species in T-lineage cells and, surprisingly, an additional 2.7-kb species in brain (12). In the work reported here we now show the structure of the gene encoding L3T4 and further characterize the pattern of transcription of this gene in brain.

MATERIALS AND METHODS

Gene Structure and Sequence. The structure and sequence of the gene encoding L3T4 were determined by using three overlapping genomic clones isolated with L3T4 cDNA probes (12) from a B10.CAS2 mouse liver genomic library (gift of P. Jones, Stanford University). The nucleotide sequence was determined by the dideoxynucleotide chain-termination method (14) with genomic clone fragments subcloned into phage M13 vectors mp10, mp11, mp18, and mp19 (15).

S1 Nuclease Mapping. RNA from C57BL/6 mouse thymus, liver, and brain was isolated by the guanidine thiocyanate method of Chirgwin *et al.* (16). RNA (50–100 μg) was hybridized with antisense single-stranded M13 probes containing the insert of L3T4 cDNA clones pcL3T4-C7 or pcL3T4-12.2 (ref. 12; unpublished data) and treated with S1 nuclease as described (17). The resultant S1 nuclease-resistant hybrids were precipitated with ethanol, and aliquots were treated with 5 ng of ribonuclease type A-1 (Sigma) at 37°C for 15 min. The samples were electrophoresed on 1.5% agarose gels and transferred to nitrocellulose filters by the procedure of Southern (18). Duplicate samples hybridized to pcL3T4-C7 were electrophoresed on an 8% polyacrylamide gel containing 7 M urea and electroblotted onto nylon membrane (Genatran 45, Plasco, Woburn, MA) for confirmation of sizing. Blots were hybridized as described (17) to the insert of cDNA clone pcL3T4-C7 (or fragments of this clone as stated) labeled with ^{32}P by random hexamer priming (19).

RESULTS

Structure of the Gene Encoding L3T4. The structure of the gene encoding L3T4 was determined by restriction mapping and sequencing of three overlapping genomic clones. The organization of the gene is illustrated in Fig. 1. It is composed of 10 exons separated by nine introns and spans 26 kb of DNA. The exons correlate only roughly with the predicted protein domains (Fig. 1). The most striking finding in this regard is the presence of a large intron (6.4 kb) in the middle of the sequence encoding the first immunoglobulin V region-like homology unit (designated "V") of L3T4. There is also a large intron (8.6 kb) within the 5′ untranslated region. The second V region-like domain (V′) is encoded in a single exon (exon V), while the sequence encoding the connecting peptide (CP) is split between two exons (VI and VII). It has been argued that the sequences encoded by each of these latter two exons may also be evolutionarily related to immunoglobulin homology units (20). If so, it is not surprising to find an intron between them. The transmembrane segment is encoded on a single exon (VIII), while the cytoplasmic tail sequence is split among three (VIII, IX, and X).

The nucleotide sequence of the exons and at least parts of all introns is shown in Fig. 2.* The coding sequence is

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.

Abbreviations: V, variable; J, joining.

*The sequence reported in this paper is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03003).

