ORIGINAL PAPER

William S. Modi · Julie Bergeron · Michael Sanford

The human MIP-1β chemokine is encoded by two paralogous genes, *ACT-2* and *LAG-1*

Abstract Human macrophage inflammatory protein-1 beta (MIP-1β) is an M_r 8,000 acidic protein that is upregulated upon stimulation in monocytes, T cells, and other lymphocytes. This protein belongs to the CC chemokine subfamily and directs the migration of specific subsets of leukocytes. The first molecular clone was isolated in 1988, and ever since there has been confusion regarding the exact number of genes encoding this and closely related proteins. PCR primers were designed from two genomic GenBank entries to conduct single-strand conformational polymorphism analysis, sequence analysis, and PCR-RFLP, and we conclude that previously isolated clones referred to as *MIP-1β* are derived from two genes, originally called *ACT-2* and *LAG-1*. The two proteins share a common length and are identical at 89 of 92 amino acids. The first two amino acid differences, V12M and L20P, occur in the signal peptide, while the third, G70S, is in the mature protein. Within the transcribed region, the genes differ at 25 of 662 nucleotides. A survey of the NCBI expressed sequence tag database reveals that both genes are expressed in a variety of tissues, and five clones representing *LAG-1* transcripts are alternatively spliced, with the 115-bp exon 2 omitted. Database searches for putative orthologues in other species revealed that the rabbit protein is about 80% similar to the two human proteins, while those of rat and mouse are 70–75% similar. Comparative sequence analysis of the human and animal proteins indicates substantially higher rates of protein evolution in the two rodents compared to human and rabbit.

W.S. Modi (✉) · J. Bergeron
Intramural Research Support Program, SAIC Frederick,
National Cancer Institute-Frederick, Frederick,
MD 21702-1201, USA
e-mail: modi@ncifcrf.gov
Tel.: +1-301-8461303, Fax: +1-301-8461909

M. Sanford
Laboratory of Genomic Diversity,
National Cancer Institute-Frederick Cancer Research and
Development Center, Frederick, MD 21702-1201, USA

Introduction

Macrophage inflammatory protein-1 alpha (MIP-1α) and macrophage inflammatory protein-1 beta (MIP-1β) were originally co-purified from endotoxin-stimulated mouse macrophages and defined as a single protein (Wolpe et al. 1988). Subsequent biochemical analyses revealed the distinctiveness of each protein and molecular cloning confirmed that they are encoded by different genetic loci in the mouse (Davatelis et al. 1988; Sherry et al. 1988).

The first human clones identified as the apparent orthologue to mouse *MIP-1b* were cDNAs from a gene called *ACT-2*, which represented a novel immune activation gene (Lipes et al. 1988) *ACT-2* was isolated by differential hybridization screening of an activated T-cell library and the protein was shown to be 76% identical at the amino acid level to murine MIP-1β (Napolitano et al. 1991). Shortly thereafter, additional human cDNA, including *pAT 744* (Zipfel et al. 1989), *hH400* (Brown et al. 1989), *G-26* (Miller et al. 1989), *HC21* (Chang and Reinherz 1989), and genomic *LAG-1* (Baixeras et al. 1990) clones corresponding to MIP-1β were isolated by other laboratories. As the sequences from these clones were aligned, numerous differences were discovered and the presence of different alleles or even distinct, closely related genes was postulated (Baixeras et al. 1990). The existence of more than one gene was also supported by Southern blot analysis of genomic DNAs (Chang and Reinherz 1989; Napolitano et al. 1991). The exact number of genes was uncertain, but Schall (1991) reported that many of the seven independently isolated *MIP-1b* cDNA clones differed slightly in the predicted proteins they encoded.

We are isolating and characterizing novel single nucleotide polymorphisms (SNPs) from various human chemokine genes. The goal is to use the SNPs as genetic

markers in association analyses with various disease phenotypes including AIDS, asthma, and kidney disorders. One approach commonly used to identify SNPs is the polymerase chain reaction (PCR)-based single-strand conformation polymorphism technique (SSCP) (Orita et al. 1989) followed by direct sequence analysis. One outcome of our search for SNPs in the human *MIP-1b* gene was the discovery that certain primer pairs were amplifying more than one gene. This prompted further investigation into the genomic structure of the *MIP-1b* gene. Results suggest that the human MIP-1β protein is encoded by two closely related paralogous genes derived from a recent duplication event.

Materials and methods

Two genomic sequences from GenBank were used for PCR primer design. X53682 contains a 2488-bp sequence originally referred to as the lymphocyte activation gene (*LAG-1*) (Baixeras et al. 1990). AC003976 is an anonymous 123,925-kb BAC clone hCIT.91_J_4 submitted by the Whitehead Institute for Biomedical Research/MIT Center for Genome Research (WICGR) (http://www.genome.wi.mit.edu/), and contains the *ACT-2* gene in addition to other genomic elements. X53682 and AC003976 were aligned using the GAP procedure in the Wisconsin Package (1997).

