LAG-3, A NOVEL LYMPHOCYTE ACTIVATION GENE CLOSELY RELATED TO CD4

By FRÉDÉRIC TRIEBEL, SETSUKO JITSUKAWA, ELENA BAIXERAS, SERGIO ROMAN-ROMAN, CATHERINE GENEVEE, EVANI VIEGAS-PEQUIGNOT,* AND THIERRY HERCEND

*From the Laboratoire d'Immunologie Cellulaire, Institut Gustave-Roussy, 94805 Villejuif, France; and *Unité de Recherche 620, Structure et Mutagenèse Chromosomique, Section de Biologie, Institut Curie, 75331 Paris, France*

A number of cell surface structures are thought to belong to the Ig superfamily (IgSF)[1] because they contain at least one domain with a characteristic folding pattern, called the Ig fold (reviewed in reference 1). Several of these molecules have critical functions in immune responses. In addition to ensuring specific antigen recognition (Ig, TCR), they may function as monomorphic ligands critical in cell-cell interactions (e.g., ICAM, CD4, CD8), receptors for viruses (e.g., CD4, ICAM), or lymphokine receptors (e.g., IL-1-R, IL-6-R).

We report here the characterization of a novel human gene, termed lymphocyte activation gene 3 (LAG-3), selectively transcribed in activated NK and T lymphocytes. It codes for a membrane protein with four extracellular IgSF domains. The sequence data, the compared exon/intron organization, and the chromosomal localization revealed that LAG-3 is closely related to CD4.

Materials and Methods

Cell Lines. The isolation and growth of the fetal $CD3^-$ $CD2^+$ F55IIIE5 (or F_5) cloned cell line has been described elsewhere (2). For mass production, the cell suspensions were plated on a feeder layer composed of irradiated allogeneic PBL plus the EBV-transformed B cell line Laz388 in V-bottomed 96-well plates at 3,000 cells per well with rIL-2 and lymphocyte-conditioned medium. 200 plates were harvested at a concentration of 3×10^6 cells/ml after 12 d in culture to give 6×10^9 cells. For the Northern blot analyses, similar culture conditions were used to produce the relevant cells.

Isolation of Membrane-bound (MB) Cytoplasmic RNA and Construction of a cDNA Sublibrary. 4×10^9 F_5 cells were washed twice in PBS and allowed to swell for 5 min at a concentration of 5×10^8 cells/ml in ice-cold hypotonic buffer medium RSB (0.01 M KCl, 0.0015 M $MgCl_2$, 0.01 M Tris-HCl, pH 7.4) (3). The cells were then ruptured mechanically with 10 strokes of a tight-fitting glass homogenizer (Dounce). The homogenate was diluted fivefold in 2.5 M sucrose TKM buffer (0.15 M KCL, 0.005 M $MgCl_2$, 0.05 M Tris-HCl, pH 7.4) and layered over 2 vol of 2.5 M sucrose TKM buffer. Two layers of sucrose TKM buffer were successively added, one with 2.05 M sucrose and a second with 1.3 M sucrose. The gradients

This work was supported in part by Association pour la Recherche sur le Cancer grants 2110 and 6338.

Address correspondence to Frederic Triebel, Laboratoire d'Immunologie Cellulaire, Institut Gustave-Roussy, 39, rue Camille Desmoulins, 94805 Villejuif, France.

[1] *Abbreviations used in this paper:* aa, amino acid; IgSF, Ig superfamily; LAG-3, lymphocyte activation gene 3; MB, membrane bound; TM, transmembrane.

were centrifuged for 5 h at 4°C in an SW 27.1 rotor (Spinco) at 25,000 rpm. The tubes were punctured at the interface between the 2.05 and 1.3 M sucrose layers. After two extractions by phenol-chloroform and precipitation by ethanol, the poly(A)$^+$ mRNA was prepared by oligo(dT) column chromatography (4). sscDNA was synthesized using reverse transcriptase primed with oligo(dT). dscDNA was synthesized from poly(A)$^+$ RNA derived from the MB-F_5 material according to the method of Gubler and Hoffman (5). After protection of internal Eco RI sites by Eco RI methylase and size selection on an ultra low temperature agarose gel (>500 bp), the dscDNAs were cloned into the Eco RI site of λ gt10 using linkers (6). The resulting λ gt10-cDNA hybrids were then packaged in vitro using a commercial packaging extract (Amersham Corp., Arlington Heights, IL). After plating on *Escherichia coli* C600 Hfl$^+$, we obtained 6 × 10^4 recombinant phages.

Isolation and Characterization of LAG-3 cDNA Clones. The MB-F_5 cDNA library (2 × 10^4 recombinant phages) was plated on *E. coli* C600/Hfl (10^3 phages per plate). Screening of filters was carried out according to standard procedures, except that Gene Screen Plus (Dupont Co., Wilmington, DE) nylon membranes were used. Hybridizations were done for 16 h at 42°C in 50% formamide, 5× SSC, 5% Denhart's solution, 100 μg/ml denatured salmon sperm DNA, and 5 × 10^6 dpm/ml ^{32}P-labeled ssDNA MB-F_5 subtracted probe. Filters were washed in 1× SSC, 0.5% SDS at 65°C twice, followed by washing in 0.1× SSC, 0.1% SDS at 42°C. The cDNA inserts were subcloned into PBS, a T3-T7 transcriptional promoter containing vector (Stratagene, La Jolla, CA). To isolate cDNA clones that contain additional 5′ sequences, we screened in turn three different TCR-γ/δ$^+$ cell-derived λ gt10 libraries containing cDNA primed either with oligo(dT), with random hexamers, or with a specific primer (nucleotides 704–688 of sequence FDC in Fig. 1). ^{32}P-labeled ssRNA FD47 probe was transcribed in vitro from the linearized pBS plasmid using T7 polymerase in the presence of 10 μM ^{32}P-UTP (800 Ci/mmol^{-1}) (7) and used to screen the first cDNA library to obtain clone FD19 (see Fig. 1). Subsequently, a 0.3-kb Bam HI–Hind III genomic fragment that corresponds to the more 5′ region of the fourth exon (Fig. 2) was labeled by the hexamer-priming method (8) and used to screen the second library to obtain clones FD61 and FD101, and the third library to obtain a near full-length 5′ sequence containing cDNA (FD191).

