Structure, 5′-Flanking Sequence, and Chromosome Location of the Human *N*-Formyl Peptide Receptor Gene. A Single-Copy Gene Comprised of Two Exons on Chromosome 19q.13.3 That Yields Two Distinct Transcripts by Alternative Polyadenylation[†,‡]

David L. Haviland, Angela C. Borel, Daniel T. Fleischer, Joie C. Haviland, and Rick A. Wetsel*

Departments of Pediatrics and Molecular Microbiology, Washington University School of Medicine, St. Louis, Missouri 63110

Received December 29, 1992; Revised Manuscript Received February 17, 1993

ABSTRACT: The *N*-formyl peptide chemoattractant receptor (fMLF-R) is a cell-surface, G-protein-coupled glycoprotein that mediates the directed locomotion of neutrophils upon binding N-formylated peptides. The fMLF-R is encoded primarily by a 1.6-kb mRNA in differentiated HL-60 and U937 cells, although larger less abundant transcripts are present. To study the origin of different fMLF-R transcripts, the genetic linkage of chemotactic receptor genes, and the regulation of fMLF-R gene expression, we determined the copy number, chromosomal location, structural organization, and 5′-flanking sequence of the human fMLF-R gene. *Bam*HI restriction fragments derived from a human fMLF-R genomic cosmid clone were isolated, subcloned, and sequenced. These data indicate that the fMLF-R structural gene is ~7.5 kb in length and is comprised of two exons separated by an ~5.0-kb intron. The first exon encodes 66 bp of the 5′-untranslated sequence, while exon 2 encodes the coding and 3′-untranslated sequences. The genomic organization of the fMLF-R gene is similar to that of the adrenergic β-1 and β-2 G-protein-coupled receptor genes in that the coding sequence is contained in a single exon. The different 3′-untranslated sequences observed in fMLF-R cDNA clones are contiguous in the genomic structure, thereby indicating that these clones are derived in part by alternative polyadenylation. Southern blot analysis using human X hamster somatic cell hybrids and in situ hybridization indicated that the h-fMLF-R gene is located on chromosome 19q13.3. Primer extension experiments using dbcAMP-differentiated U937 RNA indicated a single transcriptional initiation site. Sequence analysis 5′ of the transcriptional initiation site indicated possible cis-acting motifs that may regulate fMLF-R gene expression. These included AP-1 and CK-2 consensus sequences that bind nuclear factors of the Fos/Jun family and NF-GMb, respectively.

The directed migration of polymorphonuclear neutrophils can be mediated by a number of chemotactic factors that are released as a consequence of inflammation. Such factors include platelet-activating factor (PAF) (Hanahan, 1986), leukotriene B4 (Goldman & Goetzl, 1982), the anaphylatoxin complement fragment C5a (Chenoweth & Hugli, 1978), interleukin-8 (NAP-1/IL-8) (Baggiolini et al., 1989), and N-formylated peptides.

In addition to chemotaxis, these factors also mediate a variety of cellular and biochemical responses in neutrophils and macrophages. Such changes include bacteriocidal superoxide radical production, proteolytic enzymes released from granules, aggregation, and phagocytosis [reviewed by Snyderman and Pike (1984)]. These inflammatory factors mediate their responses through GTP-binding proteins, and the subsequent signal transduction can be abrogated by pertussis toxin (Bokoch & Gilman, 1984).

The human lymphoma cloned cell lines U937 (Sundstrom & Nilsson, 1976) and HL-60 (Collins et al., 1977) have been used extensively to study the receptors that bind the chemotactic factors. These cells can be differentiated toward macrophage-like cells with phorbol esters or toward cells with the phenotype of monocytes and neutrophils with dibutyryl-cAMP (Harris & Ralph, 1985). The receptor that binds the N-formylated peptide f-Met-Leu-Phe (fMLF-R) has recently been cloned as two distinct cDNAs (R26 and R98) from a library prepared using RNA from dbcAMP-differentiated HL-60 cells (Boulay et al., 1990b; Murphy & McDermott, 1990). Expression of the fMLF-R mRNA is increased dramatically upon differentiation of HL-60 cells with db-cAMP. In addition, higher molecular weight messages that hybridize to the fMLF-R cDNA are observed by Northern analysis and cDNA cloning in the differentiated cells (Boulay, 1990b). It has not been determined, however, if these higher weight messages are transcribed from structurally homologous genes or if these higher weight messages are products of alternative processing of the fMLF-R primary transcript. In addition, it has yet to be determined if transcription of these messages is coordinately regulated by the same promoter or if each transcript is regulated by a unique promoter. Accordingly, we have isolated and characterized a fMLF-R genomic clone isolated from a human cosmid library so that (1) the origin of the different fMLF-R transcripts can be investigated, (2) the fMLF-R gene structure can be delineated, (3) the chromosomal location of the fMLF-R gene can be determined, and (4) promoter and regulatory cis elements controlling fMLF-R gene expression can be examined.

[†] This work was supported by U.S. Public Health Service Grants AI25011, AI24836, and HL37951. D.L.H. was supported by National Institutes of Health Training Grant AI07163. R.A.W. was the recipient of Research Career Development Award AI00919 from the National Institutes of Health.

[‡] The nucleotide sequences in this paper have been deposited in GenBank under Accession Numbers L12213 and L12214.

* To whom correspondence should be addressed at the Department of Pediatrics, Box 8116, Washington University School of Medicine, 400 S. Kingshighway Blvd., St. Louis, MO 63110.