PCR was performed as follows: 30 ng of genomic DNA was amplified in a 20-µl volume in 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl2, 200 µM of each dNTP, 6 pmol of each primer, and 0.001 units of *Taq* Gold DNA polymerase (Perkin-Elmer). PCR cycling conditions consisted of an initial 94°C denaturation step for 10 min, followed by 35 cycles of 94°C, denaturation for 30 s, 57–62°C annealing for 1 min, and 72°C extension step for 1 min, ending with a 72°C step for 5 min.

PCR for SSCP was conducted as described above with the addition of 0.04 µl of a ^{32}P-dCTP (3000 Ci/mmol; New England Nuclear) to each reaction. After PCR, products were digested with a restriction enzyme in the appropriate buffer (products from primers 12/13 with *Bsa*JI and products from 3F/3R with *Aci*I; Table 1). Ten microliters of stop solution (0.05% each of bromphenol blue

and xylene cyanol in 100% foramide) was then added to each reaction. The products were heated at 95°C for 5 min, then quick chilled on ice. Products were separated in vertical slab gels using at least two of the following four electrophoretic conditions: (1) 6% acrylamide (37.5:1 acrylamide:bisacrylamide), 4°C at 50 W for 3.5 h, (2) 6% acrylamide (37.5:1 acrylamide:bisacrylamide) with10% glycerol, 4°C at 10 W for 8–12 h, (3) 5% acrylamide (19:1 acrylamide:bisacrylamide), room temperature at 40 W for 4–6 h; the first three conditions used 1×TBE buffer (0.09 M Tris base, 0.09 M boric acid, 1 mM Na$_2$EDTA, pH 7.8), or (4) 6% acrylamide (37.5:1 acrylamide:bisacrylamide), 4°C at 50 W for 3.5 h in 1×TME buffer, pH 6.8 (0.03 M Tris base, 0.035 M 2(N-morpholino)-ethanesulfonic acid, 1 mM Na$_2$EDTA pH 7.8).

PCR products were sequenced directly using Big Dye chemistry (Applied Biosystems). Products were first purified using Qiagen quick-spin columns. Electrophoresis was carried out on an ABI 377 automated fluorescent DNA sequencing machine. Nucleotide sequence alignment, consensus sequence construction, and restriction mapping were done using Sequencher software (Gene Codes).

PCR-RFLP experiments were carried out for two restriction sites distinguishing *LAG-1* from *ACT-2* within the sequenced regions. These were conducted on genomic DNA from 40–48 unrelated Caucasians. DNA was amplified using the appropriate primers, and PCR products were digested for 4 h in 40-µl volumes using the proper restriction enzyme and buffer (New England Biolabs). Products were separated in 3% agarose slab gels, stained with ethidium bromide, and photographed. Restriction fragments were compared in size to a molecular-weight ladder.

Results

Initially, two pairs of PCR primers (LAG-1 12/13 and 3F/3R) were designed from GenBank X53682 for SSCP (Table 1). When they were used on genomic DNAs from 36 European Americans and 36 African Americans, the SSCP patterns revealed unusually high levels of individual variation. Subsequent sequence analyses of products from pair 3F/3R from several individuals identified numerous sites of apparent heterozygosity. These results suggested that more than one gene was being amplified.

Table 1 PCR primers used for amplification of the human *LAG-1* and *ACT-2* genes. In parentheses are the primer locations in the GenBank reference sequences: *LAG-1* X53682; *ACT-2* AC003976.

Underlined bases represent mismatches between the GenBank reference sequences (*UTR* untranslated region)