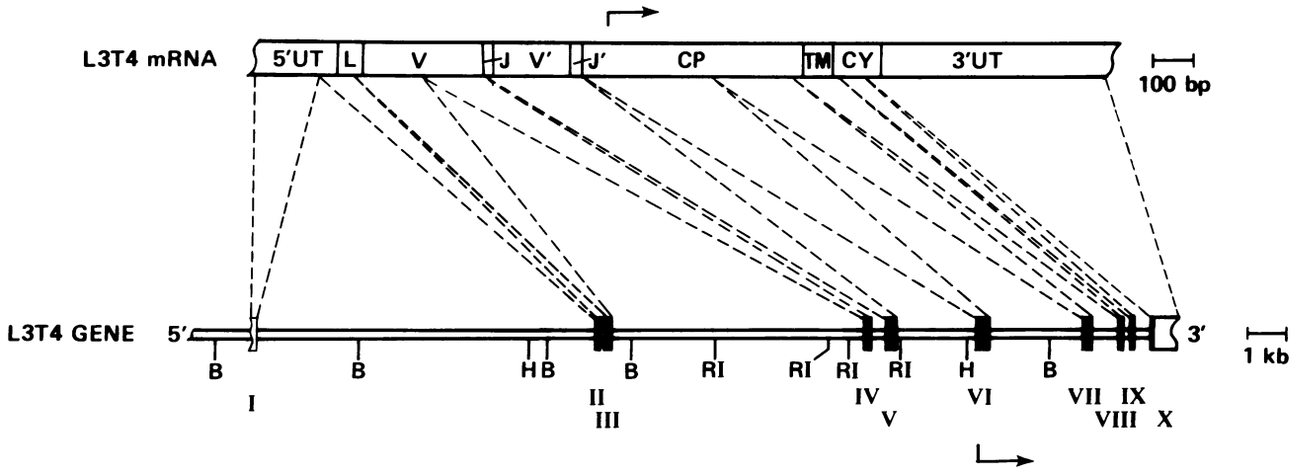

FIG. 1. Structure of the L3T4 gene. (*Lower*) The structure and partial restriction map of the L3T4 gene. The exons are indicated by boxes and numbered with Roman numerals. Shaded boxes indicate protein coding regions and open boxes represent 5′ and 3′ untranslated regions. (*Upper*) Dashed lines indicate where the exons of the L3T4 gene are represented along the structure of the L3T4 mRNA. The portions of the mRNA encoding the previously identified protein domains are indicated along the mRNA structure: UT, untranslated region; L, leader, V and V′, sequences homologous to immunoglobulin V regions; J and J′, sequences homologous to immunoglobulin J segments; CP, connecting peptide; TM, transmembrane region; CY, cytoplasmic tail. Arrows indicate the start site of the smaller L3T4 mRNA species found in mouse brain. Restriction endonuclease recognition sites: B, Bgl II; H, *Hin*dIII; RI, *Eco*RI.

identical to that previously determined for our cDNA clones with the exception of a single base change in the sequence encoding amino acid −1. At this position the genomic clone has GAG (glutamic acid), while both our cDNA clones (ref. 12; unpublished data) and that reported by Littman and Gettner (21) have GGG (glycine). Since this residue is thought to be in the signal peptide based on amino-terminal sequencing of the protein (22), this discrepancy (possibly a polymorphism) would not affect the sequence of the mature protein on the cell surface.

The intron/exon borders were determined by comparison of the genomic sequence with that of the cDNA. All of the splicing junctions obey the GT/AG rule, although the acceptor sequence at the end of intron 8 is unusually purine rich (TGGGCAG as compared with the consensus YYYXYAG, where Y is a pyrimidine and X is any nucleotide). As is typical of other members of the immunoglobulin gene superfamily, all introns divide codons between the first and second nucleotides except for those between exons encoding the cytoplasmic tail (introns 8 and 9).

Although we have not yet mapped the start site of transcription of the L3T4 gene, blot-hybridization analyses indicate that the mRNA does not extend more 5′ than the sequence shown in Fig. 2 (data not shown). There are several A+T-rich sequences within the region 5′ of the most 5′ cDNA clones described, and one or more of these may serve as a promoter sequence. However those that are more 5′ than the ones specifically demarcated in Fig. 2 are followed by an initiation codon (ATG) that, if used, would result in early termination.