The sequences of clones FD47 and FD19 were determined directly from ds supercoiled plasmids in both directions with an universal M13 primer or a reverse M13 primer using the dideoxy chain termination procedure (9) and the modified T7 polymerase (United States Biochemical Corp., Cleveland, OH). The sequences of FD61, FD101, and FD191 were determined from ssDNA after cloning in either the M13mp18 or the M13mp19 vectors. Comparison of amino acid sequences to known Ig-related domains was done using the ALIGN program with a bias of 6, a break penalty of 6, and 150 random runs (1).

Northern and Southern Blot Analysis. RNA was prepared from the cytoplasmic fraction of cells lysed in hypotonic buffer in the presence of NP-40. Total RNA was purified by two phenol-chloroform extractions, followed by two ethanol precipitations. The RNA preparations (10 mg/slot) were dissolved in water, heated at 50°C for 1 h in the presence of 1 M glyoxal, 50% DMSF, and run in a 1% agarose gel at 50 V for 3 h. The gels were then soaked in 10× SSC for 30 min, and the RNA was transferred on two Gene Screen Plus nylon membranes.

For the Southern blot analysis, high molecular weight genomic DNA samples (10 mg) were digested with either Eco RI, Hind III, Bam HI, or XbaI, subjected to electrophoresis through 0.7% agarose, and blotted in alkaline buffer onto Gene Screen Plus nylon membranes.

^{32}P-labeled DNA probes were prepared from agarose-purified fragments by the hexamer random priming method (8). The nylon filters were prehybridized in 50% formamide, 5× SSC, 5× Denhardt's solution, 100 μg/ml salmon sperm DNA, 10% polyethylene glycol 6000 for 4 h at 42°C, and then hybridized in the same buffer with 10^6 dpm/ml ^{32}P-labeled cDNA probe at 42°C for 16 h. Filters were washed twice for 30 min in 1× SSC, 0.5% SDS at 65°C, followed by a 30-min wash in 0.1× SSC, 0.1% SDS at 42°C.

Molecular Cloning of the LAG-3 Gene. Genomic DNA clones were isolated from the LY67 library made with partial Mbo I-digested DNA from a human EBV-transformed B cell line inserted into λ2001 (10). The FD47 insert was labeled using the hexamer random priming method and used to screen 2 × 10^5 plaques of this human genomic DNA library. Nine positive plaques were isolated (GD1 to GD9), and phage DNA were characterized by restriction

mapping using the FD19 probe, which contains half of the protein coding region and the 3′ untranslated region. Two overlapping 16.4-kb Eco RI and 11.5-kb Hind III DNA fragments were generated and subcloned into the plasmid pUN121 (11) to give the GD3 Eco and GD1 Hind clones. Detailed restriction maps of these subclones were constructed and compared with the restriction map of the FDC sequence shown in Fig. 1. Various restriction fragments were obtained from ultra-low temperature agarose gels and subcloned into bacteriophage M13mp18 or M13mp19. The sequences of these fragments were determined from ssDNA using the dideoxy chain termination procedure (9) and the modified T7 polymerase (United States Biochemical Corp.). 17-base oligonucleotides, whose sequences were derived from either the cDNA FD19 (positions 497–513 in Fig. 1 *B*) or from the sequence of the 5′ flanking region of the LAG-3 gene (positions −421 to −437 in Fig. 3), were synthesized and used in the sequencing project.

Chromosomal Localization by In Situ Hybridization. In situ hybridization was performed according to a previously described method (12). The GD3Eco fragment (16.4 kb) was labeled with Bio-11-dUTP (Bethesda Research Laboratories, Gaithersburg, MD). The hybridized probe was visualized on chromosomes after indirect immunofluorescence; an antibiotin antibody was used to detect the biotinylated probe and a IgG fluorescein conjugate was used as a second antibody. Chromosomes were counterstained by propidium iodide and photographed.

Results and Discussion

Isolation of a LAG-3-specific cDNA Clone from the MB-F5IIIE5 Sublibrary. Experiments were designed to isolate cDNA clones expressed selectively in an IL-2-dependent $CD3^-$ NK cell line termed F_5. The MB-F_5 expression library was prepared using mRNA extracted from MB polyribosomes on sucrose gradients. Such libraries are enriched in RNA (<10% of the total mRNA) encoding either membrane-anchored or secreted proteins. MB-F_5 was screened with a cDNA probe obtained by subtraction of the MB F_5 mRNA with mRNA from a series of transformed cells, including Jurkat (T cell leukemia), Laz 388 (EBV-transformed B lymphoblastoid cell line), K562 (erythro-myeloid leukemia), and U937 (histiocytic origin). 0.5% (120 clones) of the λ GT10 recombinant phages present in the MB-F_5 library were selected with this screening procedure. One 400-bp clone, termed FD47, was studied because it hybridized to a 2-kb transcript expressed in the F_5 cells while being absent from Jurkat, Laz 388, K562, and U937 (data not shown).

Isolation and Structure of a Near Full-length LAG-3 cDNA. To obtain longer cDNA, we prepared three additional libraries from a TCR-γ/δ^+ cloned cell line. These cells were used because their mRNA was found to hybridize strongly with FD47. A first library was obtained from mRNA primed with oligo(dT) leading to the characterization of the 1,004-bp FD19 clone (Fig. 1 *a*). Because FD19 accounted for only half the size of the transcript, a second library was prepared using random hexamer oligonucleotide priming. The clones FD61 and FD101 obtained by this approach were further informative with respect to the upstream sequence of the LAG-3 mRNA. Finally, it is with a third library primed with a specific oligonucleotide corresponding to the 5′ region of the FD19 cDNA (nucleotides 704–688 of the sequence shown in Fig. 1 *b*) that a 5′ sequence including the initiation codon ATG was obtained in clone FD191.