Primer	Sequence	Region amplified	Product size (bp)	Gene(s) amplified
LAG1-12 (791–815) LAG1-13 (1231–1252)	GCACCAAGTAAGTCTACTTTTGCAG CCCCAAATATGGTCCCTTGTAT	Intron 1	563	Both
LAG1-3F (1986–2005) LAG1-3R (2467–2486)	GCCAGATGACCTCAAAGGTC TACCAGGCAATTCCTTTTAC	Exon 3 and 3′ UTR	500	Both
LAG1-27 (224–240) LAG1-1R (826–845)	GGGAGTCCTGGTTCCCG# CTCTGCCTAGACCTTGACTC	5′ UTR and exon 1	621	*LAG-1*
LAG1-14 (1222–1242) LAG1-15 (1545–1561)	TGGGAATGGATACAAGGGACC AGCTCCCTCCCCAACGC	Exon 2	319	*LAG-1*
LAG1-18 (1951–1971) LAG1-19 (2467–2488)	GCAGAGGAAGATGCCTACCAC GGTACCAGGCAATTCCTTTTACC	Exon 3 and 3′ UTR	537	*LAG-1*
ACT2-11 (47024–47047) ACT2-12 (47265–47246)	AGTATCAGCACAGGACACAGCTG CACTGTTTGCCAGCCATGAC	5′ UTR and exon 1	240	*ACT-2*
ACT2-1B (47598–47624) ACT2-2B (47920–47900)	GCTCATGTTAGGTGGGAATGGATATTT CCCCAAAACAGGCCCTCTTTT	Exon 2	322	*ACT-2*
ACT2-3B (48342–48363) ACT2-14 (48892–48913)	GGTCAGGCAGAGGAAGATGCAA TTCCTTTTACCCCCAAATCAAAC	Exon 3	550	*ACT-2*

```
      Exon 1
    1 AGCACAGGACACAGCTAGGTTCTGAAGCTTCTGAGTTCTGCAGCCTCACCTCTGAGAAAACCTCTTTGCCACCAATACCATGAAGCTCTGCGTGACTGTC
      ----------------G----------------------------------------------------T----------------------------
                                                                                 m   k   l   c   v   t   v

                                                        Exon2
  101 CTGTCTCTCCTCGTGCTAGTAGCTGCCTTCTGCTCTCTAGCACTCTCAGCACCAATGGGCTCAGACCCTCCCACCGCCTGCTGCTTTTCTTACACCGCGA
      ------------A----------------------------C---G-----------------------------------------------------
      l   s   l   l   v   l   v   a   a   f   c   s   l   a   l   s   A   P   M   G   S   D   P   P   T   A   C   C   F   S   Y   T   A   R
                  m                               p

                                                                      Exon 3
  201 GGAAGCTTCCTCGCAACTTTGTGGTAGATTACTATGAGACCAGCAGCCTCTGCTCCCAGCCAGCTGTGGTATTCCAAACCAAAAGAGGCAAGCAAGTCTG
      --------------------------------------------------------------------------------------A------------
        K   L   P   R   N   F   V   V   D   Y   Y   E   T   S   S   L   C   S   Q   P   A   V   V   F   Q   T   K   R   G   K   Q   V   C
                                                                                                     S

      HhaI
       |
  301 CGCTGACCCCAGTGAGTCCTGGGTCCAGGAGTACGTGTATGACCTGGAACTGAACTGAGCTGCTCAGAGACAGGAAGTCTTCAGGGAAGGTCACCTGAGC
      T-----T--------A------------------------------------------------------------------------------------
      A   D   P   S   E   S   W   V   Q   E   Y   V   V   Y   D   L   E   L   N   stop
      MspI
       |
  401 CTGGATGCTTCTCCATGAGCCGCATCTCCTCCATACTCAGGACTCCTCTCCGCAGTTCCTGTCTCTTCTCTTAATGTAATCTCTTTTATGTGCTGTATTA
      -C----------------A-A--------------------------------------C-----------T------T---------C--G---

  501 TTGTATTAGGTGTTATTTCCATTATTTATATTAGTTTAGCCAAAGGATAAGTGTCCCCTATGGGGATGGTCCACTCTCACTCTTTCTCTGCTGTTGCAAA
      -------------C---------------------------------------------------------G-----G-----------------

  601 TACATGGATAACACCGTTAATTCCATGTGTTTTCATAATAAAACTTTAAAATAAAATGCAAA        LAG-1 full length
      --------------AT--G----TG---------------------------------------        ACT-2 full length
```
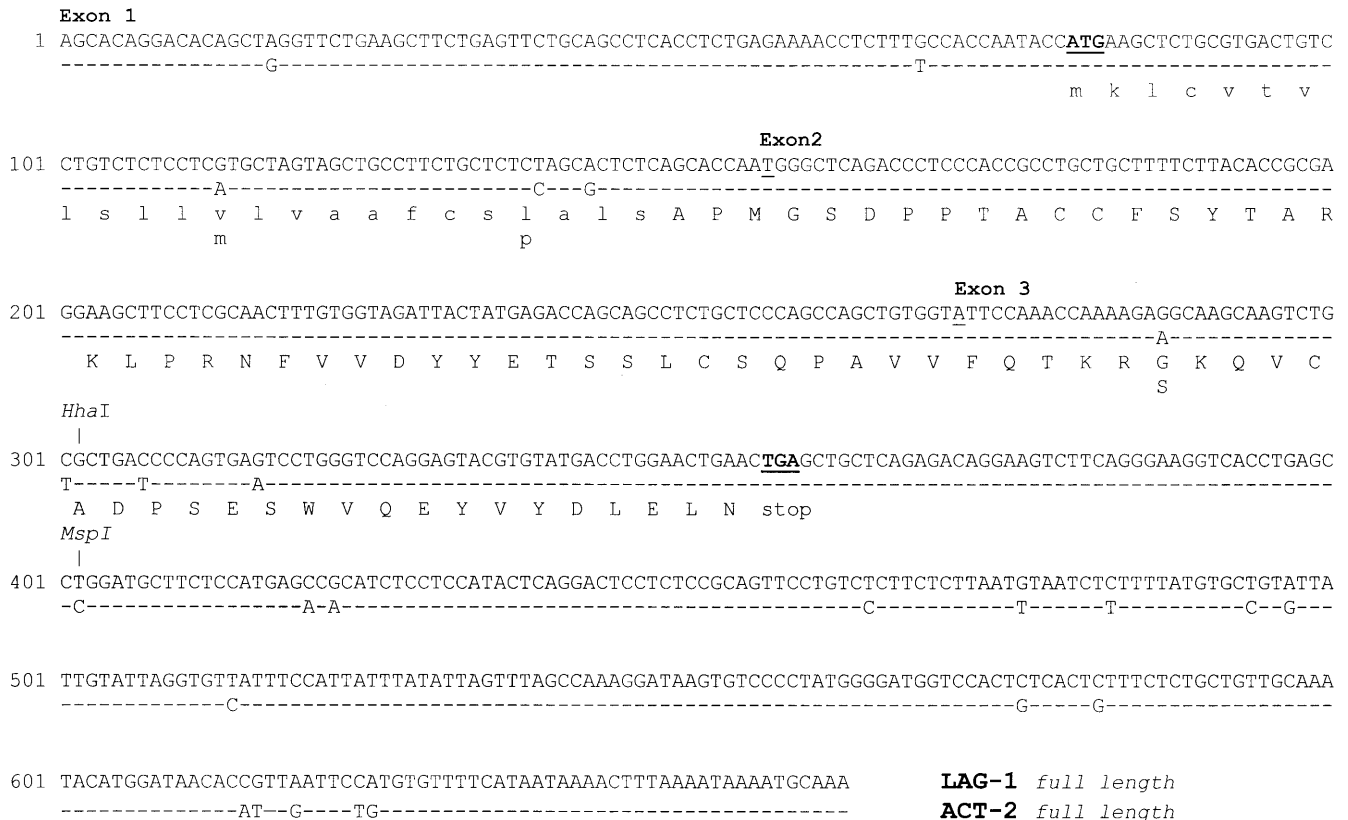
Fig. 1 Nucleotide sequence alignment of the 662-bp coding and flanking regions of the human *LAG-1* (*upper*) and *ACT-2* (*lower*) chemokine genes. Identical bases are indicated by dashes (–), differences by the appropriate nucleotides. Exon 1 begins at the RNA start site proposed by Napolitano and co-workers (1991). Initiation (ATG) and termination (TGA) codons are *underlined* and in *bold*. The first nucleotides of exons 2 (T) and 3 (A) are indicated. Positions of two restriction sites (*Hha*I, *Msp*I) distinguishing the genes are given. Amino acid translation for LAG-1 appears below the nucleotide sequences. The three residues (m, p, S) where ACT-2 differs are indicated below the LAG-1 protein. The 23 amino acid signal peptide is in *lowercase*

