Characterization and Mapping to Human Chromosome 8q24.3 of Ly-6-Related Gene 9804 Encoding an Apparent Homologue of Mouse TSA-1

Xiaochuan Shan, Annie Bourdeau, Angela Rhoton, Dan E. Wells, Edward H. Cohen, Bryan E. Landgraf and Roger G. E. Palfree

*J Immunol* 1998; 160:197-208; 
http://www.jimmunol.org/content/160/1/197

**References** This article cites 57 articles, 30 of which you can access for free at: http://www.jimmunol.org/content/160/1/197.full#ref-list-1

**Subscription** Information about subscribing to *The Journal of Immunology* is online at: http://jimmunol.org/subscription

**Permissions** Submit copyright permission requests at: http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts** Receive free email-alerts when new articles cite this article. Sign up at: http://jimmunol.org/alerts
Characterization and Mapping to Human Chromosome 8q24.3 of Ly-6-Related Gene 9804 Encoding an Apparent Homologue of Mouse TSA-1¹,²

Xiaochuan Shan,* Annie Bourdeau,³* Angela Rhoton,† Dan E. Wells,‡ Edward H. Cohen,‡ Bryan E. Landgraf,‡ and Roger G. E. Palfree*⁴

The 9804 gene, which encodes a human Ly-6 protein most similar to mouse differentiation Ag TSA-1/Sca-2, has also been called RIG-E. Like mouse TSA-1, it has a broad tissue distribution with varied expression levels in normal human tissues and tumor cell lines. Like some members of the murine Ly-6 family, the 9804 gene is responsive to IFNs, particularly IFN-α. Overlapping genomic fragments spanning the 9804 gene (5543 bp) have been isolated and characterized. The gene organization is analogous to that of known mouse Ly-6 genes. The first exon, 2296 bp upstream from exon II, is entirely untranslated. The three coding exons (II, III, and IV) are separated by short introns of 321 and 131 bp, respectively. Primers were developed for specific amplification of 9804 gene fragments. Screening of human-hamster somatic cell hybrids and yeast artificial chromosomes (YACs) indicated that the gene is distal to c-Myc, located in the q arm of human chromosome 8. No positives were detected from the Centre d’Etude du Polymorphisme Humain mega-YAC A or B panels, nor from bacterial artificial chromosome libraries; two positive cosmids (c101F1 and c157F6) were isolated from a human chromosome 8 cosmid library (LA008C01). Fluorescence in situ hybridization of metaphase spreads of chromosome 8, containing hybrid cell line 706-B6 clone 17 (CL-17) with cosmid c101F1, placed the 9804 gene close to the telomere at 8q24.3. This mapping is significant, since the region shares a homology with a portion of mouse chromosome 15, which extends into band E where Ly-6 genes reside. Moreover, the gene encoding E48, the homologue of mouse Ly-6 molecule ThB, has also been mapped to 8q24.

enhanced by IFN-α6 and IFN-γ (24, 25). Enhanced Ly-6A/E expression on activated lymphocytes or IL-1-treated YAC-1 cells is mediated through IFN-γ production, while TNF synergizes with IFN-γ in this regulation of Ly-6A/E expression (26, 27). Lymphocyte activation also induces TSA-1 expression on T cells (28).

Immunogenetic and restriction fragment length polymorphism (RFLP) analyses have demonstrated that Ly-6 molecules are encoded by genes that are tightly linked on mouse chromosome 15, distal to Myc-1, and proximal to Sis (29, 30), while in situ hybridization placed the loci in band 15E (31). On the basis of synteny, it has been proposed that the human homologues of Ly-6 genes would reside on chromosome 8, distal to Myc-1, or on chromosome 22q near Sis (29). Long range physical mapping has placed Ly-6A/E and Ly-6C in a 630-kb genomic fragment along with about 20 distinct Ly-6-related genes or pseudogenes that have regions of sufficient sequence similarity to cross-hybridize with Ly-6A cDNA probes (32). Included in this category of cross-hybridizing genes, which were first revealed as a complex multiple band pattern on Southern blots, are Ly-6F and Ly-6G. Two other members of the Ly-6 family, ThB and TSA-1 (Sca-2) are more divergent (5, 33, 34), and there is no evidence for multiple ThB- or TSA-1-like genes at the Ly-6 locus, although a TSA-1-related sequence (TSA-1-rs1) has been located on chromosome 12 (5). While immunogenetic analysis has demonstrated a very tight linkage between these and the other Ly-6 genes, the positions of the ThB and TSA-1 genes relative to Ly-6A as well as the 630-kb segment of chromosome 15 remain to be determined.

The discovery in humans of 9804 and E48 (13, 35), apparent homologues of the mouse Ly-6 molecules TSA-1 and ThB, respectively, and the report of a TSA-1/Sca-2 gene homologue in chickens (CHKSCA2A, GenBank accession no. L34554) suggest that whatever biologic functions the various mouse Ly-6 molecules possess, at least some of those functions may be suberved by related molecules in other species. Furthermore, on the assumption that the relationship to mouse Ly-6 genes extends to their local organization within the genome, it is likely that 9804 and E48 genes will be found closely linked in a human Ly-6 multigene cluster. To extend our study of 9804, and of human Ly-6-like molecules in general, and to provide information and material that can be used in studying the control of 9804 gene expression, we have isolated and characterized genomic fragments spanning all 9804 exons and over 1 kb of promoter region, and we have initiated a study of its local genomic context.