0006-2960/93/0432-4168$04.00/0

EXPERIMENTAL PROCEDURES

Materials and Reagents. Except where noted, modifying enzymes and other molecular biology reagents were purchased from Promega Biotec, Madison, WI, Boehringer Mannheim, Indianapolis, IN, or Stratagene, La Jolla, CA, and used according to the manufacturer's instructions. Modified T7 DNA polymerase was obtained from U.S. Biochemical Corp., Cleveland, OH. Hybond-N+ was obtained from Amersham Corp. Radionucleotides [α-^{32}P]dCTP (3000 Ci/mmol) and [γ-^{32}P]ATP (7000 Ci/mmol) were purchased from ICN, Costa Mesa, CA, and [α-^{32}P]CTP (800 Ci/mmol) was from New England Nuclear, Boston, MA. All synthetic oligonucleotides were synthesized using a PCR Mate, Applied Biosystems, Foster CIty, CA. U937 and HL-60 cells were cultured in RPMI-1640 medium supplemented with 10% heat-inactivated (56 °C, 30 min) defined supplemented bovine calf serum (Hyclone Laboratories, Logan, UT), 50 μg/mL gentamycin, 2 mM L-glutamine, 1 mM nonessential amino acids, and 5 mM sodium pyruvate (Gibco, BRL, Gaithersburg, MD).

Northern Analysis. HL-60 and U937 cells were differentiated for 72 h in the presence of 1 mM dbcAMP (Sigma). Total RNA was isolated by the guanidine isothiocyanate/phenol method described by Chomczynski and Sacchi (1987). RNAs were subjected to electrophoresis on 1.2% agarose/formaldehyde gels, transferred to Hybond-N+ using the method of Virca et al. (1990), and probed with a random primed labeled (Feinberg & Vogelstein, 1983) 0.76-kb *Nco*I–*Pst*I h-fMLF-R cDNA restriction fragment.

cDNA and Genomic Libraries. Using synthetic oligonucleotides with their 5′-ends at positions 647 and 1111 (relative to the published sequence; Boulay et al., 1990a), a 464-bp h-fMLF-R cDNA fragment was generated by PCR using dbcAMP-differentiated cDNA as a template. This 464-bp fragment was then used as a probe to screen a cDNA library constructed from mRNA obtained from dbcAMP-differentiated U937 cells. This library was constructed in the p-Bluescript II vector as described (Wetsel et al., 1990). From this cDNA library, full-length h-fMLF-R cDNA clones were isolated. One clone that corresponded to R26 (Boulay et al., 1990b) was digested with *Eco*RI and used as a probe to screen a normal human fibroblast cosmid library (Johnson et al., 1992).

Subcloning of Cosmid Fragments. Unique restriction fragments that hybridized with h-fMLF-R cDNA fragments were isolated by preparative digestion of the cosmid clones with *Bam*HI. Individual fragments were isolated by low-melt agarose extraction and cloned into p-Bluescript SK (Stratagene).

DNA Sequence Analysis. All DNA sequencing was performed using double-stranded templates (Tabor & Richardson, 1987). Two micrograms of template was denatured in 0.2 M NaOH and 0.2 mM EDTA, neutralized, annealed with synthetic oligonucleotides (20-mers), and sequenced employing the dideoxy chain termination method (Sanger et al., 1977) and the modified bacteriophage T7 DNA polymerase (Sequenase). The 20-mers corresponding to the h-fMLF-R cDNA were made to coding and noncoding strands with the 5′-ends at intron/exon junctions. Subcloned genomic DNA was sequenced at least once in both directions.

Determination of the Length of the h-fMLF-R Intron. The length of the intron in the h-fMLF-R gene was determined by polymerase chain amplification strategies. Sense, W274 5′-(CCCAGACCTAGAACTACCCAGAGCAAGACCACAGCTGGTGAACAG) and anti-sense, W242 5′-(CTGACGTTTATGGTCTATTCACTACTATAGGTCCTTCTCT) oligonucleotides that correspond to DNA sequences bordering the fMLF-R intron were used as primers. One microgram of either human genomic or fMLF-R cosmid DNA was initially denatured at 95 °C for 3 min with 1 μg of each primer in a 100-μL solution containing 10 mM Tris, pH 8.3, 50 mM KCl, 2.5 mM $MgCl_2$, 0.1% gelatin, 200 μM dNTPs, and 3 units of Hot Tub polymerase (Amersham). Following the initial denaturation, the DNA was amplified by melting at 95 °C for 1 min, annealing at 68 °C for 2 min, and polymerizing at 72 °C for 5 min. Fifty cycles of amplification were performed using a Tempcycler (Coy Laboratory Products, Ann Arbor, MI), followed by a final elongation at 72 °C for 10 min.

Genomic Southern Analysis. Genomic Southern analysis was conducted as described (Haviland et al., 1991). Ten micrograms of either U937 or HL-60 high molecular weight DNA was digested overnight with the restriction endonucleases described in Figure 2A. All samples, including radiolabeled λ X *Hin*dIII and ϕX-174 X *Hae*III, were subjected to electrophoresis on a 1% agarose gel. Samples were then transferred to Hybond-N+ using the alkaline transfer protocol as per manufacturer's instructions. The DNA samples were cross-linked to the membrane with a Stratalinker (Stratagene) and prehybridized and probed using a random primed labeled 0.76-kb restriction fragment (*Nco*I–*Pst*I) of the h-fMLF-R cDNA.