Using these overlapping clones that spanned a total length of 1,871 bp, we established the nucleotidic sequence, termed FDC (Fig. 1 *b*), of the LAG-3 transcript. The LAG-3 mRNA has a long open reading frame encoding a 498-amino acid pro-

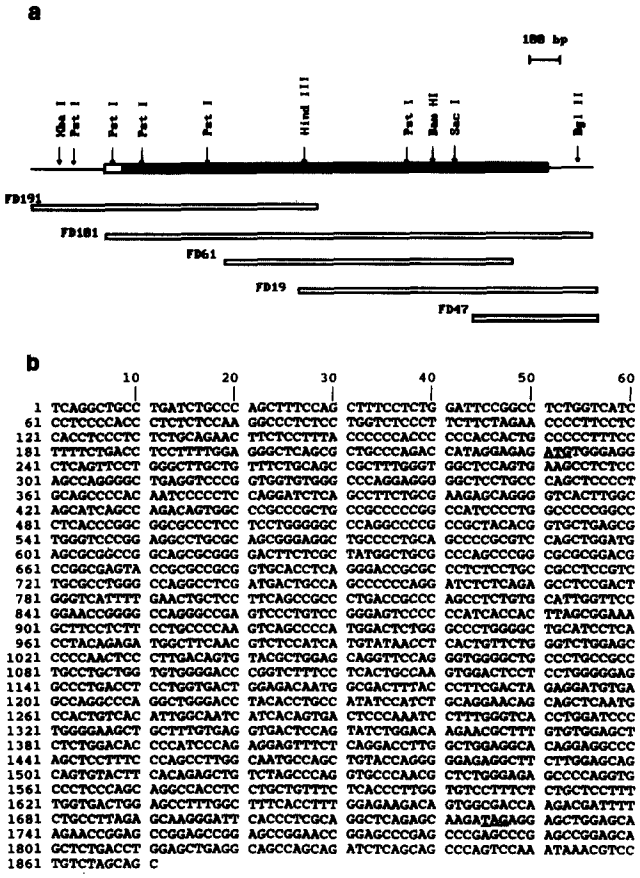
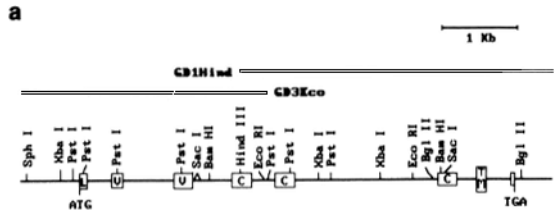


FIGURE 1. (*a*) Restriction map of FDC, a sequence derived by contiguity from overlapping human LAG-3 cDNA clones. FD47 and FD19 were isolated from independently derived oligo-dT-primed cDNA libraries. FD61 and FD101 were obtained from a hexamer random-primed cDNA library. FD191 is derived from an LAG-3-specific oligomer-primed (positions 704– 688 of sequence FDC) cDNA library. The protein-coding region is boxed. The shaded region represents the mature polypeptide after cleavage of the leader peptide. (*b*) Nucleotide sequence of lymphocyte cDNA clones encoding the human LAG-3 antigen. The methionine initiation codon and the stop codon are underlined. This sequence, termed FDC, was derived by contiguity from five overlapping LAG-3 cDNA (clones FD47, FD19, FD61, FD101, and FD191). These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession numbers X51984 and X51985.

tein. The sequence surrounding the presumed start codon (GAGATGT) is compatible with the Kozak consensus sequence (13). In the 3'-untranslated region, there is a single consensus polyadenylation site at position 1619. Using an alignment program based on a dynamic programming algorithm (14), the analysis of the LAG-3 cDNA revealed no significant local homologies with sequences available in the GenBank database.

Isolation and Structure of the LAG-3 Gene. Nine genomic clones (25–35 kb) were isolated from a human genomic DNA library with the FD19 probe. These clones were characterized by Southern blot hybridization and were found to display the same restriction map using the various LAG-3 probes. Overlapping fragments of two selected subclones, designated GD1 Hind and GD3 Eco (Fig. 2 *a*), were characterized by further restriction mapping and partial nucleotide sequencing. The map of the LAG-3 locus, from the 5' Sph I to the 3' Bgl II sites, is shown in Fig. 2 *a*. The LAG-3 gene spans ~6.6 kb and includes eight exons. A series of sequences (data not shown) was performed to study the precise exon-intron organization. All the introns displayed the appropriate GT and AG splicing donor/acceptor sites at their extremities (data not shown).



b

```
CCTGCAATGTCATTCTTTGAGCTCAGTTCCTCATCTCTGTCATGGAGAGCATTAGATTTCATGAATTCATACTAAGTGTCCAATACAGTGCTTAGCA -901
CGTAATGAAGCCTCAATACAATGTAGTTATTCTCCATGCCCCACAAAGCTGCATGCCTAGCCTCAGACCTACCATTTTTTGGGGTGCAGTAAGGCTTCCT -801
GTCCACCATGTTCCCAGGGACATTGTACTGATGGGTGGAAAGGCAGGTCTAAAGGGGTCACGAAGTTCTGGGAGGTTAAGGGAACGAGGAAGGAGATTGA -701
GCAACAAGGAAAGAGCTTGCCAAGAAGGAGGTGTGAATATTGGGACTGAGGAGGCAGCTTAGAGATGGGCAAGGGGGCAGTTCCAGGCAGAAATGGTTCG -601
TGGAGGCAGAAGGTCCCTGGGAGAGGGAGCAGTCTGGAGGGTGGGGCAGGGGCGAGGAGGGGGAGGTGGGGAGACCCAGGACTGAGGAAGTAAACAAGGG -501
GAGCGCCACCACAGAGGTGGAGAGGTGGAGGGTGCTGCTGCTGGGAATCAACCCCCTCAGACTTTCCACTGCGAAGCGAAACCGTAAGCCCTGGGGTGCG -401
GGGGGCGGGCCGGGAGGAGGGGAAGTGGGGAAGGTGGAGGGAAGGCCGGGCACAGGGGTGAAGGCCCAGAGACCAGCAGAACGGCATCCCAGCCACGACG -301
GCCACTTTGCTCTGTCTGCTGTCCGCCACGGCCCTGCTCTGTTCCCTGGGACACCCCCGCCCCCACCTCCTCAGGCTGCCTGATCTGCCCAGCTTTCCAG -201
CTTTCCTCTGGATTCCGGCCTCTGGTCATCCCTCCCCACCCTCTCTCCAAGGCCCTCTCCTGGTCTCCCTTCTTCTAGAACCCCTTCCTCCACCTCCCTC -101
TCTGCAGAACTTCTCCTTTACCCCCCACCCCCCACCACTGCCCCCTTTCCTTTTCTGACCTCCTTTTGGAGGGCTCAGCGCTGCCCAGACCATAGGAGAG -1
+1
ATG
```

FIGURE 2. (*a*) Structure of the human LAG-3 gene in chromosomal DNA. The signal peptide (L), V-like (V), C-like (C), transmembrane (TM), and cytoplasmic (CYT) regions are boxed. The map was constructed by single and double endonuclease digestions of λ 2001 clones GD1 and GD3 and their subclones GD3 Eco and GD1 Hind, as well as from nucleotide sequencing. (*b*) Nucleotide sequence of the promoter region of the human LAG-3 gene. The arrow indicates the first nucleotide of the FDC cDNA sequence. The ATG (+1), the inverted CCAAT box, and the GC box (Sp1-binding site) are underlined.