Materials and Methods

Cell culture

The U-937 human histiocytic lymphoma cell line (ATCC CRL 1593, American Type Culture Collection, Rockville, MD) and PBMCs (from normal healthy volunteers) were maintained or assayed in RPMI 1640 medium supplemented with 10% heat-inactivated FBS, 2 mM L-glutamine, 50 U/ml penicillin, and 50 μg/ml streptomycin. U-937 cells were subcloned, via limiting dilution, at a density of 0.1 cell/well in multiple 96-well tissue culture plates. The subcloned cells from growth-positive wells were expanded and evaluated for costimulatory activity as described below.

The CTLL-2 indicator cell line generously supplied by K. A. Smith (Division of Allergy and Immunology, Cornell University Medical Center, New York, NY) was maintained in RPMI 1640 medium containing 10% heat-inactivated FBS, 2 mM L-glutamine, 10 mM nonessential amino acids, 10 mM sodium pyruvate, 50 mM 2-ME, 50 U/ml penicillin, and 50 μg/ml streptomycin. This medium was further supplemented with 15% conditioned medium from Con A/PHA-treated rat splenocytes.

Costimulation assay

Human T cells, from the heparinized whole blood of healthy volunteers, were purified using T cell enrichment columns (R&D Systems, Minneapolis, MN) following the manufacturer’s directions. The assessment of U-937 costimulatory activity was accomplished by coculturing these cells with purified T cells (10^6/well) in a 96-well flat-bottom microtiter tray precoated with goat anti-mouse Fab (Sigma Chemicals, St. Louis, MO) at 10 μg/ml. Any unbound goat anti-mouse Fab was removed by extensively washing the culture wells with PBS before the addition of the T cells, U-937 cells, and 0.1 μg/ml of soluble anti-CD3 (OKT3, Ortho Diagnostics, Raritan, NJ). Culture supernatants, harvested after 24 h, were assessed for IL-2 production using the murine CTLL-2 cell line according to published methods (36).

Preparation of total and poly(A)^+ mRNA

The preparation of total RNA was based upon the procedures of Chomczynski et al. (37), Poly(A)^+ mRNA was isolated using an mRNA separation kit (Clontech, Palo Alto, CA) following the manufacturer’s instructions.

Generation of 32P-labeled DNA probes

Single-stranded cDNA, prepared from 1 μg of U-937 clone 6 poly(A)^+ mRNA, was generated using a cDNA cycle kit (Invitrogen, San Diego, CA) following the manufacturer’s directions. 32P-labeled DNA complementary to this was made using a random-primed DNA labeling kit (Boehringer Mannheim, Indianapolis, IN). A subtracted probe, enriched for unique or more highly expressed parental U-937 DNA sequences relative to clone 6 sequences, was generated using a Subtractor 1 kit (Invitrogen) following the manufacturer’s instructions. All other 32P-labeled probes were prepared from double stranded DNA using the random primed DNA labeling kit.

cDNA library construction and screening

A U-937 cDNA library was made in the vector pCDNAI (Invitrogen). The library, which had approximately 1.6 × 10^9 independent clones, was amplified once, frozen in aliquots, and stored at −70°C. For screening, a sample of the library was plated at approximately 7000 colonies/100-mm diameter plate. Duplicate filters of these colonies were generated and prepared for hybridization with two different probes. All filters were prehybridized in Cak solution (50 mM NaH2PO4 (pH 7.4), 0.75 M NaCl, 6 mM EDTA, 1% SDS, 0.1% Ficoll, 0.1% polyvinyl pyrrolidone, 0.1% BSA, Torula yeast RNA (Sigma)) at 100 μg/ml for 0.5 to 2 h at 65°C. One of each duplicate filter was hybridized with the clone 6 cDNA probe at 2 × 10^6 cpm/ml of Cak for 16 to 24 h at 65°C. The other filter was hybridized with the subtracted cDNA probe at 3.0 to 1.0 × 10^6 cpm/ml. Hybridizations were performed in Cak for 16 to 24 h at 65°C. Autoradiography of the filters was performed using Biomax MR film (Eastman Kodak, Rochester, NY). Colonies were picked from plates on the basis of having stronger signals with the clone 6 subtracted probe than with the clone 6 probe. Bacteria from selected colonies were plated at low density (5 × 100/100-mm plate) and rescreened with the two probes as described above. A tertiary rescreening was done to ensure purification of the clones.

cDNA analysis

Dideoxynucleotide sequencing was performed using Sequenase Version 2.0 DNA Sequencing kits (United States Biochemical, Cleveland, OH) following the manufacturer’s instructions. Oligonucleotides used as sequencing primers were either purchased from Invitrogen or were custom made by National Biosciences (Plymouth, MN). A large and a small fragment of the 9804 cDNA, released upon digestion with BstXI, were subcloned as blunt-ended fragments into EcoRV-digested pBluescript II SK’ (Stratagene, La Jolla, CA). These two constructs were used for 9804 sequencing, while the construct containing the large BstXI fragment was used also in the generation of random-primed 32P-labeled probes.