BLAST searches of the NCBI database with X53682 revealed similarity to AC003976. Alignment showed the two sequences to be 95.6% identical over the entire 2488 bp of X53682. The X53682 sequence aligned with the region of AC003976 between bp 46388 to bp 48875. To validate the sequence differences between the two clones in the protein-coding regions, primers were designed from each sequence with terminal mismatches between the two GenBank sequences. PCR amplification of genomic DNA and direct sequencing of products from six to ten individuals were carried out using the following primers: LAG-1 27/1R, LAG-1 14/15, LAG-1 18/19, ACT-2 11/12, ACT-2 1B/2B, and ACT-2 3B/14 (Table 1). These primers are located in noncoding regions and amplify each of the three coding regions and some flanking material. Results indicated that the sequences derived from clone X53682 (*LAG-1*) differed from the sequences obtained from clone AC003976 (*ACT-2*).

The sequences of the coding and associated 5′ and 3′ flanking regions of the two genes are shown in Fig. 1.

The transcribed regions were identical in length (662 bp), with all differences being attributable to base substitutions as opposed to insertion-deletions. A total of 155 bp, 115 bp, and 392 bp were sequenced for the putative transcribed regions of each exon, respectively. Translated amino acid sequences are included in Fig. 1; both full-length proteins contain 92 amino acids. The signal peptides are 23 amino acids in length. The number of nucleotide (nt) and amino acid (aa) differences between X53682 and AC003976 in each exon are – exon 1: 5 nt, 2 aa; exon 2: 0 nt, 0 aa; exon 3: 20 nt, 1 aa. Two of the three amino acid differences occur in the signal peptide, V12M and L20P. The third, G70S, is found in the mature protein.

Two restriction site differences in exon 3 (*Hha*I and *Msp*I) distinguishing the genes are shown in Fig. 1. These sites were used in PCR-RFLP assays of genomic DNA from 40–48 Caucasians to demonstrate that the variation found between the two genes identified by sequencing may be extended to a larger number of individuals. Analyses revealed that *Hha*I should cut *LAG-1* at position 301, and *Msp*I should cut *ACT-2* at position 402. Both of these sites were tested using primers LAG-1 18/19 and ACT-2 3B/14 and the expected results were found (RFLP results not shown).

Discussion

The LAG-1 and ACT-2 proteins differ by three amino acids, two in the signal peptide and one in the mature

protein. Among the differences in the signal peptide, the valine to methonine substitution is a conservative one. The leucine to proline change involves two nonpolar residues, with proline containing a large aromatic ring. The glycine to serine change at position 70 in the mature peptide involves two polar, uncharged residues with glycine containing -H and the larger serine, $-CH_2OH$ as side groups.