Structure of L3T4 mRNA in Brain. We and others (12, 23) have demonstrated previously that the L3T4 gene is transcribed in mouse brain and that the predominant L3T4 mRNA in brain is approximately 1 kb smaller than that found in thymus, spleen, lymph node, and T-cell lines. The most likely explanation seemed to be an alternative form of splicing of the L3T4 mRNA, since only a single gene is present in the mouse genome (12). Knowing the structure of the L3T4 gene, we further investigated the mechanism of generation of this smaller L3T4 mRNA in brain by S1 nuclease mapping and by hybridization of probes from different parts of the L3T4 cDNA to S1 nuclease hybrids or to RNA blots. As shown in Fig. 3, a 1280-base-pair (bp) probe, extending from the sequence encoding amino acid 8 through 7 bp of 3′ untranslated region, was fully protected from S1 nuclease digestion after hybridization to thymus RNA, but only 700 bp were protected after hybridization to brain RNA (plus a much smaller amount of fully protected fragment), and no protection was seen with liver RNA. Blots of the 700-bp S1 nuclease hybrid between brain mRNA and this fragment hybridized to a probe containing the most 3′ 229 bp of this cDNA fragment and not to a probe containing the most 5′ 303 bp (data not shown). Similar hybridization results were found on blots of brain mRNA, in which the 2.7-kb brain mRNA species was additionally found not to hybridize to probes containing the 5′ untranslated region (data not shown). To determine whether the 700-bp protected fragment seen on S1 nuclease analysis of brain mRNA extended fully through the 3′ end of the S1 probe, we used a longer cDNA probe containing 770 bp of 3′ untranslated region and found a correspondingly longer protected fragment. Taken together, these results indicate that the approximate start of the L3T4 mRNA in brain is within the sequence encoding amino acid 200 (±10 bp), and that this mRNA extends 3′ from this point colinearly with the L3T4 mRNA found in thymus (Figs. 1 and 2). We cannot completely exclude the possibility that there might be a small amount of additional noncontiguous sequence at the 5′ end of this brain transcript, but if so, it does not hybridize to any portion of the L3T4 mRNA found in thymus. Furthermore, the 5′ end that we have identified for the brain L3T4 mRNA does not correlate with an intron/exon border of the L3T4 gene as defined by the splicing pattern in T-lineage cells. Therefore, the brain transcript cannot be the result of a simple alternative splicing pattern skipping over exons I through V. Although we cannot rule out an unusual splice from a short upstream exon unique to brain to a poorly conserved acceptor site within the coding sequence of exon VI, it appears most likely that the brain mRNA results from use of an alternative promoter and transcription start site within the 5′ portion of exon VI. In this regard it is notable that there is a TATAA sequence that begins at position −32 relative to the predicted start site of the brain mRNA (Fig. 2), and this could represent a promoter active in brain. Similarly, there is a methionine codon (ATG) 43 bp 3′ of the predicted brain mRNA start site (Fig. 2), and this could represent a translational start site for the brain mRNA.

FIG. 2. *(Legend appears at the bottom of the opposite page.)*

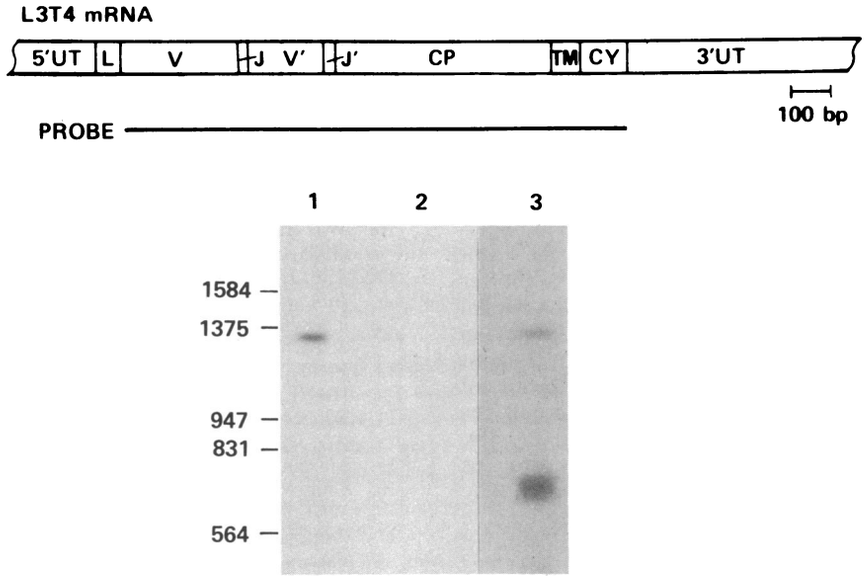

FIG. 3. S1 nuclease analysis of brain mRNA. The origin of the probe used for the S1 nuclease mapping shown is indicated along the structure of the L3T4 mRNA. The autoradiogram shows the fully protected probe fragment with thymus RNA (lane 1), no protection with liver RNA (lane 2), and a predominant 700-bp protected fragment plus some full-length fragment with brain RNA (lane 3). Exposures were with an intensifying screen at −70°C for 30 min for thymus and liver and 3 days for brain. Longer exposures of the liver lane did not reveal any protected fragments. Size markers are shown to the left of the autoradiogram in bp and represent products of a *Hin*dIII/*Eco*RI digest phage λ.

DISCUSSION

In contrast to the gene encoding the alternative T-cell subset marker Ly-2, which is only about 5 kb in length (24), the gene encoding L3T4 is a much larger one, primarily because of the size of the first and third introns. The gene is similar in organization to other members of the immunoglobulin gene superfamily with the notable exception of intron 3, which divides the sequence encoding the amino-terminal V-like region of L3T4 into two exons. This division of an immunoglobulin-like homology unit between two exons (a feature also found in the human homolog; ref. 21) is unique among immune system proteins within the immunoglobulin gene superfamily. The most likely explanation for this finding is the insertion of an intron after the divergence of this gene from immunoglobulin and T-cell receptor genes, but prior to the separation of mouse and human species. A similar conclusion has recently been drawn by Littman and Gettner (21).