Primer Extension Assay. Primer extension assays were performed as described (Kingston, 1989). Briefly, an antisense synthetic oligonucleotide W242 (40-mer) was made with its 5′-end at position 166 (Boulay et al., 1990b). Approximately 100 ng of the oligonucleotide was end labeled using [γ-^{32}P]ATP and polynucleotide kinase (30 units/μL) (U.S. Biochemicals, Cleveland, OH) for 1 h at 37 °C. Free [γ-^{32}P]-ATP was removed from the labeled oligonucleotide by G-25 spin column chromatography (Boehringer Mannheim Biochemicals) and the radioactivity quantitated. The end-labeled oligonucleotide (10^5 and 5×10^5 cpm) was hybridized to 50 μg of total RNA obtained from dbcAMP-differentiated U937 cells for 15 h at 50 °C in a 50-μL volume of formamide buffer (40 mM PIPES, 1 mM EDTA, 0.4 M NaCl, and 50% deionized formamide, pH 6.4). As a negative control, 50 μg of yeast tRNA (Boehringer Mannheim Biochemicals) was incubated with the radiolabeled oligonucleotide using the same hybridization conditions as used in the experimental samples. Ten micrograms of yeast tRNA was added to all samples, and the RNA was precipitated with 100% ethanol and a 25% volume of 3 M sodium acetate, centrifuged for 15 min at 10000*g*, washed with 70% ethanol, and air-dried. The RNA was then dissolved in 30 μL of 5 mM Tris (pH 8.2) containing 50 mM KCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, and 2 mM deoxynucleotide triphosphates. Thirty units of AMV reverse transcriptase (Life Sciences, St. Petersburg, FL) was added to each sample, and the solution was incubated at 42 °C for 90 min. The mRNA was then digested with RNase A (100 μg/mL) for 30 min at 37 °C. The samples were ethanol precipitated, and the extended primers were analyzed by electrophoresis on 6% acrylamide/urea gels. For comparative purposes, the oligonucleotide used in the primer extension analysis was used as sequencing primer with the cDNA.

Chromosome Assignment and in Situ Hybridization. A 0.76-kb (*Nco*I–*Pst*I) h-fMLF-R cDNA restriction fragment was random primed labeled and used to probe Southern blots of hamster X human somatic cell hybrid DNAs restricted with *Bam*HI. Two panels of 31 different hamster X human

somatic cell hybrids (Bios Corp., New Haven, CT) were hybridized as described above and were then scored for the presence or absence of the human fMLF-R gene, and the percentage asynteny was calculated for each of the human chromosomes. In situ hybridization (Bios Corp., New Haven, CT) was constructed using the method of Lichter et al. (1990) with minor modifications. The human fMLF-R cosmid was nick translated with digoxigenin-11-dUTP, combined with sheared human DNA, and hybridized to human metaphase chromosomes in 50% formamide, 2× SSC, and 10% dextran sulfate. Specific hybridization was detected with anti-digoxigenin antibodies conjugated to FITC, and the slides were then counterstained with propidium iodide and analyzed.

RESULTS AND DISCUSSION

Cloning of U937 fMLF-R cDNA. A cDNA library constructed using differentiated U937 mRNA was screened for fMLF-R clones as described under Experimental Procedures. Over 50 positive hybridizing clones were identified, 11 of which were plaque purified and subjected to Southern and partial sequence analyses. The longest of the 11 cDNA clones, UF1, was fully sequenced and was found to extend 16 bp upstream of the HL-60 cDNA clones R26 and R98 (Boulay et al., 1990b). The remaining sequence of UF1 was identical to those of the two HL-60 cDNAs except at base positions where variations existed in the R26 and R98 sequences. At these positions, UF1 shared characteristics of both HL-60 cDNAs. For example, UF1 was identical to R26 in the 5′-untranslated region and at nucleotide 301 (nonsilent variation). In addition, UF1 was polyadenylated at the same nucleotide position as R26. However, the UF1 sequence was identical to R98 at nucleotide positions 1037 (nonsilent coding substitution) and 1175 (base-pair deletion in the 3′-untranslated region).

Northern Analysis of HL-60 and U937 RNA. RNA from undifferentiated and differentiated HL-60 and U937 RNA was isolated and screened by Northern analysis for the presence of fMLF-R transcripts as described under Experimental Procedures. A faint signal of 1.6 kb was detected in the RNA of undifferentiated HL-60 and U937 cells (Figure 1). Upon differentiation, the abundance of the 1.6-kb message increased approximately 100-fold in both cell types. In addition, higher molecular weight messages of 2.3, 3.1, 4.8, and 6.0 kb were apparent in the differentiated U937 and HL-60 RNA samples. These data demonstrate that HL-60 and U937 cells express very similar sizes of RNA that hybridize to the fMLF-R cDNA. In addition, these data demonstrate that expression of the 1.6-, 4.8-, and 6.0-kb RNA increases similarly in both HL-60 and U937 cells when they differentiate toward cells with monocytic or neutrophilic phenotype; however, the abundance of the 2.3- and 3.1-kb RNA is markedly increased in the differentiated U937 cells over that observed in differentiated HL-60 cells.

Determination of the Copy Number of the Human fMLF-R Structural Gene. High molecular weight DNA from HL-60 cells was prepared and digested to completion with five restriction enzymes, subjected to agarose electrophoresis, blotted onto charged nylon membranes, and probed with a random primed labeled 0.76-kb fMLF-R cDNA restriction fragment (*Nco*I–*Pst*I). The results of this experiment are shown in Figure 2A. With each restriction digest, the probe hybridized to a single DNA band. Identical results have been obtained using high molecular weight genomic DNA isolated from U937 cells (data not shown). These results indicate that the human fMLF-R structural gene is present in the human genome as a single copy.

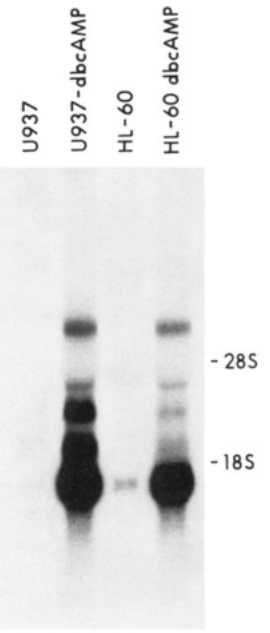

FIGURE 1: Northern blot of undifferentiated and differentiated HL-60 and U937 RNA hybriized with an fMLF-R cDNA probe. Samples (25 μg) of total RNA from promyelocytic HL-60 and U937 cells, either undifferentiated or differentiated for 72 h with 1 mM dbcAMP, were electrophoresed in a denaturing formaldehyde/agarose gel and transferred to Hybond-N+. The membrane was probed with a ^{32}P-labeled 0.76-kb (*Nco*I–*Pst*I) restriction fragment.