In the promoter region (Fig. 2 *b*), no TATA box is found upstream of the 239-bp 5′ untranslated sequence. An Sp 1 binding site that contains the typical GGGCGG core hexanucleotide (15) is located at position −389 from the translation initiation site. An ATTGG sequence (CCAAT box in reverse), known to be an important component of several promoters (16), is present at position −662.

To further characterize the LAG-3 gene, DNAs from the K562 myeloid tumor cell line and from polyclonal IL-2-dependent T and NK cell lines were digested by either Eco RI, Hind III, Bam HI, or Xba I restriction enzymes. Southern blot hybridizations were performed using the FDC probe (1,871 bp) that had been constructed by fusing the 5′ Eco RI/Hind III fragment of clone FD191 to the 3′ Hind III/Eco RI fragment of clone FD19. Three Eco RI fragments (2, 8.2, and 10 kb), two Hind III fragments (5.7 and 9.5 kb), three Bam HI fragments (2.8, 4, and 13 kb), and three Xba I fragments (3, 4, and 6 kb) were detected (data not shown). These results, in accordance with the restriction map of the genomic clones, indicated that the LAG-3 gene is present in the human haploid genome as a single copy gene.

Expression of the LAG-3 Gene. The 1,004-bp clone FD19 was used as a probe to analyze the expression of the LAG-3 gene in a series of hematopoietic and non-hematopoietic cells (Fig. 3). There was no message in the transformed cell lines of T, B, and myeloid origin (Jurkat, Laz 388, K562, U937; Fig. 3 *a*, lanes *1–4*) that

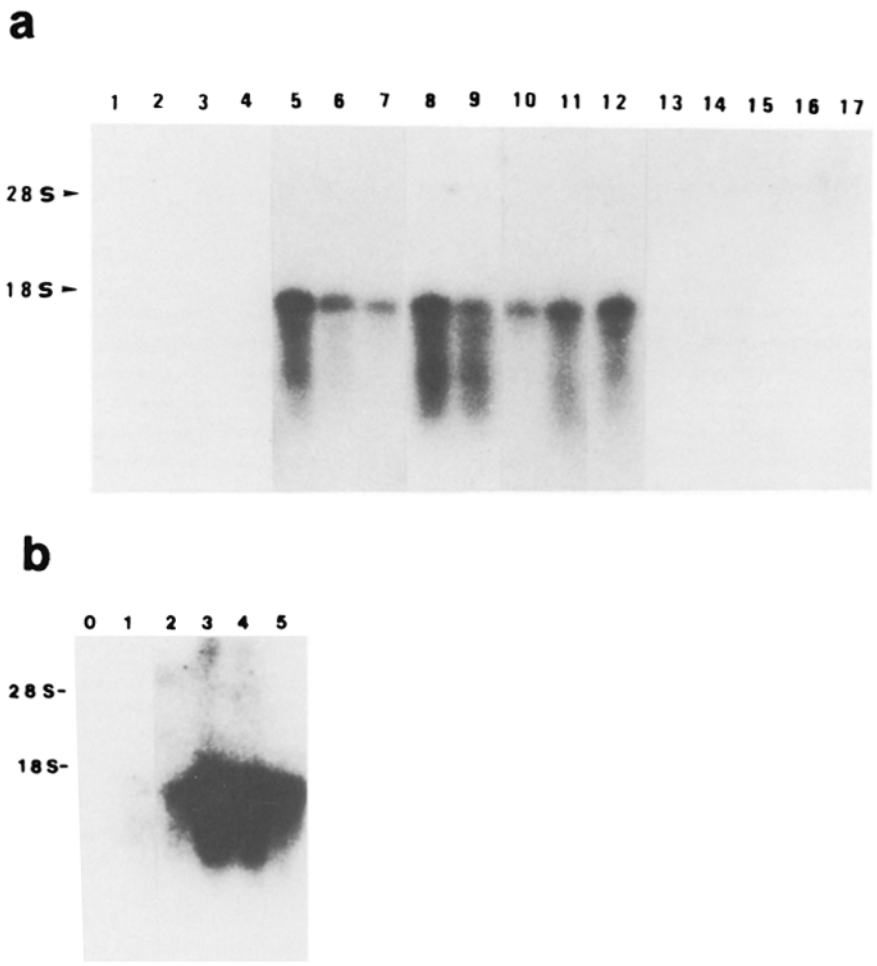


FIGURE 3. Expression of the LAG-3 gene. (*a*) Cellular distribution as deduced from RNA blot hybridization. Total cytoplasmic RNA were prepared from the following cells: lane *1*, Jurkat; lane *2*, Laz 388; lane *3*, K562; lane *4*, U937; lanes *5*, *6*, and *7*, three $CD3^-$ NK cell lines (F55IIIE5, SIIH4, SIIIG5); lanes *8*, *9*, *10*, and *11*, four $CD3^+$ TCR-α/β^+ cloned cell lines F55C8 and F55B4 ($CD4^+$), and SIF8 and F55IIIG5 ($CD8^+$); lane *12*, the $CD3^+$ TCR-γ/δ^+ BK clone; lane *13*, macrophages derived from PBL; lane *14*, purified T cells; lane *15*, SKN-AS, a neuroblastoma cell line; lane *16*, brain tissue; lane *17*, PLC-PRF5, a hepatoma cell line. (*b*) Kinetics of LAG-3 expression in T cell blasts. Lane *0*, uninduced lymphocytes; lanes *1–5*, PBL activated with PHA for 1, 2, 3, 4, and 5 d, respectively.

have been used for the subtraction procedures. Additional T cell tumors including CEM and MOLT-4 were also found to be negative (data not shown), as well as peripheral blood monocytes (Fig. 3 *a*, lane *14*), two cultured neuroblastoma cell lines (one shown in lane *15*), fresh brain tissue (lane *16*), and a tumor cell line of hepatic origin (lane *17*).