Interferon treatment of U-937 cells and PBMCs

U-937 cells and PBMCs (5 × 10^5 cells/ml) were incubated in culture medium in the presence or absence of either 100 U/ml IFN-γ (Biosource, Carpenteria, CA) or 1000 U/ml IFN-α (Biosource) for 20 h. The cells were collected for RNA extraction and Northern blot analysis as described below.

Northern analyses

Human multiple-tissue Northern blots were purchased from Clontech and processed following the manufacturer’s instructions. RNA from various adult and fetal tissues was kindly supplied by A. Bateman (Endocrine Laboratory, Royal Victoria Hospital, Montreal, Canada). Northern analyses were performed on total RNA or poly(A)^+ mRNA.
Genomic fragment analysis

A human genomic cosmID clone, CMC9 (CytoMed, unpublished observation), which hybridized with 9804 cDNA, was digested with restriction enzymes (New England Biolabs, Mississauga, Ontario, Canada) singly or in combination, and the products were separated on 0.8% agarose gels. The DNA bands were transferred to Zeta-Probe nylon membranes (Bio-Rad Laboratories, Hercules, CA) by capillary transfer in 0.4 M NaOH, and the membranes were prepared for hybridization following the manufacturer’s recommended procedure. 3P-labeled probes were prepared from DNA fragments using a random primer labeling kit (Boehringer Mannheim Canada, Laval, Quebec, Canada). Hybridizations were performed at 65°C in 1 mM EDTA, 0.5 M NaHPO4 and 7% SDS, pH 7.2, for 16 h, then washed at 65°C twice for 30 min in 1 mM EDTA, 40 mM NaHPO4, pH 7.2, containing 5% SDS, and twice more in the same buffer containing 1% SDS. Autoradiography was performed by exposing X-ray film (NEN-DuPont, Wilmington, DE) to the membranes at -70°C.

DNA fragment cloning and sequencing

The cosmid DNA was digested with PstI, BamHI, or SacI. After agarose gel purification of specific fragments, or without separation of digestion products, fragments were ligated into pBluescript II KS+ (Stratagene) at the PstI, BamHI, or SacI sites. XL-1 Blue bacterial cells (Stratagene) were transformed with the ligation product, and white colonies, which grew on agar plates containing ampicillin, X-galactoside, and isopropylthiogalactoside, were screened by colony filter hybridization using 32P-labeled probes generated from 9804 cDNA or from previously cloned genomic fragments. The fragment PCR5 was isolated by blunt end ligation of a PCR-generated fragment into the pBluescript II KS+ Smal site. This fragment spans exon II and exon IV of the 9804 gene, and could be amplified from cosmid or genomic DNA using primers 9804LEAD (5’ cgcggcttetctgatag 3’) and 9804TR5 (5’ gccggcactagc 3’), and the former was sequenced. The sequencing product was cloned into the M13 vector and partially sequenced. From this sequence, two 200 bp flanking primers were designed to amplify a genomic fragment in a Hybrid Thermal Reactor (Hybaid Ltd., Middlesex, U.K.): denaturation was at 94°C for 10 min before adding enzymes, followed by 30 cycles of 0.5 min at 94°C, 0.5 min at 50°C, and 1 min at 72°C, and a final incubation for 10 min at 72°C. Primers for amplification of an E48 gene fragment were E48F1 (5’ aggcctctctgctgtgctg 3’) and E48R1 (5’ gcggtcaggggccaaa 3’), and the temperature profile was: denaturing at 94°C for 10 min before enzyme addition, followed by 5 cycles of 2 min at 94°C, 3 min at 72°C, 35 cycles of 0.5 min at 94°C, 0.5 min at 65°C, and 1 min at 72°C, and a final 72°C incubation for 10 min. DNA sequencing was performed by the dideoxy chain termination method (T7 Sequencing Kit, Pharmacia Biotech, Baie D’urfe, Quebec, Canada) with flanking primers and internal oligonucleotide primers, which were synthesized at the Sheldon Biotechnology Centre (McGill University).

PCR screening of human-hamster somatic cell hybrids and libraries

Primers 9804ABF1 (5’ cttgactctggctgtagc 3’) and 9804X3R (5’ cacaatttgccgctgctg 3’) were based on sequences in intron 1 and exon III of the 9804 gene and were used in PCR under the following temperature profile: 99°C for 10 min before adding enzyme; 94°C; 1 min; 64°C, 30 s; and 72°C, 1 min for 5 cycles; 94°C, 30 s; 60°C, 30 s.; and 72°C, 1 min for 30 cycles; 72°C, 10 min. The primer pair 9804X3F1/9804X4R1 (9804X3F1 was 5’ cagactctctggctgtagc 3’) was used in PCR under the following temperature profile: 99°C for 10 min before adding enzyme; 94°C, 1 min; 60°C, 30 s; 72°C, 30 s, for 35 cycles; 72°C, 10 min. These two primer pairs were employed in screening BIOSMAP PCRable DNA from human-hamster hybrid cell panel 1 (BIOS Corp., New Haven, CT). Amplified fragment specificity was confirmed by hybridization with internal oligonucleotide probes. The pair 9804ABF1/9804X3R was subsequently used in screening human-hamster hybrid cell lines CL-17, 3.8/1-1, MC2F, 21q-1, and T/UC (38, 39), the distal portion of a large YAC contig spanning S2q3.2-S2q4.1, CEPH mega-YAC A and B libraries, a BAC library (Research Genetics, Huntsville, AL), and a human chromosome 8 cosmID library (LA08NC01). A 4.3 kb EcoRI fragment of a positive cosmID clone (c101F1) was subcloned into the M13 vector and partially sequenced. From this sequence, two 200 bp flanking primers AR431 (5’ gattgtagctggctgctg 3’ forward and 5’ gat gttgtagctggctgctg 3’ reverse) were developed for further sequencing of other cosmIDs. The following PCR cycle profile was used: initial denaturation for 5 min at 94°C, followed by 30 cycles of 0.5 min at 94°C, 1 min at 60°C, and 1 min at 72°C, and a final incubation 4 min at 72°C.