A nuclear magnetic resonance structure is available for the ACT-2 protein (Lodi et al. 1994). Evaluation of the three-dimensional structure seen in the protein database (http://www.rcsb.org/pdb/cgi/explore.cgi?pid= 4377977407482&pdbId=1HUM) shows that S70 is positioned in a β-pleated sheet following a tight β turn. The oxygen atom of the hydroxyl side chain is hydrogen bonded with the hydrogen atom of T67. Furthermore, the oxygen atom of S70 is hydrogen bonded with the amino group of K71. Thus, this network of hydrogen bonds stabilizes the turn and β sheet. If S70 is mutated to glycine, the network of hydrogen bonds is broken, and due to the increased flexibility of glycine, the β turn may also be destabilized. Since MIP-1β (ACT-2) is known to bind the seven-transmembrane G protein-coupled receptor CCR5 (Laurence et al. 2000), we might assume that LAG-1 binds this receptor also. The effect of G70S on receptor binding is unknown; however, mutations at a single amino acid (F13) have been shown to affect ACT-2 binding (Laurence et al. 2000).

Among human chemokine genes there are several additional examples of very similar paralogous gene/protein pairs or triplets. These include MIP-1α and LD78β (Irving et al. 1990), lymphotactin-1α and 1β (Yoshida et al. 1996), platelet factor 4 (PF4) and platelet factor 4 variant 1 (PF4V1), and GROα, GROβ, and GROγ (Haskill et al. 1990). In all of these examples, the mature proteins in each group are at least 95% identical. At first glance this would suggest some redundancy in function; however, the variant amino acids may be adequate for functional distinctiveness or there may be substantial differences in transcriptional regulatory regions of the genes (Hogan et al. 1994). Pairs of closely related chemokine genes have also been reported in other species and this reflects the fact that this multigene family has undergone a series of tandem duplications events, some very recent, throughout the evolutionary history of mammals (Modi and Yoshimura 1999; Tasaki et al. 1999).

One important question involves the tissue location and timing of expression, and patterns of inducibility of each of these two genes. Earlier studies showed that ACT-2 (and/or LAG-1) is not strongly expressed in resting lymphocytes but is rapidly induced in stimulated T cells, B cells, and monocytes (Napolitano et al. 1991; Ziegler et al. 1991), and in CD3⁻ natural killer cells (Baixeras et al. 1990). These studies used full-length cDNA clones as hybridization probes in Northern blots, which would hybridize to mRNA from both genes, and would not pick up low-level transcripts.

Another approach for evaluating expression is to examine the NCBI human expressed sequence tagged (EST) database. Using the LAG-1 sequence of Fig. 1 as a query sequence, the 28 best matches were tabulated (Table 2). Inspection of these revealed three important observations.

1. All EST clones could be classified as either LAG-1 or ACT-2 based upon the intergenic differences outlined in Fig. 1. Furthermore, there is no evidence that more than two expressed sequences exist in the database.
2. Both LAG-1 and ACT-2 are expressed in a greater variety of tissues (e.g., parathyroid tumor, prostate, kidney, pineal gland, spleen, germ cell tumors) than was previously expected (this could be due to lymphocyte contamination).
3. Five of the LAG-1 clones derived from four different libraries are alternatively spliced transcripts. Each of these clones is missing exon 2 that contains 115 nucleotides (Table 2). Not only is the predicted protein shorter due to the loss of exon 2, but since the exon 1/2 and 2/3 junctions split codons, there is a frameshift beginning with exon 3. This results in a different amino acid sequence encoded by exon 3, the utilization of an upstream termination codon, and the deletion of two terminal amino acids (Fig. 2). From Fig. 2 it is important to also note that the alternatively spliced protein is missing three of the four cysteine residues critical for intrastrand disulfide bonding. If proper folding is prevented, this shorter protein may not bind CCR5.

Both LAG-1 (Baixeras et al. 1990) and ACT-2 (Napolitano et al. 1991) have been assigned to the long arm of Chromosome (Chr) 17 using in situ chromosome hybridization. The ACT-2 gene is located in the region 46,388–48,875 bp on the 123,925-bp AC003976 clone; however, LAG-1 is not found on this clone. A search of the NCGBI website for Chr 17 (http://www.ncbi.nlm.nih. gov/genome/seq/chr.cgi?CHR=17&SRT=ppos&MIN=100 &ORG=Hs) reveals that neither AC003976 nor X53682 is found on any larger genomic contig. This means that the closest these two genes could be to one another is about 46 kb. This is consistent with the results of Tasaki and co-workers (1999) who presented a physical map of the beta chemokine cluster based upon BAC and YAC clones. They showed that AT744.2 (LAG-1) and MIP-1b/AT744.1 (ACT-2) are separated by the LD78b and LD7g5′F genes and are about 80 kb apart.