Resting lymphocytes (Fig. 3 *b*, lane *1*) as well as resting purified T cells (Fig. 3 *a*, lane *13*) were negative. A series of normal T and NK IL-2-dependent clones and polyclonal cell lines described previously (2, 17, 18) were assessed. The transcription

of the LAG-3 gene was detected as a single species mRNA of ~2 kb in all these activated lymphocytes. Fig. 3 *a* displays data representative of the various subpopulations tested: lanes *5–7*, three $CD3^-$ NK cell lines; lanes *8–11*, four $CD3^+$ TCR-α/β^+ lines, including two $CD4^+$ and two $CD8^+$ cells; lane *12*, 1 $CD3^+$ TCR-γ/δ^+ cell line.

These data supported the view that the LAG-3 protein corresponds to a broadly expressed (i.e., not restricted to a unique subset) lymphocyte activation antigen. The kinetic parameters of this expression were tested after PHA stimulation of PBL. Fig. 3 *b* shows that the expression of the LAG-3 gene is undetectable in resting PBL and peaks at days 3–4 after PHA stimulation.

Structure of the LAG-3 Protein. The characteristics of the LAG-3 protein, shown in Fig. 4, have been deduced from the structure of the gene and from the analysis of its translation product. It appears as a type I membrane protein encoded by 498 amino acids (aa). The mature protein includes 470 aa with a predicted molecular mass of 51,295 daltons and a pI of 10.9. The hydrophobic leader peptide (28 aa) is encoded by both exon I (19 aa) and exon II (9 aa out of 50). The extra-cellular region is encoded by exons II (41 aa out of 50), III (101 aa), IV (90 aa), V (92 aa), and VI (81 aa); the transmembrane region (TM) by exon VII (44 aa); and the highly charged cytoplasmic region by exon VIII (21 aa). The extracellular region includes eight cysteine residues and four potential N-linked glycosylation sites (Asn-X-Ser, Thr).

The peptidic stretch encoded by exons II and III corresponds to a V-SET IgSF domain (1), including β strands A, B, C, C′, C″, D, E, F, and G, with two unusual features. First, this V domain includes an extra loop of ~30 amino acids encoded by the first part of exon III. This loop shown in Fig. 4 *b* joins β strand C to β strand C′ and contains 10 proline residues. It seems that such an insertion might be compatible with IgSF fold since it does not disrupt the core of the fold that is considered to consist of β strands A, B, E and G, F, C (19). Differences in V and C domains occur in the middle of the fold, and also an alternatively spliced exon that can be inserted in the coding sequence for N-CAM domain 4 occurs in this region (20). The second unusual point is that the downstream cysteine of domain 1 seems likely to be in the β strand G position rather than in β strand F (residue 121), as is almost invariably the case (Fig. 4 *b*). The pattern of Asp × Gly × Tyr × Cys is very characteristic in the β strand F position, and this is present, except that an Ala is in the Cys position. It seem possible than a disulphide bond as shown in Fig. 4 *b* might form and it should be noted that an unusual disulphide bond of a different kind has been reported in the V-like domain of CD8α chain (21).

An Arg-Gly-Asp (RGD) sequence is found in β strand E (Fig. 4 *b*). This motif is known to represent a potential adhesiotope (22), but it is dubious that its forms the core of a major binding site since, at this position (+101), such a sequence is probably exposed within the IgSF fold.

Exons IV, V, and VI encode IgSF-related domains (1) with 51, 50, and 42 aa, respectively, between the two conserved cysteine residues. These three domains have C-type folds and show sequence patterns characteristic of the C2-SET (1). They were scored against C2-SET sequences with the ALIGN program (23, 24): of 57 sequences tested, scores of >3 SD were obtained 32, 41, and 11 times for domain 2, 3, and 4, respectively. Domain 4 belongs to the truncated C2-SET since it lacks strand D.

LAG-3 domains 1 and 2 were aligned by eye with domains 3 and 4, taking into

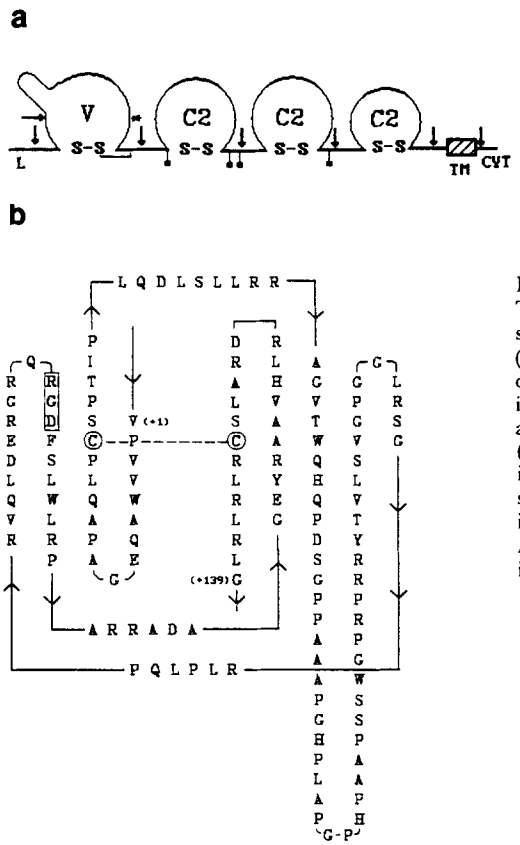


FIGURE 4. (*a*) Schematic diagram of LAG-3. The domains are designed as L (leader), V (V set IgSF domain), C2 (C2 set IgSF domain), TM (transmembrane), and CYT (cytoplasmic). The cysteine residues that form disulfide linkages are indicated. The position of introns is marked by arrows. The carbohydrate sites (▲) and the RGD (*) sequence (putative cell attachment sites) are indicated. (*b*) A model for LAG-3 domain 1. The sequence of the first Ig domain (aa +1 to +139) is displayed according to the format used by Amzel and Poljak (36). The disulphide bond is indicated and the RGD sequence is boxed.

account identities and structural considerations (Fig. 5 *a*). Since domain 1 contains an extra loop of sequence, the alignment was started at aa 91 in this domain and at aa 276 in domain 3. Out of 129 possible matches between residues, there are 34 identities, 35 similarities, and 9 breaks (alignment score + 8.5 SD). Moreover, in β strand F of domains 2 and 4 there is a sequence W × C that is most unusual at this position, where the sequence Y or F × C is commonly found (1). Taken together, the data suggests that LAG-3 has evolved by gene duplication from a pre-existing two-domain structure resembling that of an Ig L chain.