Cosmid library filter screening

The cosmid library LA08NC01 on Hybond-N filters was hybridized with positive cosmID c101F1 using a random prime labeling method. Repeated sequences were blocked by incubating each probe with 200 μl of human placental DNA (5 mg/ml), 100 μl NaPO4 (1 M, pH 7.2), 40 μl NaCl (5 M), and 1 μl EDTA (0.5 M) in 400 μl total volume for 20 to 30 min at 65°C. The library filters were prehybridized at 65°C for 30 min, then hybridized with probe at 65°C overnight. Prehybridization and hybridization buffer was composed of 0.25 M NaPO4, pH 7.2, 0.25 M NaCl, 5% SDS, 10% polyethylene glycol, and 1 mM EDTA. The filters were autoradiographed after washing three times at 65°C in 0.05 M NaPO4, 0.5% SDS, 1 mM EDTA.

Fluorescence in situ hybridization

Cosmid c101F1 was labeled with fluorescein-linked dUTP using the Prime-In-Fluorescence labeling kit (Stratagene) according to the manufacturer’s instruction. Repeated sequences were blocked by incubating with human placental DNA. Metaphase chromosome spreads were prepared from colcemide-treated CL-17 cells, denatured, dehydrated, hybridized, and washed according to the protocol from a Chromosome In Situ Hybridization Kit (Stratagene). A chromosome-specific centromere probe was used to identify human chromosome 8 (39). Slides were mounted with antifade solution containing propidium iodide and observed under an Olympus BH-RFCA microscope (oil immersion, 1500X). Microphotographs were taken using Kodak Ektachrome 400 ASA film.

Data analysis

Compilation and analysis of DNA sequence information were performed using MacVector Sequence Analysis Software version 4.1 (International Biotechnologies-Eastman Kodak, New Haven, CT) with ENTREZ. Sequences release 5.0, 6.0, or 7.0 (National Center for Biotechnology Information, Bethesda, MD). The genomic sequence was examined for gene structure by alignment with 9804 cDNA sequences and inspection for splice donor and acceptor sites (40). Related sequences were searched using the E-mail server blast@ncbi.nlm.nih.gov (41). Clustal W sequence alignment (42) was performed through the following Web site: http://dot.igen.bcm.tmc.edu/9331/multalign/options/clustalw.html. Promoters were predicted by use of the programs Proscan Version 1.7 at Web page: http://bimas.dctnr.nih.gov/molbio/proscan (43), and Eukaryotic Promoter Prediction analysis by Neural Network (EPPNN) at Web page: http://www.hgc.lbl.gov/projects/promoter.html. Potential transcription regulatory elements were inspected by Web Signal Scan Service: http://bimas.dctnr.nih.gov/molbio/signal (44). Polypeptide sequence distance analyses and phenogram construction were performed using the PHYLIP software package from Joe Felsenstein (downloaded with the documentation from the PHYLIP home page at Web site: http://evolution.genetics.washington.edu/phylip.html). PROTDIST analysis was performed on aligned amino acid sequences of Ly-6 domains using the Dayhoff PAM. The resulting matrix then served as input to KITSCH for tree computation. Optimizations were performed using Global branch rearrangement, whereby input sequences were jumbled 24 times. Analysis of several other programs and options within the PHYLIP package, along with use of alternate positional alignment choices, resulted in only slight differences affecting the relative placement of the most divergent Ly-6 superfamilies.

Results

Cloning of the 9804 cDNA

The 9804 cDNA clone6 was isolated as a result of differential screening of a parental U-937 cDNA library for genes that were expressed at much lower levels in a functionally distinct subclone (clone 6) (Fig. 1). In the system under study, first described by Johnson and Jenkins (45), the human monocytic cell line, U-937, provides a B7-independent form of T cell costimulation. As measured by IL-2 production by human T cells receiving a primary stimulation through anti-CD3 and immobilized secondary Ab, clone 6, in contrast to the parental line, was unable to provide the

[6] The 9804 gene is covered by U.S.A. patent number 548612. During preparation of reports on the cDNA and genomic cloning of 9804, we discovered partial sequence of other clones corresponding to the same gene appearing in the EST database. Capone et al. also discovered and published a compilation of these sequences (55). Independently, Mao et al. discovered this gene and called it RIG-1 on the basis of its retinoic acid inducibility in a promyelocytic leukemia cell line (54). Because of these reports, presentation of our cDNA data now would be redundant.
necessary costimulation. A reasonable explanation, which provided the motivation for the differential screening project, was that clone 6 was deficient in expression of one or more components of a novel costimulatory pathway. Among the cDNA clones identified in this way (see Materials and Methods), 9804 was considered the most promising candidate for an element of a costimulation pathway, because it encoded a protein related to the mouse Ly-6 family of cell surface Ags, and Ly-6 molecules are thought to be involved in intercellular signaling events. At present, however, we have no direct evidence that the 9804 protein plays a role in an alternative costimulation pathway. In sequence alignments, 9804 appeared at first most similar to Ly-6A, but following reports of the TSA-1 (SCA-2) cDNA sequence (5, 33), it became clear that 9804 was most likely the human homologue of mouse TSA-1 (Figs. 2 and 3).