One important issue concerns the gene duplication event creating LAG-1 and ACT-2. Aligning X53682 and AC003976 reveals an overall similarity of about 95.6%. This means 2.2% divergence along each branch of an evolutionary tree since the two genes last shared a common ancestor. If one assumes a neutral substitution rate of 0.2–0.4% per site per million years (Wu and Li 1985), then the duplication event creating LAG-1 and ACT-2 occurred 6–12 million years ago (MYA). Old World monkeys are thought to have last shared a common ancestor with anthropoid apes about 30 MYA, and humans with gorilla and chimpanzee, 5–12 MYA (Sibley and Alquist 1987). This predicts that the duplication event coincided

Table 2 Expressed sequence tagged (EST) sequences for the *LAG-1* and *ACT-2* genes. *Asterisked* entries are alternatively spliced clones lacking exon 2 (see Fig. 2)

Accession number	Insert size (bp)	Tissue of origin	cDNA library
LAG-1			
AI031546	607	Parathyroid tumor	Soares parathyroid tumor NbHPA
AI523695	632	Prostate	NCI-CGAP Pr28
AW293656	466	Mixture of 21 normalized or subtracted libraries	NCI-CGAP Pr28
AA767773	492	Tonsillar cells enriched for germinal-center B cells	NCI-CGAP GCB1
AI669076	527	Prostate	NCI-CGAP Pr28
AI261560	509	Kidney	NCI-CGAP Kid11
AI224410	388	Mixed small- and large-cell follicular lymphoma	NCI-CGAP Lym12
AW088643	403	Pooled, mixed ovarian tumors	NCI-CGAP Ov23
AI928485	456	Kidney	NCI-CGAP Kid11
AA701133	463	Pineal gland	Soares pineal gland N3HPG
AA777287	503	Liver and spleen	Soares fetal liver spleen 1NFLS S1
AI765277*	471	Two pooled kidney clear-cell tumors	NCI-CGAP Kid12
AI688619*	484	Twelve pooled colon tumors	NCI-CGAP Co3
AI688323*	454	Twelve pooled colon tumors	NCI-CGAP Co3
AI436708*	472	Kidney	NCI-CGAP Kid11
AI301910*	455	Two pooled kidney clear-cell tumors	NCI-CGAP Kid5
ACT-2			
AW006861	609	Kidney	NCI-CGAP Kid11
AI302152	642	Two pooled kidney clear-cell tumors	NCI-CGAP Kid5
AI862912	459	Kidney	NCI-CGAP Kid11
AA887236	550	Neuroendocrine lung carcinoid	NCI-CGAP Lu5
AW515711	649	Pooled germ-cell tumors	NCI-CGAP GC6
AI950224	648	Pooled germ-cell tumors	NCI-CGAP GC6
AW242796	458	Kidney	NCI-CGAP Kid11
AI341064	455	Pooled germ-cell tumors	NCI-CGAP GC6
AI343539	459	Two pooled kidney clear-cell tumors	NCI-CGAP Kid5
AI347734	469	Two pooled kidney clear-cell tumors	NCI-CGAP Kid5
AI765207	528	Two pooled kidney clear-cell tumors	NCI-CGAP Kid5
AA362106	520	Macrophage	Macrophage I

```
      Exon 1
  1 AGCACAGGACACAGCTAGGTTCTGAAGCTTCTGAGTTCTGCAGCCTCACCTCTGAGAAAACCTCTTTGCCACCAATACCATGAAGCTCTGCGTGACTGTC
                                                                                     m  k  l  c  v  t  v
                                                   Exon3
101 CTGTCTCTCCTCGTGCTAGTAGCTGCCTTCTGCTCTCTAGCACTCTCAGCACCAAATTCCAAACCAAAAGAGGCAAGCAAGTCTGCGCTGACCCCAGTGA
     l  s  l  l  v  l  v  a  a  f  c  s  l  a  l  s  A  P  N  S  K  P  K  E  A  S  K  S  A  L  T  P  V  S
201 GTCCTGGGTCCAGGAGTACGTGTATGACCTGGAACTGAACTGAGCTGCTCAGAGACAGGAAGTCTTCAGGGAAGGTCACCTGAGCCTGGATGCTTCTCCA
     P  G  S  R  S  T  C  M  T  W  N
301 TGAGCCGCATCTCCTCCATACTCAGGACTCCTCTCCGCAGTTCCTGTCTCTTCTCTTAATGTAATCTCTTTTATGTGCTGTATTATTGTATTAGGTGTTA
401 TTTCCATTATTTATATTAGTTTAGCCAAAGGATAAGTGTCCCCTA
```
LAG-1 *alternatively spliced form*

Fig. 2 Nucleotide and putative amino acid sequence of the alternatively spliced *LAG-1* transcripts lacking exon 2 observed in the NCBI EST database. Initiation and termination codons for the full-length transcripts from Fig. 1 are *underlined* and in **bold**. The upstream termination codon used in the alternatively spliced form is *double underlined*. The first nucleotide of exon 3 (A) is indicated

with the radiation of the anthropoids. To more accurately date the duplication event and determine which gene is ancestral, requires comparing *LAG-1* and *ACT-2* sequences from other species.