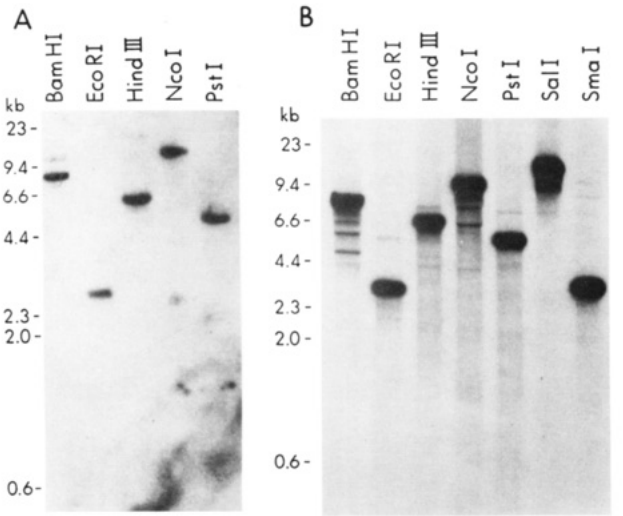

FIGURE 2: Southern blot analysis of the human fMLF-R gene. Panel A: Genomic Southern blot analysis of HL-60 DNA. Twenty micrograms of high molecular weight genomic DNA isolated from HL-60 cells was digested to completion with the restriction enzymes shown above, Southern blotted onto Hybond-N+, and probed with an [α-^{32}P]dCTP random primed labeled (*Nco*I–*Pst*I) restriction fragment of the fMLF-R cDNA. Panel B: Southern blot analysis of the h-fMLF-R-T2 cosmid. Four micrograms of cosmid DNA was digested to completion with the restriction enzymes shown above, Southern blotted onto Hybond-N+, and probed as described in panel A.

Isolation of the Human fMLF-R Gene. A human fibroblast genomic cosmid library was screened for fMLF-R clones using the UF1 full-length h-fMLF-R cDNA as a probe. One fMLF-R cosmid clone was identified and colony purified. Cosmid DNA that was isolated from this clone was digested with five erestriction enzymes (Figure 2B), subjected to agarose gel electrophoresis, blotted onto nylon membranes, and probed as was the blot in Figure 2A. The blot of the fMLF-R cosmid DNA gave the same banding pattern as the genomic DNA blot (panel A) when hybridized with the fMLF-R cDNA

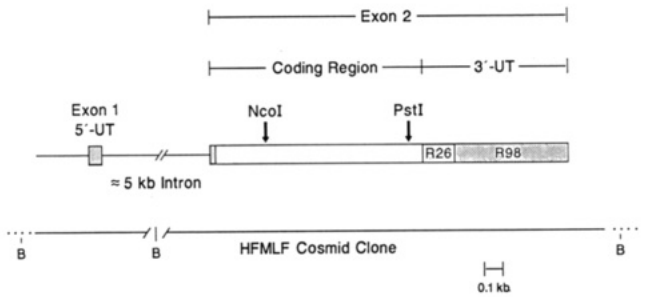

FIGURE 3: Structural organization of the human fMLF-R gene. The h-fMLF-R-T2 cosmid was digested with *Bam*HI, and the individual fragments that hybridized to the fMLF-R cDNA were subcloned into p-Bluescript-KS-II+ and sequenced using synthetic oligonucleotides as primers. The size of the single intron present in the fMLF-R gene was determined using polymerase chain reaction (PCR). The coding region is shown as an open box. The 5′-untranslated and alternatively processed 3′-untranslated regions are indicated by hatched boxes.

probe. These results demonstrate that the human fMLF-R cosmid DNA is similar in structure to the fMLF-R gene contained in the HL-60 and U937 cells and suggest that the entire human fMLF-R structural gene is contained in the fMLF-R cosmid clone.

Structural Organization of the h-fMLF-R Gene. The human fMLF-R cosmid was preparatively digested with *Bam*HI, and the individual fragments that hybridized to the fMLF-R cDNA were subcloned into p-Bluescript-KS-II+. The intron–exon boundaries and exon sequence were determined by DNA sequence analysis employing oligonucleotides as primers. The resulting of the mapping data are shown in Figure 3. The human fMLF-R structural gene (~7.5 kb) is encoded by only two exons. Exon 1 encodes 66 bp of 5′-untranslated sequence; exon 2 encodes 11 bp of 5′-untranslated sequence, the coding sequence, and 3′-untranslated sequences. Exons 1 and 2 are separated by an intron of approximately 5 kb, as determined by PCR analysis (Experimental Procedures).

Allelic Variation and Alternative Polyadenylation of the Human fMLF-R Gene. Two fMLF-R cDNA clones, R26 and R98, have been isolated using mRNA isolated from HL-60 cells (Boulay et al., 1990b). These two cDNA differ in their 5′-untranslated sequences by a 16-bp deletion, in their coding sequences by two nonsilent base-pair changes, and in their 3′-untranslated sequences by a large variation in length and a single base-pair deletion. In addition, we report here the sequence of another distinct fMLF-R cDNA clone, UF1, that was isolated from a U937 library. As described above, UF1 contains a combination of the variations described in R26 and R98. It was not clear, until the present study, if these cDNAs are the products of alternative splicing, of allelic variations, or of three separate fMLF-R genes. The data presented here argue that the 5′ and coding sequence differences, as well as the 3-untranslated base-pair deletion, result from allelic variation because (1) the genomic and cosmid Southern mapping data demonstrate that the fMLF-R is encoded by only a single-copy gene and (2) there was no evidence in the Southern or genomic cloning data of multiple coding exons that would be required for alternative splicing. Recent sequence analyses of PCR-amplified genomic DNA from HL-60 and U937 cells indicate that there are at least three different allelic variations of the fMLF-R gene (data not shown). For example, U937 cells are homozygous for a single allelic variant that encodes the UF1 fMLF-R cDNA, while HL-60 cells are heterozygous for two other allelic variants that encode the R26 and R98 fMLF-R cDNAs.