Expression of 9804 mRNA in human tissues

Figure 4 shows the results of Northern blot analysis of RNA obtained from a variety of human tissues. There is little 9804 mRNA in adult or fetal skeletal muscle and low levels in the pancreas and testis. The highest levels in adults are found in the liver, kidney, ovaries, and PBLs. In another experiment, PBMCs expressed modest uninduced levels (see Fig. 5). In contrast to adult liver, fetal liver contains a very low level of 9804 mRNA. Other tissues containing moderate levels include: adult brain, placenta, lung, and colon. Lower levels are present in the heart, spleen, thymus, prostate, and small intestine. Fetal spleen has a relatively high amount of 9804 mRNA as compared with adult. This same high level is evident also in fetal adrenal glands. Moderate levels are found in fetal skin, and, as in the adult, it is also present in moderate amounts in fetal gastrointestinal tissue.

Effects of IFN on 9804 mRNA levels in U-937 cells and PBMCs

IFN-α and IFN-γ are known modulators of Ly-6 expression (12, 25, 26). As an additional test for potential relatedness of 9804 to the Ly-6 gene family, mRNA from IFN-treated U-937 cells and PBMCs was evaluated via Northern analysis (Fig. 5). Whereas IFN-α strongly enhanced 9804 mRNA levels in both U-937 and PBMCs, IFN-γ had little effect in U-937, and produced only a modest increase of 9804 mRNA in the PBMC preparation.

Isolation and sequencing of overlapping 9804 gene fragments

Restriction analysis of a cosmid, CMC9, containing the 9804 gene showed two PstI fragments of 1255 and 1390 bp that hybridized with a 9804 cDNA probe (not shown). Subcloning and sequencing these fragments (P19 and P40) demonstrated that together they accounted for the entire sequence of the original 9804 cDNA clone in three exons. These exons are labeled II, III, and IV in Figure 6. The positions of the intervening sequences were in accord with the highly conserved features found in the organization of other Ly-6-related genes. Forward and reverse oligonucleotide primers corresponding to sites flanking the 9804 coding region were used in the generation of PCR fragment PCR5. The sequence of this fragment overlapped P19 and P40, contained a single PstI site, and proved that the P19 and P40 sequences were indeed juxtaposed as depicted in Figure 6. We had accounted for the original 9804 cDNA sequence in just these three exons, but other Ly-6-related genes usually have a short upstream exon (exon I) that contains only 5′-untranslated sequence. Database searches revealed several cDNA sequences from the Washington University/Merck Expressed Sequence Tag (EST) Project that corresponded to the 9804 gene. A few of these had more extended 5′ ends. Comparison of these sequences with our 9804 cDNA clone and genomic sequences confirmed the presence of another exon containing 5′-untranslated sequence. To locate this exon I, two more upstream PstI fragments (P26 and P36) and overlapping BglII fragments (B7 and B18) were isolated. The relationships of these fragments to each other and to other fragments that were cloned to facilitate sequencing of this region are shown in Figure 6. The full sequence of the region spanned by these four PstI fragments, a distance of 5543 bp, was determined and is presented in Figure 7. The sequence includes the four exons of the 9804 gene with 483 bp downstream of its 3′ end, and 1167 bp upstream of the 5′ end of the most 5′ extended cDNA sequence known at this time (Fig. 7).

Structure of the 9804 gene

The gene is composed of four exons that are separated by three introns of variable lengths. A comparison of the genomic sequence with the cDNA permitted location of the intron/exon boundaries, and these were found at positions that corresponded well with those of other Ly-6 genes. The splice sites are compared with the eukaryote consensus splice site sequences in Table I. Exon I, containing only untranslated sequence, is about 80 bp in length, although the actual transcription initiation site has yet to be determined. Exon II is 107 bp long and contains 55 bp of 5′ untranslated sequence and the first 52 bp of coding region. This, together with the first 8 bp of exon III, encodes the 20-amino acid leader sequence. The remainder of the polypeptide sequence is encoded in the following 112 bp of exon III and the first 125 bp of exon IV (including the stop codon). Exon IV contains a further 716 bp of 3′-untranslated sequence. Intron 1, which contains an inverted Alu repeat sequence (bases 2870 to 3188) about 360 bases upstream of exon II (Fig. 7), is comparatively large (2298 bp), whereas introns
2 and 3 are only 321 and 131 bp, respectively. In Figure 8, we provide a dot plot comparison of the gene sequence with the recently published mouse TSA-1 genomic sequence (46). The spacing of exons is very similar, but the greatest degree of sequence similarity is evident in the region encompassing exons II to IV, which encode the TSA-1 and 9804 polypeptides. For use in gene mapping studies, we have designed primers based on the published E48 cDNA sequence that direct amplification of a fragment between exons III and IV of the E48 gene. Analysis of the PCR fragment amplified from human genomic DNA with this primer pair revealed an E48 intron 3 remarkably similar in size to that of the 9804 gene. Alignments (Fig. 9) and distance comparisons (not shown) of the short 9804, E48, TSA-1, and ThB intron 3 sequences was consistent with the close relationship between 9804 and TSA-1, and showed no significant similarity between the 9804 and E48 introns, since in pairwise alignments, the identity scores were only marginally reduced by randomizing the input sequences (not shown).