Unfortunately, inspection of GenBank indicates that these genes are unavailable for any other primate species. However, there are five mouse (M23503, M35590, X62502, AF128218, AF128219), one rat (U06434), and one rabbit (D17402) MIP-1β sequence available. Aligning these amino acid sequences reveals no variation among the five mouse entries, but a number of differences between mouse MIP-1β, rat MIP-1β, rabbit MIP-1β, human ACT-2, and human LAG-1 (Fig. 3). Examination of the amino acid composition of the three positions distinguishing the two human genes in the three animal species shows that residue 12 is uninformative while positions 20 and 70 provide conflicting interpretations regarding ancestry. Thus, due to the limited amount of comparative sequence data for these genes, we are unable to determine whether *LAG-1* or *ACT-2* is the ancestral human gene. Possibly, both of these, or other closely

Fig. 3 Amino acid sequence alignment of the human LAG-1 (X53682), human ACT-2 (AC009736), rat (D17402), mouse (M23503), and rabbit (U06434) MIP-1β proteins carried out using the Wisconsin Package (1997) PILEUP program

```
               1                                                    50
 human LAG-1   MKLCVTVLSL LVLVAAFCSL ALSAPMGSDP PTACCFSYTA RKLPRNFVVD
 human ACT-2   MKLCVTVLSL LMLVAAFCSP ALSAPMGSDP PTACCFSYTA RKLPRNFVVD
   rat MIP-1b  MKLCVSAFSL LLLVAAFCDS VLSAPIGSDP PTSCCFSYTS RKIHRNFVMD
 mouse MIP-1b  MKLCVSALSL LLLVAAFCAP GFSAPMGSDP PTSCCFSYTS RQLHRSFVMD
rabbit MIP-1b  MKLGVTVLSV ALLVAALCPP ALSAPMGSDP PTACCFSYTL RKLPRHFVID

               51                                                   92
 human LAG-1   YYETSSLCSQ PAVVFQTKRG KQVCADPSES WVQEYVYDLE LN
 human ACT-2   YYETSSLCSQ PAVVFQTKRS KQVCADPSES WVQEYVYDLE LN
   rat MIP-1b  YYETSSLCSQ PAVVFLTKKG RQICADPSEP WVNEYVNDLE LN
 mouse MIP-1b  YYETSSLCSK PAVVFLTKRG RQICANPSEP WVTEYMSDLE LN
rabbit MIP-1b  YFETTSLCSQ PAVVFQTKKG RQVCANPSES WVQEYVDDLE LN
```

Table 3 Kimura (1983) protein distances between the human ACT-2, human LAG-1, and the MIP-1β proteins from three species of mammals. Distances are the corrected (based on Dayhoff PAM matrix) number of substitutions per amino acid

	Human ACT-2	Human LAG-1	Rat MIP-1β	Mouse MIP-1β	Rabbit MIP-1β
Human ACT-2	0.00	3.30	27.27	28.84	19.84
Human LAG-1		0.00	25.73	28.84	19.84
Rat MIP-1β			0.00	18.44	33.76
Mouse MIP-1β				0.00	33.76
Rabbit MIP-1β					0.00

related genes, are present in various species, but due to limited data collection they have not yet been isolated for any species except human.

To extend the analysis of the animal sequences, evolutionary distances were computed for all pairwise comparisons (Table 3). This table indicates that the human proteins are more similar to that of the rabbit than to those of rat or mouse. Furthermore, the rabbit and human proteins are about 20% diverged while those of rat and mouse show about 18.5% divergence. Primates (human) and lagomorphs (rabbit) are thought to have last shared a common ancestor about 75 MYA, while rat and mouse diverged about 15 MYA. If these dates are accurate, then the rate of protein evolution in the two rodents is about five times greater (18.5/15 vs 20/75) than that seen in human and rabbit. Accelerated rates of molecular evolution have been reported for many rodent genes (Wu and Li 1985), and this may reflect the extraordinary ability of rodents to diversify rapidly and expand into novel ecological niches. However, the genes being compared between species may not be orthologous (resulting from speciation), but are, rather, paralogous (resulting from duplication).

Gene nomenclature

The formal gene symbols *SCYA4* and *SCY4L* are currently used to reference the following clones or genes: *CCL4*, *CCL4L*, *AT744.1*, *AT744.2*, *ACT-2*, *G-26*, *HC21*, *H400*, *MIP-1b*, *LAG-1*, *SISg*, *MAD-5* (see http://cytokine.medic.kumamotou.ac.jp/CFC/CK/CCG/CCG.html). One option for future naming is that *MIP-1*β be synonymous with *SCYA4* and *ACT-2*, and *LAG-1* with *SCYA4L*. It would probably be least confusing if the names *MIP-1*β and *LAG-1* were used in the future. It may be noted that commercially available MIP-1β protein from R&D Systems (Minneapolis, Minn; catalogue

No. 271-BME), or PeproTech (Rocky Hill, N.J.; catalog No. 300-09) is ACT-2.