One of the most prominent characteristics of LAG-3 is its structural relationship with CD4. Indeed, an alignment of sequences with some notable patches of identity can be made as shown in Fig. 5 *b*. Two large gaps are inserted to allow for the extra loop of sequence in LAG-3 domain 1 and for the fact that CD4 domain 3 is of V-SET type, while LAG-3 domain 3 is of C2-SET type. The patches of similarity include the start of domain 1 (nine identities and six similarities out of 17 possible matches), and the very unusual W × C sequence in LAG-3 domains 2 and 4, which are also present at the corresponding positions in CD4. This sequence motif is not seen in an equivalent position in any other IgSF domains. Overall, there are 87 identities and 82 similarities out of 338 residues aligned (19 breaks) when the extracellular regions of LAG-3 and rat CD4 are compared. As in the LAG-3 structure, notable

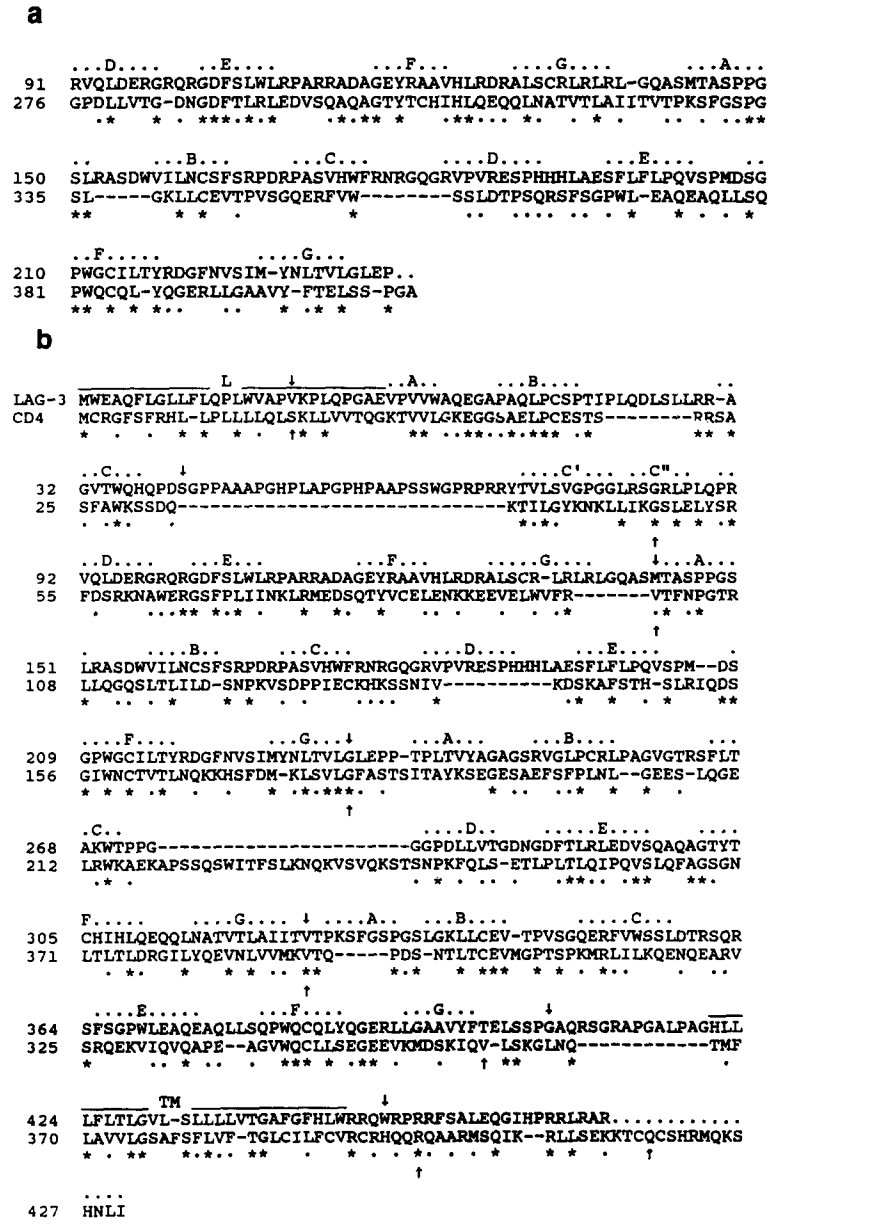


FIGURE 5. (*a*) Sequence alignments showing internal homology in the LAG-3 molecule. Amino acid sequences in domain 1 (starting at position 119 after the extra loop) and domain 2 are lined up with the corresponding positions in domains 3 and 4. Identities are indicated by (*) and similarities by (·). (*b*) Alignments of LAG-3 and rat CD4 (37) amino acid sequences, and position of introns. The dotted lines above the sequences indicate the positions of putative β strands in the four IgSF domains. The strands are labeled as in reference 1. The putative leader (L) and transmembrane (TM) sequences are marked by a straight line above the sequence. The position of introns is marked by arrows above the sequence (for LAG-3) and below the sequence (for CD4 according to reference 38 for human CD4). Identities are indicated by an asterisk and similarities by a period.

patches of internal sequence homologies are found in the CD4 molecule between domains 1 and 2, as well as between domains 2 and 4 (25). More generally, the exon-intron organization of LAG-3 and CD4 is very similar: both genes include an intron within the first IgSF domain, and the position of the introns (marked by arrows in Fig. 5 *b*) in LAG-3 is very close to that of CD4.

It has been previously suggested that CD4 has evolved by gene duplication from a pre-existing two-IgSF structure (25). The present findings strengthen this hypothesis and support the view that CD4 and LAG-3 may have shared a common four-domain evolutionary ancestor.