We have identified previously two sequences similar to immunoglobulin *J* (joining) segments in the predicted L3T4 protein sequence. One, (L3T4 J) followed the amino-terminal V-like region (L3T4 V), while the other (L3T4 J′) followed a second, foreshortened V-like region of the protein (L3T4 V′). Our analysis of the L3T4 gene structure shows that there is an intron within the first *J*-like sequence, making its evolutionary relationship to immunoglobulin *J* sequences more questionable. In contrast, the *J′* sequence is on the same exon as *V′* and is immediately followed by an intron. Therefore, it is more likely that the sequence similarity of L3T4 *J′* to immunoglobulin *J* segments is evolutionarily significant. We have not found any sequences closely resembling the heptamer–spacer–nonamer sequences that are thought to be signals for rearranging immunoglobulin and T-cell receptor genes near the 3′ ends of the V- or J-like segments of the L3T4 gene. These results are consistent with our previous conclusion that the L3T4 gene does not rearrange (12) and suggest that either the L3T4 gene split off from immunoglobulin genes before the ability to rearrange had been acquired or the L3T4 gene subsequently lost that ability.

The gene encoding L3T4 may be one of a growing family of immunoglobulin-related genes (e.g., *Thy-1, OX2*) whose expression is shared between the hematopoietic system and the central nervous system (25, 26). The L3T4 mRNA in brain consists primarily of a shorter species than that in T cells (although a much smaller amount of full-sized mRNA was also detected). The data presented here indicate that the smaller form of L3T4 mRNA in brain is most likely the result of transcription from a different start site, entirely excluding the first five exons of the gene and beginning within exon VI. This mRNA could potentially be translated into a protein in the same frame as L3T4. Such a protein is predicted to be only 217 amino acids in length as compared to 431 for mature L3T4. It would be missing the signal peptide and the amino-terminal 214 amino acids of mature L3T4 but would continue linearly through the cytoplasmic tail. The amino terminus of this predicted protein does not have a typical signal sequence, so it is possible that the protein, if produced, remains inside the cell or perhaps uses another sequence (e.g., the usual transmembrane sequence) as a signal peptide. We have attempted to stain fixed mouse brain sections with

FIG. 2 (*on opposite page*). Nucleotide sequence of the L3T4 gene. The complete nucleotide sequence of the exons and partial nucleotide sequence of all introns are shown. The deduced amino acid sequence is indicated above the nucleotide sequence for all protein coding regions of the gene. The numbers in the right margin refer to the number of the last amino acid begun on that line. The border between the signal peptide and mature protein sequence has been shifted three amino acids compared to that previously reported (12) based on identification of the amino terminus by protein sequencing (22). Horizontal arrows indicate the beginning of protein coding segments (labeled as in Fig. 1), introns [denoted intervening sequences (IVS) 1–9], and untranslated regions. The GT (donor) and AG (acceptor) consensus sequences at the mRNA splicing junctions are underlined. An asterisk (*) denotes the start of our most 5′ cDNA clone. Another reported cDNA clone starts 27 bp more 5′ (21). Two A+T-rich sequences that are candidate promoters are overlined in this 5′ region. A vertical arrow indicates (within ±10 bp) the start site of the shorter L3T4 transcript in mouse brain. The preceding TATAA sequence, which may be a promoter element, is underlined. The ATG (methionine codon) which could serve as a translation start site for this brain transcript is underlined and overlined.

monoclonal antibodies against several epitopes of the L3T4 molecule but have not been able to convincingly demonstrate protein expression to date. It is possible that no protein is made from this unusual transcript, or that our available monoclonals do not recognize the protein translated from the foreshortened mRNA. In any case, the predicted protein would almost certainly not function in the same manner as the full-length L3T4 on T lymphocytes, since it is missing the amino-terminal portion that is most closely related to immunoglobulin recognition units. In the human system significant levels of the full-length CD4 mRNA as well as a shorter species have been described in brain (23). Monoclonal antibodies specific for CD4 have been found to stain human brain sections (27, 28). Although it remains controversial, it is likely that the staining in human brain is a result of expression of the full-length CD4 molecule on macrophages and possibly also on glial and/or neuronal cells (27, 28). Notably, if the shorter transcript in human brain correlates with the one we have observed in the mouse, it could not be translated into protein because an equivalent methionine codon is not present. In mouse brain we have identified astrocytes as at least one cell type that expresses L3T4 transcripts (unpublished data).