References

Baixeras E, Roman-Roman S, Jitsukawa S, Genevee C, Mechiche S, Viegas-Pequignot E, Hercend T, Triebel F (1990) Cloning and expression of a lymphocyte activation gene (LAG-1). Mol Immunol 27:1091–1102

Brown KD, Zurawski SM, Mosmann TR, Zurawski G (1989) A family of small inducible proteins secreted by leukocytes are members of a new superfamily that includes leukocyte and fibroblast-derived inflammatory agents, growth factors, and indicators of various activation processes. J Immunol 142:679–687

Chang HC, Reinherz EL (1989) Isolation and characterization of a cDNA encoding a putative cytokine which is induced by stimulation via the CD2 structure on human T lymphocytes. Eur J Immunol 19:1045–1051

Davatelis G, Tekamp-Olson P, Wolpe SD, Hermsen K, Luedke C, Gallegos C, Coit D, Merryweather J, Cerami A (1988) Cloning and characterization of a cDNA for murine macrophage inflammatory protein (MIP), a novel monokine with inflammatory and chemokinetic properties [published erratum appears in J Exp Med (1989) 170:2189]. J Exp Med 167:1939–1944

Haskill S, Peace A, Morris J, Sporn SA, Anisowicz A, Lee SW, Smith T, Martin G, Ralph P, Sager R (1990) Identification of three related human GRO genes encoding cytokine functions. Proc Natl Acad Sci USA 87:7732–7736

Hogan M, Sherry B, Ritchlin C, Fabre M, Winchester R, Cerami A, Bucala R (1994) Differential expression of the small induc-

ible cytokines GRO alpha and GRO beta by synovial fibroblasts in chronic arthritis: possible role in growth regulation. Cytokine 6:61–69

Irving SG, Zipfel PF, Balke J, McBride OW, Morton CC, Burd PR, Siebenlist U, Kelly K (1990) Two inflammatory mediator cytokine genes are closely linked and variably amplified on chromosome 17q. Nucleic Acids Res 18:3261–3270

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge, UK

Laurence JS, Blanpain C, Burgner JW, Parmentier M, LiWang PJ (2000) CC chemokine MIP-1 beta can function as a monomer and depends on Phe13 for receptor binding. Biochemistry 39:3401–3409

Lipes MA, Napolitano M, Jeang KT, Chang NT, Leonard WJ (1988) Identification, cloning, and characterization of an immune activation gene. Proc Natl Acad Sci USA 85:9704–9708

Lodi PJ, Garrett DS, Kuszewski J, Tsang ML, Weatherbee JA, Leonard WJ, Gronenborn AM, Clore GM (1994) High-resolution solution structure of the beta chemokine hMIP-1 beta by multidimensional NMR. Science 263:1762–1767

Miller MD, Hata S, De Waal Malefyt R, Krangel MS (1989) A novel polypeptide secreted by activated human T lymphocytes. J Immunol 143:2907–2916

Modi WS, Yoshimura T (1999) Isolation of novel GRO genes and a phylogenetic analysis of the CXC chemokine subfamily in mammals. Mol Biol Evol 16:180–193

Napolitano M, Modi WS, Cevario SJ, Gnarra JR, Seuanez HN, Leonard WJ (1991) The gene encoding the Act-2 cytokine: genomic structure, HTLV-I/Tax responsiveness of 5′ upstream sequences, and chromosomal localization. J Biol Chem 266: 17531–17536

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc Natl Acad Sci USA 86:2766–2770

Schall TJ (1991) Biology of the RANTES/SIS cytokine family. Cytokine 3:165–183

Sherry B, Tekamp-Olson P, Gallegos C, Bauer D, Davatelis G, Wolpe SD, Masiarz F, Coit D, Cerami A (1988) Resolution of the two components of macrophage inflammatory protein 1, and cloning and characterization of one of those components, macrophage inflammatory protein 1 beta. J Exp Med 168: 2251–2259

Sibley CG, Alquist JE (1987) DNA hybridization evidence of hominoid phylogeny: results from an expanded data set. J Mol Evol 26:99–121

Tasaki Y, Fukuda S, Lio M, Miura R, Imai T, Sugano S, Yoshie O, Hughes AL, Nomiyama H (1999) Chemokine PARC gene (SCYA18) generated by fusion of two MIP-1alpha/LD78alpha-like genes. Genomics 55:353–357

Wisconsin Package (1997) Version 9.1. Genetic Computer Group (GCG), Madison, Wis

Wolpe SD, Davatelis G, Sherry B, Beutler B, Hesse DG, Nguyen HT, Moldawer LL, Nathan CF, Lowry SF, Cerami A (1988) Macrophages secrete a novel heparin-binding protein with inflammatory and neutrophil chemokinetic properties. J Exp Med 167:570–581

Wu CI, Li WH (1985) Evidence for higher rates of nucleotide substitution in rodents than in man. Proc Natl Acad Sci USA 82:1741–1745

Yoshida T, Imai T, Takagi S, Nishimura M, Ishikawa I, Yaoi T, Yoshie O (1996) Structure and expression of two highly related genes encoding SCM-1/human lymphotactin. FEBS Lett 395:82–88

Ziegler SF, Tough TW, Franklin TL, Armitage RJ, Alderson MR (1991) Induction of macrophage inflammatory protein-1 beta gene expression in human monocytes by lipopolysaccharide and IL-7. J Immunol 147:2234–2239

Zipfel PF, Balke J, Irving SG, Kelly K, Siebenlist U (1989) Mitogenic activation of human T cells induces two closely related genes which share structural similarities with a new family of secreted factors. J Immunol 142:1582–1590

Note added in proof A BLAT search of the University of California at Santa Cruz Human Genome Project Working Draft web site [http://genome.ucsc.edu/] reveals that ACT-2 and LAG-1 are located on chromosome 17 about 651 kb apart.