Primer extension analysis of the 5′ end of 9804 mRNA has not yet worked for us, and we think this is due to the high GC content of exon I. In an attempt to identify potential transcription initiation sites, we performed promoter analyses on the first 1245 bases of the sequence (Fig. 7); we employed the programs ProScan and EPPNN (see Materials and Methods). The former failed to predict any promoter for the gene. The latter also failed at the default cutoff value, but identified five potential transcription initiation sites when a reduced cutoff value was chosen. These results are presented in Table II, and the sites are indicated by bold capital letters in Figure 7. Four of these sites lie between 278 and 230 relative to the most 5′ cDNA sequence. While there is no TATA or CAAT consensus sequence, the promoter region is GC-rich (78% within 100 bp of the predicted exon I), and contains several potential Sp1 sites, an example of which (GC box) is indicated in Figure 7. Many potential transcription factor binding sites can be found in this region, but we have presented only potential IFN-α-stimulated response elements (ISRE) and retinoic acid response elements (RARE) in Figure 7. In addition, a sequence comparison of these potential response elements with ISRE and RARE known from several different genes is presented in Table III (47–49).

**FIGURE 2.** Alignment of the predicted 9804 polypeptide Ly-6 module sequence with several members of the Ly-6 family and superfamily. Only the Ly-6 modules of each are aligned (motif starts at XXCCXC, usually immediately after the signal peptide, and ends at CXXCCXC). In the case of uPAR, which contains three Ly-6 modules, they are labeled 1 to 3 starting from the outermost, N-terminal domain. Alignments were generated with the Clustal W program and slightly adjusted to align Cys 7 in CD59 and uPAR 3, which were out of alignment by two places.

**FIGURE 3.** Phenogram showing diagrammatically the degrees of similarity between the Ly-6 modules aligned in Figure 5. Programs from the PHYLIP software package were used for this analysis, as described in Materials and Methods.

Chromosomal localization of 9804 gene

Primer pairs were designed for specific PCR amplification of 9804 gene fragments from human genomic DNA (see Materials and Methods). In preliminary screening of DNA from the BIOS-MAP..
human-hamster hybrid cell panel 1, amplification of specific fragments correlated with the presence of human chromosome 8 in the hybrid (data not shown). A more refined study was performed on human-hamster cell hybrids CL-17, 3;8/4-1, MC2F, 21q11, and TL/UC, which contain chromosome 8 deletions or translocations (38, 39). The portion of chromosome 8 present in each of these chromosomes is shown in Figure 10 along with the PCR results. Amplification of the 9804 gene fragment was seen for template DNA from CL-17, 21q11, and TL-UC in addition to the control total human DNA. No amplification was seen when the template DNA was derived from 3;8/4-1, MC2F, or total hamster DNA or when no DNA template was included. These results sublocalize the primers to the telomeric end of the q arm in interval I-9 of chromosome 8 (50).

The primer pair 9804ABF1/9804X3R was then applied in screening YACs in the distal portion of a large YAC contig (51). This contig extends into interval I-9 and about 2 megabases distal to the c-Myc gene located in 8q24.2. All YACs within this contig were negative, indicating that the primers were distal to c-Myc.

The primers were then used to screen CEPH mega-YAC A and B libraries as well as a BAC library (Research Genetics) and a chromosome 8 cosmid library, LA08NC01 (52). Although no positives were seen for the YAC or BAC libraries, two positive cosmids, c101F1 and c157F6, were isolated from the cosmid library. These cosmids were subsequently used to screen the same library and develop an overlapping cosmid contig (Fig. 11). Overlaps between cosmids were confirmed by comparison of EcoRI restriction enzyme digestion patterns and by direct hybridization of cosmids to each other. PCR using the sequence tagged site primer pair AR431, developed from sequences near one end of cosmid c101F1, amplified fragments from c101F1, c93F9, c16E6, and c182H9, but not c157F6 and c189A10 (Fig. 11).

Cosmid c101F1 was fluorescently labeled and hybridized to metaphase spreads of hybrid cell line CL-17, which contains intact human chromosome 8 as its only human material. Figure 12 shows a typical metaphase spread of chromosome 8, distinguished by

FIGURE 4. Northern blots of 9804 mRNA in RNA extracts from various human adult and fetal tissues. A, A human multiple-tissue Northern blot (Clontech). B, Total RNA from adult and fetal (18 wk) tissues. An attempt was made to load equal amounts (10 μg) of each. To show loading differences and check on the quality of the RNA preparations, the intensity of ethidium bromide staining was monitored and is presented below the blot.