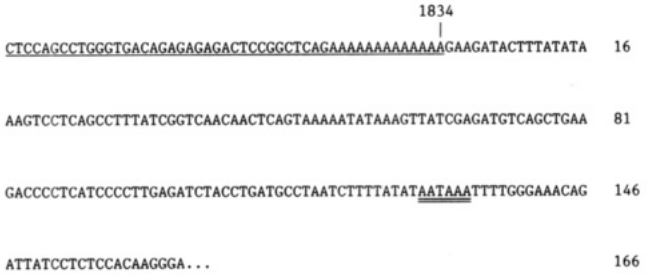

FIGURE 4: Sequence of the 3′-end of the human fMLF-R gene. The single-underlined sequence is that of the published fMLF-R R98 cDNA sequence (Boulay et al., 1990b). The double-underlined sequence is the consensus sequence for a polyadenylation signal.

There is apparently one sequence difference in the R98 fMLF-R cDNA that is not the result of allelic variation, however. This nonallelic difference is the additional 3′-untranslated sequence in the R98 cDNA. These additional 3′-untranslated nucleotides appear to be derived from alternative polyadenylation of the fMLF-R primary transcript. This conclusion is drawn from the fact that the different 3′-untranslated sequences are contiguous in the human fMLF-R gene structure (Figures 3 and 4).

Origin of Different fMLF-R cDNA Clones. As shown above (Figure 1), differentiated U937 and HL-60 RNA contain five RNA species that hybridized to the human fMLF-R cDNA probe under conditions of high stringency. The most intensely hybridizing RNA is 1.6 kb, while other messages of 2.3, 3.1, 4.8, and 6.0 kb are detected. From all experimental evidence, the UF1 and R26 cDNA clones are derived from the 1.6-kb message, since these cDNAs are found in high abundance in U937 and HL-60 cDNA libraries and are of the appropriate size [accounting for the poly(A) tail]. The 1.9-kb R98 cDNA was felt originally to be too small to be derived from the 2.3-kb message; however, the fMLF-R genomic sequence indicates that the R98 clone described by Boulay et al (1990b) was primed during first-strand synthesis in an A-rich region of the 3′-untranslated sequence of the mRNA instead of the poly(A) tail (Figure 4). In the genomic sequence there is a polyadenylation signal 128 bp downstream of the A-rich region. It is possible, therefore, that the R98 cDNA was derived from the 2.3-kb mRNA, although it cannot be excluded at the moment that the 2.3-kb message seen on Northern blots instead is an fMLF-R homologue. Presently, cDNAs corresponding to the 3.1-, 4.8-, and 6.0-kb mRNAs have not been characterized.

Determination of the Transcription Initiation Site. To determine the transcriptional initiation site in the fMLF-R gene, primer extension was performed as described under Experimental Procedures. The primer-extended products were subjected to electrophoresis adjacent to a DNA sequencing ladder using the same oligonucleotide as a primer and the full-length R26 fMLF-R cDNA in p-Bluescript-SK-II+ (Figure 5). One major transcriptional initiation site (TIS) was found and is designated as +1 in Figure 6. The 5′-end of our longest fMLF-R cDNA clone, UF1, isolated from the U937 library corresponded exactly to the TIS.

Determination of the Human fMLF-R Genomic Flanking Sequence. To search for possible promoter and regulatory cis elements upstream of the fMLF-R structural gene, the DNA sequence 5′ of the TIS was determined from a *Bam*HI subclone of the h-fMLF genomic cosmid clone. Three hundred and ninety-four base pairs upstream of the TIS were sequenced at least twice on both strands, and the sequence is shown in Figure 6. DNA sequence motifs are not present that clearly

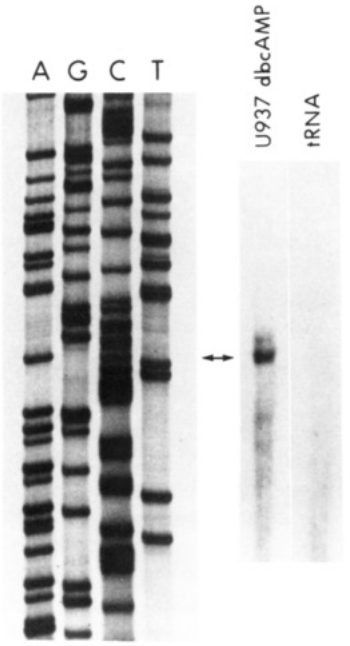

FIGURE 5: Determination of the transcriptional initiation site in the fMLF-R gene by primer extension. For the primer extension analysis a ^{32}P-labeled oligonucleotide, complementary to the 5′-translated end of the human fMLF-R mRNA, was hybridized to 50 μg of dbcAMP-differentiated U937 mRNA and to 50 μg of yeast tRNA as control. The primer-extended products were subjected to electrophoresis adjacent to a DNA sequencing ladder using the same oligonucleotide as a primer and the fMLF-R cDNA clone.

```
ctcaccacagtgcttgctgctttctttaccttgacccttggagggagcaggggcccggacacttggat   -270

ttcttggcccttgttgttgagagcactgaacctctgcatcccacagagactgaggctgagaaatacag   -202

 CK-2
tcaggtacatgagtttctaaacaggcccagcccactgtcctaatgccattaaagcagacagtatattg   -134

                                  NF-κB
gtgtattcttggggccatcaaaaatcagaagaagctcagacttcctatttcctgctacccagctggtt   -66

                                                                AP-1
tcagttcctttacccctcctcctgttccttggtgtatgttttgctgcaatcattagagcctgagtCAC   +3
                                                                  +1

                                                        (EXON 1)
TCTCCCCAGGAGACCCAGACCTAGAACTACCCAGAGCAAGACCACAGCTGGTGAACAGTCCAGgtaag   +66

            INTRON 1                    (EXON 2)    M  E  T  N  S
aaacatattc... (~5 kb) ...actgcactatttcagGAGCAGACAAGATGGAGACAAATTCC..   +91
```