Chromosomal Location of the LAG-3 Gene. In light of the likely relationship between LAG-3 and CD4 genes, it was of interest to compare their location on the genome. The human CD4 gene has been shown to be located on chromosome 12 (26, 27), with a regional assignment corresponding to band pter-p12 (27). In situ hybridization was performed on human chromosomes using a 16.4-kb biotinylated LAG-3 genomic probe and indirect immunofluorescence. Green fluorescent spots were recurrently observed on the distal segment of the short arm of a middle-sized chromosome (Fig. 6, *a* and *b*). R-banding of the metaphases showing these fluorescent spots indicated that the LAG-3 probe is localized on chromosome 12 on the p13.3 band (Fig. 6 *c*). More precisely, the LAG-3 gene position on chromosome 12 is likely to be assigned to the p13.32 band because green spots were not distally located on p13.3 in the majority of the cells examined. The frequency of metaphases with at least one spot on one of the four chromatides of chromosomes 12 was 75%. Very few other spots (two or three per cell, according to the probe concentration) were observed in the karyotype on the other chromosomes.

Thus, the positions of both LAG-3 and CD4 genes in the human genome are very similar since the regional assignment of CD4 (pter-p12) includes band p13.3. This observation confirmed that both genes are likely to derive from a common evolutionary ancestor.

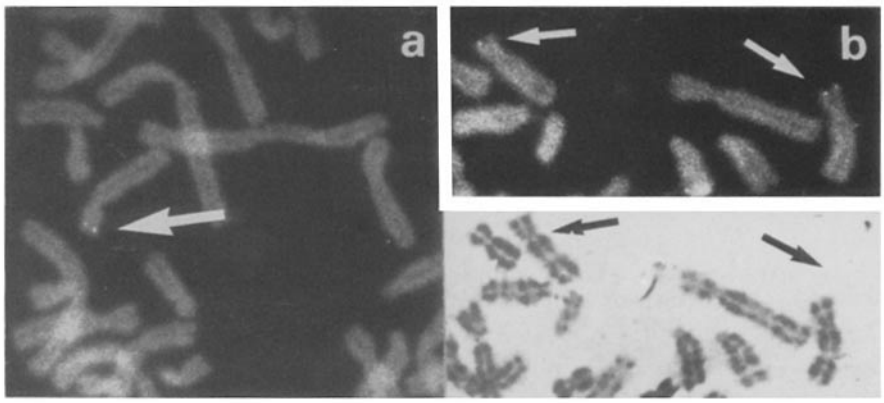


FIGURE 6. Chromosomal localization of the LAG-3 gene by in situ hybridization of biotinylated LAG-3 probe on human chromosomes. Arrows indicate chromosome 12. (*a* and *b*) Fluorescein detection of hybridized sites and propidium iodide counterstaining. (*c*) R banding of chromosomes shown in *b*.

In light of the present data, one cannot exclude that the LAG-3 protein may have biological properties similar to that of CD4 with respect to interactions with MHC gene products (28, 29) or with viral proteins (30, 31).

More generally, LAG-3 may function like many other IgSF molecules as a ligand for membrane-anchored or soluble proteins. Among IgSF members whose expression is upregulated upon cell activation, examples include ICAM-1, known to be involved in cell-cell interactions (32), and IL-1-R or IL-6-R, which function as receptors for growth factors (33, 34). Note that ICAM-1 has also been shown recently to represent a major receptor for rhinoviruses (35). Further studies will be needed to test these working hypotheses on the LAG-3 function.

Summary

We have identified a novel human gene of the Ig superfamily, designated LAG-3. Expression of this gene is undetectable in resting PBL, while it is found (a 2-kb message) in activated T and NK cells. The LAG-3 gene includes eight exons; the corresponding cDNA encodes a 498-amino acid membrane protein with four extracellular IgSF domains. The first one belongs to the V-SET; it is particular since it includes an extra loop in the middle of the domain and an unusual intrachain disulphide bridge. The three other domains belong to the C2-SET. Strong internal homologies are found in the LAG-3 molecule between domains 1 and 3, as well as between domains 2 and 4. It is therefore likely that LAG-3 has evolved by duplication of a preexisting gene encoding a two IgSF-domain structure. The compared analysis of LAG-3 and CD4, with respect to both their peptidic sequence as well as their exon/intron organization, indicated that the two molecules are closely related. This point is strengthened by the finding that both genes are located on the distal part of the short arm of chromosome 12.

We thank Alan F. Williams for his precious help in the detailed comparison between LAG-3 and other IgSF molecules, particularly CD4.

Received for publication 21 December 1989.

References

1. Williams, A. F. 1987. A year in the life of the immunoglobulin superfamily. *Immunol. Today.* 8:298.
2. Nowill, A., P. Moingeon, A. Ythier, M. Graziani, F. Faure, L. Delmon, M. Rainaut, F. Daffos, C. Bohuon, and T. Hercend. 1986. Natural killer clones derived from fetal (25 wk) blood: probing the human T cell receptor with WT31 monoclonal antibody. *J. Exp. Med.* 163:1601.
3. Mechler, B., and T. H. Rabbitts. 1981. Membrane-bound ribosomes of myeloma cells IV. mRNA complexity of free and membrane-bound polysomes. *J. Cell Biol.* 88:29.
4. Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidilic acid cellulose. *Proc. Natl. Acad. Sci. USA.* 69:1408.
5. Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene.* 25:263.
6. Huynh, T. V., R. A. Young, and R. W. Davis. DNA Cloning: A practical approach. D. Glover, editor. IRL Press. Oxford. 49-78.
7. Triebel, F., M. Graziani, F. Faure, S. Jitsukawa, and T. Hercend. 1987. Cloned human

CD3⁻ lymphocytes with NK-like activity do not express nor rearrange TCR gamma genes. *Eur. J. Immunol.* 17:1209.

8. Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6.
9. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.
10. Dariavach, P., G. Lefranc, and M. P. Lefranc. 1987. Human Immunoglobulin Clambda6 gene encodes the Kern + Oz-lambda chain and Clambda4 and Clambda5 are pseudogenes. *Proc. Natl. Acad. Sci. USA.* 84:9074.
11. Triebel, F., R. Breathnach, M. Graziani, T. Hercend, and P. Debre. 1988. Evidence for expression of two distinct T cell-receptor beta chain transcripts in a human diphtheria toxoid specific T cell clone. *J. Immunol.* 140:300.
12. Viegas-Pequignot, E., B. Dutrillaux, H. Magdelenat, and M. Coppey-Mosian. 1989. Mapping of single-copy DNA sequences on human chromosomes by in situ hybridization with biotinylated probes: enhancement of detection sensitivity by intensified-fluorescence digital imaging microscopy. *Proc. Natl. Acad. Sci. USA.* 86:582.
13. Kozak, M. 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 12:857.
14. Goad, W. B., and M. I. Kanehisa. 1982. Pattern recognition in nucleic acid sequences. I. A. general method for finding local homologies and symmetries. *Nucleic Acids Res.* 10:247.
15. Kadonaga, J. T., K. A. Jones, and R. Tijan. 1986. Promoter-specific activation of RNA polymerase II transcription by Sp1. *TIBS (Trends Biochem. Sci.)* 11:20.
16. Mcknight, T. D., D. C. Alexander, M. S. Babcock, and R. B. Simpson. 1986. Nucleotide sequence and molecular evolution of two tomato genes encoding the small subunit of ribulose-1, 5-bisphosphate carboxylase. *Gene (Amst.).* 48:23.
17. Triebel, F., F. Faure, M. Graziani, S. Jitsukawa, M. P. Lefranc, and T. Hercend. 1988. A unique V-J-C rearranged gene encodes a γ protein expressed on the majority of $CD3^+$ TCR-α/β-circulating lymphocytes. *J. Exp. Med.* 167:694.
18. Faure, F., S. Jitsukawa, F. Triebel, and T. Hercend. 1988. CD3/Ti gamma A: a functional gamma receptor complex expressed on human peripheral lymphocytes. *J. Immunol.* 140:1372.
19. Lesk, A. M., and C. Chothia. 1982. Evolution of proteins formed by beta-sheets. II. The core of the Ig domains. *J. Mol. Biol.* 160:325.
20. Santoni, M. J., D. Barthels, G. Vopper, A. Boned, C. Goridis, and W. Wille. 1989. Differential exon usage involving an unusual splicing mechanism generates at least eight types of NCAM cDNA in mouse brain. *EMBO (Eur. Mol. Biol. Organ.) J.* 8:395.
21. Kirszbaum, L., J. A. Sharpe, N. Goss, J. Lahnstein, and I. D. Walker. 1989. The alpha-chain of murine CD8 lacks an invariant Ig-like disulfide bond but contains a unique intrachain loop instead. *J. Immunol.* 142:3931.
22. Ruoslahti, E., and M. D. Pierschbacher. 1986. Arg-Gly-Asp, a versatile cell recognition signal. *Cell.* 44:517.
23. Dayhoff, M. O., W. C. Barker, and L. T. Hunt. 1983. Establishing homologies in protein sequences. *Meth. Enzymol.* 91:524.
24. Williams, A. F., and A. N. Barclay. 1988. The immunoglobulin superfamily domains for cell surface recognition. *Annu. Rev. Immunol.* 6:381.
25. Williams, A. F., S. J. Davis, Q. He, and A. N. Barclay. 1989. Structural diversity in domains of the immunoglobulin superfamily. *Cold Spring Harbor Symp. Quant. Biol.* In press.
26. Kozbor, D., J. Finan; P. C. Nowell, and C. M. Croce. 1986. The gene encoding the T4 antigen maps to human chromosome 12. *J. Immunol.* 136:1141.
27. Isobe, M., K. Huebner, P. J. Maddon, D. R. Littman, R. Axel, and C. M. Croce. 1986.

The gene encoding the T-cell surface protein T4 is located on human chromosome 12. *Proc. Natl. Acad. Sci. USA.* 83:4399.

28. Doyle, C., and J. L. Strominger. 1987. Interaction between CD4 and class II MHC molecules mediates cell adhesion. *Nature (Lond.).* 330:256.
29. Mazerolles, F., A. Durandy, D. Piatier-Tonneau, D. Charron, L. Montagnier, C. Auffray, and A. Fischer. 1988. Immunosuppressive properties of synthetic peptides derived from CD4 and HLA-DR antigens. *Cell.* 55:497.
30. Clayton, L. K., R. E. Hussey, R. Steinbrich, H. Ramachandran, Y. Husain, and E. Reinherz. 1988. Substitution of murine for human CD4 residues identifies amino acids critical for HIV-gp120 binding. *Nature (Lond.).* 335:363.
31. Peterson, A., and B. Seed. 1988. Genetic analysis of monoclonal antibody and HIV binding sites on the human lymphocyte antigen CD4. *Cell.* 54:65.
32. Simmons, D., M. W. Makgoba, and B. Seed. 1988. ICAM, an adhesion ligand of LFA-1, is homologous to the neural cell adhesion molecule NCAM. *Nature (Lond.).* 331:624.
33. Sims, J. E., C. J. March, D. Cosman, M. B. Widmer, H. R. MacDonald, C. J. McMahan, C. E. Grubin, J. M. Wignall, J. L. Jackson, S. M. Call, D. Friend, and S. K. Dower. 1988. cDNA expression cloning of the IL-1 receptor, a member of the immunoglobulin superfamily. *Science (Wash. DC).* 241:585.
34. Yamasaki, K., T. Taga, Y. Hirata, H. Yawata, Y. Kawanishi, B. Seed, T. Taniguchi, T. Hirano, and T. Kishimoto. 1988. Cloning and expression of the human interleukin-6 (BSF-2/IFNbeta 2) receptor. *Science (Wash. DC).* 241:825.
35. Greve, J. M., G. Davis, A. M. Meyer, C. P. Forte, S. C. Yost, C. W. Marlor, M. E. Kamarck, and A. McClelland. 1989. The major human rhinovirus receptor is ICAM-1. *Cell.* 56:839.
36. Amzel, L. M., and R. J. Poljak. 1979. Three dimensional structure of Igs. *Annu. Rev. Biochem.* 48:961.
37. Clark, S. J., W. A. Jefferies, N. Barclay, J. Gagnon, and A. F. Williams. 1987. Peptide and nucleotide sequences of rat CD4 (W3/25) antigen: evidence for derivation from a structure with four immunoglobulin-related domains. *Proc. Natl. Acad. Sci. USA.* 84:1649.
38. Maddon, P. J., S. M. Molineaux, D. E. Maddon, K. A. Zimmerman, M. Godfrey, F. W. Alt, L. Chess, and R. Axel. 1987. Structure and expression of the human and mouse T4 genes. *Proc. Natl. Acad. Sci. USA.* 84:9155.