FIGURE 5. Northern blots showing the effects of IFN on 9804 mRNA levels in U-937 cells (A) and PBMCs (B). Cells were incubated for 20 h with no IFN, with IFN-γ (100 U/ml), or with IFN-α (1000 U/ml). 10 μg of total RNA was applied per well.
hybridization with a chromosome 8-specific centromere probe, and showing positive fluorescence due to c101F1 hybridization in the 8q24.3 region close to the telomere.

Discussion
Not only because of the size of the mouse Ly-6 family, but also because the independent regulation of each member generates distinct patterns of expression during hemopoiesis and immune responses, suggesting that they provide distinct functions affecting various stages of leukocyte development, it was reasonable to expect that we would find homologues of at least some of these molecules in human. Since the first characterization of an Ly-6 gene, attempts have been made to detect their human counterparts through interspecies homology cloning procedures. These attempts have been hampered by an apparently rapid divergence of the genes between species that has prevented detectable cross-hybridization of Ly-6A-like genes much beyond a subset of the rodentia (X. Shan and R. G. E. Palfree, unpublished results). So far, related genes have been cloned in this way only from rat (53). Little homology cloning has been attempted with ThB or TSA-1 because reports of their sequences were comparatively recent. In fact, the TSA-1 cDNA sequence became available only after we had sequenced 9804. The discovery and characterization of the 9804 gene as a likely human homologue of mouse ThB, and 9804 forms a subgroup with TSA-1/Sca-2. Furthermore, the members in each subgroup appear to be at least as closely related to each other as Ly-6C is to Ly-6A.

With respect to chromosomal location, the mapping of the 9804 gene to human chromosome 8 in the q24.3 region, telomeric to c-myc, is significant, since this region is homologous to the mouse Ly-6 gene locus on mouse chromosome 15 (29). Furthermore, the E48 gene has also been mapped to 8q24 (13). Taken together, these data strongly support the recognition of 9804 and E48 as homologues of mouse TSA-1 and ThB, and suggest the existence of a multigene family for human Ly-6 in the region q24 of human chromosome 8. It is likely, therefore, that additional human Ly-6 genes more related to Ly-6A will be found in the vicinity of 9804 and E48.

Although the precise function is unknown for any member of the Ly-6 family, an indication that 9804 performs a function like that of TSA-1 may be obtained by comparing their expression levels in various tissues. TSA-1/Sca-2, as assessed by immunofluorescence staining with mAbs, is expressed on 75% of bone marrow leukocytes, mature B cells, a subset of thymic medullary epithelial cells, thymic dendritic cells, and immature thymocytes (7). Northern analysis, however, reveals that the mRNA is present in a wide range of tissues (5). Our Northern blot analysis of 9804 expression included more tissues than reported in the TSA-1 study, so a detailed comparison is not yet possible, but it is evident that both mRNAs are broadly distributed in a variety of tissues. There is also a parallel between the levels of TSA-1 and 9804 mRNA expression among certain tissues. For example, low levels are expressed in both testis and heart, while high levels are found in liver and lung. Furthermore, both can be detected in kidney, spleen, and brain. These data are consistent with the prediction that 9804 is a functional homologue of mouse TSA-1/Sca-2.

Studies on the function of TSA-1, largely restricted to the thymus, reveal that anti-TSA-1 can inhibit the transition from CD4+CD8- to CD4- thymocytes. This inhibition thus favors the development of CD8+ cells in organ culture (14). The interpretation of these data is difficult because Abs can mimic either natural ligands (stimulate) or block the functional engagement of these molecules (inhibit). Moreover, several distinct cell types may be

FIGURE 6. Genomic organization, restriction map, cloned fragments, and sequencing strategy of 9804. The exons are boxed; translated portions of these are solid boxes. Arrowed lines indicate the subcloned fragments and arrowheads the internal primers used for sequencing. The restriction enzyme sites are designated as follows: PstI (P), NcoI (N), MscI (M), SacI (S), BglII (Bg), BamHI (Bm), SmaI (Sm).
affected in the thymus. Even so, the above data indicate that there are probably cells in the thymus that are responsive to an interaction between TSA-1 and a natural TSA-1 ligand. This suggests that the 9804 gene product also participates in intercellular communication relevant to developmental processes. A recent report, describing the cloning of a retinoic acid-inducible gene named RIG-E (54), which is identical with 9804, provides additional evidence to support this conclusion.

FIGURE 7. Nucleotide sequence of the 9804 gene. Exons are denoted by upper case letters and underlined. Donor and acceptor sites within the introns are denoted by italic boldface type. Predicted transcription initiation sites are indicated in single boldface capital letters. A GC box, one of several potential Sp1 sites, and the polyadenylation signal are boxed. The translation start and termination codons are indicated in boldface and are double underlined. Potential ISRE and RARE in the promoter region are indicated in boldface and labeled above the sites. An Alu repetitive sequence in intron 1 is underlined with a dashed line.
The chromosomal genes for Ly-6A/E, Ly-6C, Ly-G, Ly-6F, ThB, and most recently TSA-1 have been cloned and characterized. The structures of several other Ly-6-related genes are also known. Apart from ThB, which appears to lack the usual exon I and contains part of the 5′ untranslated sequence, these genes typically possess four exons. The boundaries of the other three exons correspond to particular sites within the encoded polypeptide. Exon II terminates approximately three codons before the end of the signal peptide, while the boundary between exons III and IV corresponds to a position in the loop between the fifth and sixth of the 10 cysteine residues in the mature Ly-6 domain (6). We have found this same organization in the structure of the 9804 gene. Furthermore, the similar spacial arrangement of 9804 and TSA-1 exons is consistent with the close homology indicated by their sequence similarity. There is considerable variation in the sizes of corresponding introns in Ly-6-related genes. Intron 3 varies from approximately 7 kb in the CD59 gene down to 162 bp in the ThB gene, and now 131 bp in the 9804 gene. We have been able to compare these with the corresponding intron from the ThB-like E48 gene (Fig. 9), because preliminary work in a study of the region of chromosome 8 containing the two new human Ly-6 genes produced a PCR fragment of the E48 gene that included this intron. Interestingly, the sizes of the E48 intron (132 bp) and 9804 intron 3 are almost identical and are much smaller than intron 3 in mouse Ly-6A (~1 kb) or Ly-6C (~1.7 kb) genes. There is, however, little sequence similarity between these small 9804 and E48 introns, so no evolutionary relationship can be inferred.