FIGURE 6: Nucleotide sequence of the 5′-flanking region of the human fMLF-R structural gene. The sequence shown is derived from a *Bam*HI subcloned fragment derived from the h-fMLF-R-T2 cosmid genomic clone. Exon 1 and part of exon 2 are shown with a part of the intron (lowercase letters). The amino acid sequence encoded in exon 2 is shown above the appropriate codons. The major transcriptional initiation site is shown as +1. Sequences that correspond to AP-1, CK-2, and NF-κB consensus sequences are boxed.

indicate that the 5′-flanking sequence contains a cis promoter that regulates fMLF-R gene expression. For example, classical "CAAT" and "TATA" boxes are not present at positions –20 to –50 bp upstream of the TIS, nor are there CG islands that are commonly found in the promoters of many "house-keeping genes"; however, since fMLF-R expression is induced by phorbol esters and cAMP, there are sequence motifs in the fMLF-R 5′-flanking region which suggest that there are enhancer elements within this sequence that might regulate fMLF-R gene expression. For example, AP-1 and CK-2 consensus sequences are present at positions –1 to –7 and –196 to –202, respectively. The AP-1 sequence is very similar to the cyclic AMP response motif (CTGACGTCAG) and binds nuclear transcription proteins from the Fos and Jun family (Sassone-Corsi et al., 1990). The CK-2 sequence binds NF-GMb, a phorbol-inducible nuclear protein that regulates expression of the granulocyte/macrophage colony-stimulating factor gene (Shannon et al., 1988). In addition to the AP-1 and CK-2 sequence motifs, there is a sequence motif at positions –90 to –99 that is almost identical to the consensus sequence GRGRNTYYMY which binds the nuclear regulatory protein NF-κB (Serfling et al., 1989). Assessment of any actual role that these sequence motifs play in regulating expression of the human fMLF-R gene must await functional experiments performed with deletion and mutant reporter fusion gene constructs prepared from the fMLF-R 5′-flanking sequence.

Table I: Synteny Table of the Human fMLF Gene and Human Chromosomes in Hamster X Human Hybrid Cells[a]

	fMLF gene/human chromosome				
human chromosome	+/+	+/–	–/+	–/–	asynteny
1	3	4	0	17	17
2	0	7	1	16	33
3	0	7	4	13	46
4	1	6	1	16	29
5	7	0	15	2	63
6	1	6	3	14	38
7	2	5	0	17	21
8	1	6	4	13	42
9	0	7	3	14	42
10	0	7	3	14	42
11	1	6	2	15	33
12	2	5	2	15	29
13	5	2	2	16	17
14	5	2	2	15	17
15	2	5	2	17	29
16	0	7	2	15	38
17	1	6	1	16	29
18	1	6	3	14	38
19	6	1	0	17	4
20	1	6	2	15	33
21	5	2	2	15	17
22	2	5	1	16	25
X	0	7	3	14	42
Y	1	6	6	11	50

[a] +/+, hybridization signal and chromosome both present; +/–, hybridization signal present but chromosome absent; –/+, hybridization signal absent but chromosome present; and –/–, hybridization signal and chromosome both absent.

Chromosome Localization and in Situ Hybridization. The chromosome localization of the h-fMLF-R structural gene was determined using Southern blots of human X hamster somatic cell hybrids. Two panels of restricted (*Bam*HI) DNA were scored after hybridization with a 0.76-kb (*Nco*I–*Pst*I) h-fMLF-R cDNA fragment. The percentage synteny was calculated and is shown in Table I. The analysis indicated concordance between the presence of fMLF-R sequences and human chromosome 19 (Table I). The h-fMLF-R structural gene was further sublocalized by fluorescence in situ hybridization to metaphase chromosome preparations. The hybridization resulted in specific labeling of the long arm of chromosome 19 (Figure 7). Measurements of hybridized chromosomes indicated that the h-fMLF-R structural gene is located at band q13.3.

Comparison to Other Gene G-Protein-Coupled Proteins. The human fMLF-R gene shares similar genomic organization with some but not all G-protein-coupled receptors. For example, the rat β-1 (Shimomura & Terada, 1990), the rat β-2 (Buckland et al., 1990), and the human β-2 adrenergic