We thank Dr. Pat Jones for the B10.CAS2 genomic library and Diane Bet for assistance in preparation of the manuscript. This work was supported by National Institutes of Health Grants GM34991 and AI11313 and a grant from the Weingart Foundation. S.D.G. is a recipient of Postdoctoral Fellowship 1 F32 CA07877 from the National Cancer Institute, B.T. is a recipient of a postdoctoral fellowship from the National Multiple Sclerosis Society, and J.R.P. was a fellow of the John A. Hartford Foundation.

1. Engleman, E. G., Benike, C. J., Metzler, C., Gatenby, P. A. & Evans, R. (1981) *J. Immunol.* **127,** 2124–2129.
2. Krensky, A. M., Clayberger, C., Reiss, C. S., Strominger, J. L. & Burakoff, S. J. (1982) *J. Immunol.* **129,** 2001–2003.
3. Biddison, W. E., Rao, P. E., Talle, M. A., Goldstein, G. & Shaw, S. (1982) *J. Exp. Med.* **156,** 1065–1083.
4. Dialynas, D. P., Wilde, D. B., Marrack, P., Pierres, A., Wall, K. A., Havran, W., Otten, G., Loken, M. R., Pierres, M., Kappler, J. & Fitch, F. W. (1983) *Immunol. Rev.* **74,** 29–56.
5. Reinherz, E. L., Meuer, S. C. & Schlossman, S. F. (1983) *Immunol. Rev.* **74,** 83–112.
6. Swain, S. L. (1983) *Immunol. Rev.* **74,** 129–142.
7. Marrack, P., Endres, R., Shimonkevitz, R., Zlotnik, A., Dialynas, D., Fitch, F. & Kappler, J. (1983) *J. Exp. Med.* **158,** 1077–1091.
8. Greenstein, J. L., Kappler, J., Marrack, P. & Burakoff, S. J. (1984) *J. Exp. Med.* **159,** 1213–1223.
9. Bank, I. & Chess, L. (1985) *J. Exp. Med.* **162,** 1294–1303.
10. Wassmer, P., Chan, C., Logdberg, L. & Shevach, E. M. (1985) *J. Immunol.* **135,** 2237–2242.
11. Tite, J. P., Sloan, A. & Janeway, C. A., Jr. (1986) *J. Mol. Cell. Immunol.* **2,** 179–190.
12. Tourvieille, B., Gorman, S. D., Field, E. H., Hunkapiller, T. & Parnes, J. R. (1986) *Science* **234,** 610–614.
13. Field, E. H., Tourvieille, B., D'Eustachio, P. & Parnes, J. R. (1987) *J. Immunol.* **138,** 1968–1970.
14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
15. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78.
16. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
17. Zamoyska, R., Vollmer, A. C., Sizer, K. C., Liaw, C. W. & Parnes, J. R. (1985) *Cell* **43,** 153–163.
18. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
19. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
20. Clark, S. J., Jefferies, W. A., Barclay, A. N., Gagnon, J. & Williams, A. F. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 1649–1653.
21. Littman, D. R. & Gettner, S. N. (1987) *Nature (London)* **325,** 453–455.
22. Classon, B. J., Tsagaratos, J., McKenzie, I. F. C. & Walker, I. D. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 4499–4503.
23. Maddon, P. J., Dalgleish, A. G., McDougal, J. S., Clapham, P. R., Weiss, R. A. & Axel, R. (1986) *Cell* **47,** 333–348.
24. Liaw, C. W., Zamoyska, R. & Parnes, J. R. (1986) *J. Immunol.* **137,** 1037–1043.
25. Williams, A. F. & Gagnon, J. (1982) *Science* **216,** 696–703.
26. Clark, M. J., Gagnon, J., Williams, A. F. & Barclay, A. N. (1985) *EMBO J.* **4,** 113–118.
27. Pert, C. B., Hill, J. M., Ruff, M. R., Berman, R. M., Robey, W. G., Arthur, L. O., Ruscetti, F. W. & Farrar, W. L. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9254–9258.
28. Funke, I., Hahn, A., Rieber, E. P., Weiss, E. & Riethmuller, G. (1987) *J. Exp. Med.* **165,** 1230–1235.