The full sequence of our 9804 cDNA clone is contained in three exons, but the most common structure for Ly-6-related genes includes an additional exon, exon I, which contains the beginning of the 5′ untranslated sequence. Shortly after data began to accumulate from the Washington University-Merck EST project, routine searches of GenBank showed an accumulation of sequences in the database that overlapped with the 9804 sequence. Several were found to begin upstream of the original 9804 cDNA sequence, and these defined the 3′ end of exon I. As yet, however, we have no direct knowledge of the transcription initiation site. In a promoter analysis that required a lower cutoff value than the default, the EPPNN program located four potential initiation sites between 30 and 78 bases upstream of the most 5′-extended 9804 sequence that we have found so far in the EST database. As is the case for mouse Ly-6 genes, there is no TATA box in the presumed promoter of the 9804 gene, but there are several potential Sp1 binding sites within 220 nucleotides upstream of the closest predicted initiation site, and this region is GC-rich (~80%).

In comparison with E48, which is expressed exclusively in the skin (13), 9804 has much broader tissue distribution in terms of mRNA expression. We have shown that 9804 mRNA levels are enhanced strongly by IFN-α and slightly by IFN-γ in normal human PBMCs. As seen in the case of Ly-6A, this increased mRNA level is probably accompanied by elevated surface 9804 protein expression on IFN-α-treated PBMC, and it may indicate a requirement for the 9804 function, particularly during antiviral immune responses. IFN-α, but not IFN-γ, was also effective in inducing higher 9804 mRNA levels in U-937 cells, and we have found the same to be true for most malignant cell lines tested so far, although not in the B cell lines Daudi and Raji (35). Recently, Capone et al. (55) also discovered the TSA-1-like cDNA sequences in the EST database, and performed Northern analyses that complement ours by showing expression in lymphoid cells, including T and B cells.
macrophages, and dendritic cells. This distinctive expression pattern will be under the control of response elements for tissue-specific and inducible transcription factors. A large number of potential transcription factor binding sites were revealed by computer analysis of the 5' flanking region of the 9804 gene, most of which will have no physiologic relevance. But we also looked especially for retinoic acid response elements, \( g \)-IFN-activated sites (GAS), and sequences similar to known ISRE, because of the inducibility of 9804 gene expression by IFNs (shown above) and retinoic acid (54). Furthermore, an ISRE and GAS are known to be involved in the induction of mouse Ly-6A expression by IFN-\( \gamma \) and IFN-\( \alpha \) (56–58). No typical GAS site was found in the 9804 promoter region, which is consistent with there being only a slight effect of IFN-\( \gamma \) on 9804 mRNA levels. Such an effect could be indirect. A potential ISRE and two possible RARE half-sites were found in the promoter region of 9804 (Fig. 7). The regulatory elements for the induction of 9804 expression by IFNs and retinoic acid are currently under investigation, and we are employing DNA footprinting to determine which other elements in this promoter region may contribute to the control of 9804 expression.

Our discovery of 9804 resulted from efforts to identify unique U-937 cell components that would be responsible for the intercellular interaction leading to T lymphocyte costimulation. This relationship of the cDNA to a family of molecules known to be...
involved in cellular communication and cell adhesion encourages the speculation that the 9804 protein on U-937 may indeed affect intercellular interactions with responding T cells. Our initial attempts to substantiate this hypothesis have not produced convincing results. Abs to 9804 will greatly facilitate future studies.

A final note concerning nomenclature: as in our first presentation of this work at a recent meeting (35, 59), we have retained the original name “9804,” resisting the temptation to rename it HuSca-1 or HuTSA-1, for example, because of the misleading connotations associated with such nomenclature. In this, we are in agreement with Brakenhoff et al. (13) for whom HuThB (human thymocyte and B cell Ag) would have been a totally inappropriate name for the skin-specific E48 Ag. Even the name RIG-E places too much emphasis on one of many defining properties of the gene: its retinoic acid inducibility. Indeed, our data are consistent with a very strong induction of the 9804 gene by IFN-α. Until a more acceptable name is agreed upon, therefore, we shall continue to use the term 9804.

Acknowledgments

The authors thank Nicholas Brown for help with the in situ hybridization, and Drs. H.P.J. Bennett, A. Bateman, and F. Congote for sharing reagents and helpful discussions.

References