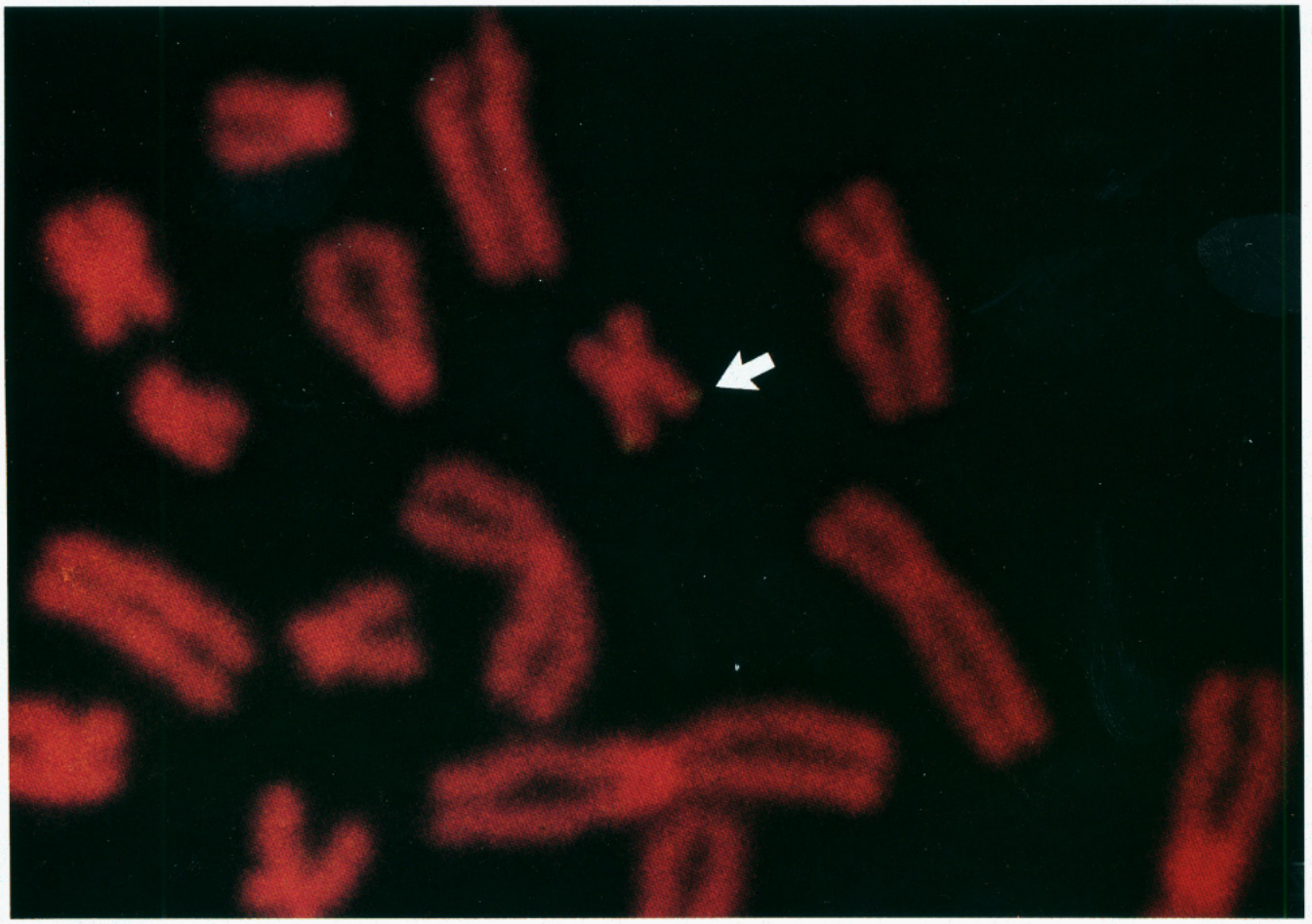

FIGURE 7: Chromosome localization of the human fMLF-R gene by in situ hybridization. The fluorescence hybridization of the nick-translated digoxigenin-11-dUTP human fMLF-R cosmid probe to human metaphase chromosomes is shown. The signal was detected using an anti-digoxigenin FITC antibody. The identity of chromosome 19 was confirmed by cohybridizing a genomic probe from the E2A locus previously localized to 19p13. This resulted in specific labeling of both 19p13 and 19q13 (data not shown), indicating that the h-fMLF-R structural gene resides on chromosome 19q13.

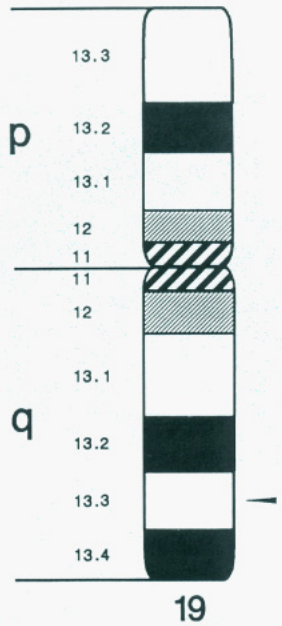

FIGURE 8: Schematic diagram of human chromosome 19 showing the band location of the human fMLF-R gene.

receptors (Emorine et al., 1987) are all coupled to effector systems using GTP-binding proteins, and all contain their protein coding sequences in a single exon. In contrast, the murine tachykinin receptors (substance K and substance P) (Sudelin et al., 1992) and the rat dopamine $D2_A$ and D_3 receptors (Sokoloff et al., 1990; O'Malley et al., 1990) are G-protein coupled, traverse the membrane seven times, but are interrupted with four and six introns, respectively.

There are several other eukaryotic genes that have intronless coding regions in addition to the fMLF-R and adrenergic receptor genes. These include the RNS2 and RNS3 genes of the RNase superfamily (Hamann et al., 1990) and the interferon $\alpha 1$ and $\beta 1$ genes (Nagata et al., 1980; Lawn et al., 1981). The RNS2 and RNS3 genes of the RNase superfamily are organized as the fMLF-R gene in that they have a single intron separating a 5′-untranslated exon from the coding exon; in contrast, the two interferon genes contain in a single exon all coding and untranslated sequences.

Concluding Remarks. The data described here indicate that the human fMLF-R gene is approximately 7.5 kb in length, is comprised of two exons, and is contained in a single copy on chromosome 19 band q13.3 (Figure 8). The genes encoding the chemotactic receptor C5a-R and two fMLF-R homologues, FPRH1 and FPRH2, have been localized also to chromosome 19, although their exact band locations were not determined (Murphy et al., 1992; Bao et al., 1992). We are currently investigating the possibility that the C5a-R, fMLF-R, FPRH1, and FPRH2 genes are tightly linked on chromosome 19 by chromosome walking studies employing genomic cosmid and YAC clones known to contain the human fMLF-R gene. In addition, this investigation could possibly

identify other genes that are involved in chemotaxis of inflammatory cells.

Delineation of the genomic structure indicated that the three characterized human fMLF-R cDNA clones (R26, R98, and UF1) are generated by a combination of allelic variation and alternative polyadenylation. The differences in the 5′-untranslated and coding regions result from allelic variation, while the differences in length of the 3′-untranslated sequences result from alternative polyadenylation. Northern blotting data indicate that as many as five distinct RNA (1.6, 2.3, 3.1, 4.8, and 6.0 kb) contained in differentiated HL-60 and U937 cells hybridize to the human fMLF-R cDNA. Collective evidence strongly suggests that R26 and UF1 cDNAs are derived from the 1.6-kb RNA and that possibly R98 is derived from the 2.3-kb RNA. It is not presently known how many of the other fMLF-R hybridizing RNA are derived from alternative processing of the fMLF-R gene and how many are derived from closely homologous fMLF-R genes. Now that the fMLF-R gene has been isolated, identifying which RNA are derived from the fMLF-R gene should be possible.

The determination of the human fMLF-R 5′-flanking sequence represents the first attempt to examine DNA sequences required for expression and regulation of genes encoding chemotactic receptors. Continued investigation of fMLF-R gene expression should result in the identification of cis and trans factors required for expression of many different chemotactic receptors, since many of these receptors are coordinately expressed and regulated by a similar range of inflammatory mediators.

ACKNOWLEDGMENT

We thank Drs. Harvey R. Colten and Robert C. Strunk for critical evaluation of the data and text and William Whitehead for determining some of the DNA sequence.

REFERENCES

Baggiolini, M., Walz, A., & Kunkel, S. L. (1989) *J. Clin. Invest. 84*, 1045–1049.

Bao, L., Gerard, N. P., Eddy, R. L., Jr., Shows, T. B., & Gerrard, C. (1992) *Genomics 13*, 437–440.

Bokoch, G. M., & Gilman, A. G. (1984) *Cell 39*, 301–308.

Boulay, F., Tardif, M., Brouchon, L., & Vignais, P. (1990a) *Biochem. Biophys. Res. Commun. 168*, 1103–1109.

Boulay, F., Tardif, M., Brouchon, L., & Vignais, P. (1990b) *Biochemistry 29*, 11123–11133.

Buckland, P. R., Hill, R. M., Tidmarsh, S. F., & McGuffin, P. (1990) *Nucleic Acids Res. 18*, 682.

Chenoweth, D. E., & Hugli, T. E. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 3943–3947.

Chomczynski, P., & Sacchi, N. (1987) *Anal. Biochem. 162*, 156–159.

Collins, S. J., Gallo, R. C., & Gallagher, R. E. (1977) *Nature 270*, 347–349.

Emorine, L. J., Marullo, S., Delavier-Klutchko, C., Kaveri, S. V., Durieu-Trautman, O., & Strosberg, A. D. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6995–6999.

Feinberg, A. P., & Vogelstein, B. (1983) *Anal. Biochem. 132*, 6.

Goldman, D. W., & Boetzl, E. J. (1982) *J. Immunol. 129*, 1600–1604.

Hamann, K. J., Ten, R. M., Loegering, D. A., Jenkins, R. B., Heise, M. T., Schad, C. R., Pease, L. R., Gleich, G. J., & Barker, R. L. (1990) *Genomics 7*, 535–546.

Hanahan, D. J. (1986) *Annu. Rev. Biochem. 55*, 483–509.

Harris, P., & Ralph, P. (1985) *J. Leukocyte Biol. 37*, 407–422.

Haviland, D. L., Haviland, J. C., Fleischer, D. T., & Wetsel, R. A. (1991) *J. Biol. Chem. 266*, 11818–11825.

Johnson, C. J., Densen, P., Hurford, R. K., Colten, H. R., & Wetsel, R. A. (1992) *J. Biol. Chem. 267*, 9347–9353.

Kingston, R. E. (1989) in *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K., Eds.) pp 4.8.1–4.8.3, Greene Publishing Associates and Wiley-Interscience, New York.

Lawn, R. M., Adelman, J., Franke, A. E., Houck, C. M., Gross, M., Najarian, R., & Goeddel, D. V. (1981) *Nucleic Acids Res. 9*, 1045–1052.

Lichter, P., Tang, C. C., Call, K., Hermanson, G., Evans, G. A., Housman, D., & Ward, D. C. (1990) *Science 247*, 64–69.

O'Malley, K. L., Mack, K. J., Gandelman, K. Y., & Todd, R. D. (1990) *Biochemistry 29*, 1367–1371.

Murphy, P. M., & McDermott, D. (1990) *J. Biol. Chem. 266*, 12560–12567.

Murphy, P. M., Ozcelik, T., Kennedy, R. T., Tiffany, H. L., McDermott, D., & Franke, U. (1992) *J. Biol. Chem. 267*, 7637–7643.

Nagata, S., Mantei, N., & Weissman, C. (1981) *Nature 287*, 401–408.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Sassone-Corsi, P., Ransone, L. J., & Verma, I. M. (1990) *Oncogene 5*, 427–431.

Serfling, E., Barthelmas, R., Pfeuffr, I., Schenk, B., Zarius, S., Swoboda, R., Mercurio, F., & Karin, M. (1989) *EMBO J. 8*, 465–473.

Shannon, M. F., Gamble, J. R., & Vadas, M. A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 674–678.

Shimomura, H., & Terada, A. (1990) *Nucleic Acids Res. 18*, 4591.

Snyderman, R., & Pike, M. C. (1984) *Annu. Rev. Immunol. 2*, 257–281.

Sokoloff, P., Giros, B., Martres, M. P., Bouthenet, M. L., & Schwartz, J. C. (1990) *Nature 347*, 146–151.

Sundelin, J. B., Provvedini, D. M., Wahlestedt, C. R., Laurell, H., Pohl, J. S., & Paterson, P. A. (1992) *Eur. J. Biochem. 203*, 625–631.

Sundstrom, C., & Nilsson, K. (1976) *Int. J. Cancer 17*, 565–577.

Tabor, S., & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4767–4771.

Virca, G. D., Northemann, W., Shiels, B. R., Widera, G., & Broome, S. (1990) *BioTechniques 8*, 370–371.

Wetsel, R. A., Fleischer, D. T., & Haviland, D. L. (1990) *J. Biol. Chem. 265*, 2435–2440.