Human Chemotaxis Receptor Genes Cluster at 19q13.3–13.4. Characterization of the Human C5a Receptor Gene†

Norma P. Gerard,*,‡ Lu Bao,‡ He Xiao-Ping,‡ Roger L. Eddy, Jr.,§ Thomas B. Shows,§ and Craig Gerard‡

Ina Sue Perlmutter Laboratory and Department of Pediatrics, Children's Hospital, Departments of Medicine, Beth Israel and Brigham and Women's Hospitals, Center for Blood Research, and Thorndike Laboratory of Harvard Medical School, Boston, Massachusetts 02115, and Department of Human Genetics, Roswell Park Memorial Institute, Buffalo, New York 14263

Received August 26, 1992; Revised Manuscript Received November 16, 1992

ABSTRACT: The human C5a anaphylatoxin and formyl peptide receptor genes, as well as two genes with high sequence identity to the formyl peptide receptor, FPRH1 and FPRH2, have been mapped to chromosome 19 (Lu et al., 1992). Further analysis reveals that these genes are present in the 19q13.3 band adjacent to the 13.3–13.4 interface. MRNAs for the C5a and formyl peptide receptors, as well as for FPRH1, are expressed in cAMP differentiated U937 cells and human eosinophils, while all four transcripts are expressed in human lung. This observation opens the possibility for coordinate regulation of these genes. In order to initiate the mapping of fine structure at this locus, genomic clones have been analyzed. All four of the genes have a similar structure, with the receptor protein encoded in a single exon. Detailed characterization of the C5a receptor gene reveals a two exon structure, with the 5′ untranslated sequence and initiating methionine located in the first exon. An intron of ~9 kb separates exon 1 from the receptor-encoding exon 2. The region of genomic DNA flanking the 5′ untranslated sequence possesses promoter activity when transfected into the myeloid-derived rat basophilic leukemia RBL-1 cells, but the same region is inactive when transfected into nonmyeloid cells. Deletional analyses indicate that C5a receptor 5′ flanking region contains both cell-specific suppressor and promoter regions.

Inflammation is classically characterized by swelling, redness, pain, and warmth. These symptoms and signs are related to the local influx of white blood cells and increased vascular permeability, which are brought about by an ever-growing list of humoral and tissue-derived mediators (Harvath, 1991; Oppenheim, et al., 1991). The application of expression cloning techniques has led to the identification of receptors for the major chemotactic and pro-inflammatory ligands including bacterial formyl peptides (Boulay et al., 1991a,b; Murphy & McDermott, 1991), platelet-activating factor (PAF)[1] (Nakamura et al., 1991; Ye et al., 1991; Kunz et al., 1991), members of the II-8/Gro family (Holmes et al., 1991; Murphy & Tiffany, 1991; Beckmann et al., 1991), and C5a anaphylatoxin (Gerard & Gerard, 1991; Boulay et al., 1991). Analysis of the predicted structures of these receptors indicates that all are members of the seven transmembrane segment receptor gene family which transduce signals via GTP-binding proteins (Dohlman et al., 1991; Gilman, 1987).

Despite obvious similarities in the structure of these receptors, however, the profiles for stimulation of the chemotactic response, activation of the NADPH oxidase, and degranulation of lysosomal particles are quite different for each agonist (Dohlman et al., 1991; Gilman, 1987). Multiple reasons may explain these findings. First, the number of receptor sites varies enormously, with C5a and formyl peptide receptors most abundant on polymorphonuclear neutrophils, at 100–200 000 sites/cell, followed by IL-8 with 20–40 000 sites/cell, and still lower numbers for the lipid receptors for PAF and leukotriene B_4 (LTB_4), at <10 000 sites/cell (Nelson et al., 1981, 1983; Hwang et al., 1986; O'Flaherty et al., 1986, 1990; 1991; Goldman et al., 1986; Gerard et al., 1989; Schepers et al., 1992). Coupling to different GTP-binding proteins may provide a second potential reason for the broad range of biologic effects or possibly receptor coupling with different affinities to a pool of transducing proteins (Goldman & Goetzel, 1984; McLeish et al., 1989; Beaulieu et al., 1992). Differential metabolism of ligands by proteases, lipases, or oxygenases provides yet another aspect of control (Ward & Ozols, 1976; Connelly et al., 1985; Tarbet et al., 1991; Soberman et al., 1988).

Biologically, the C5a and formyl peptide ligands behave most similarly when compared with other mediators. Analysis of the receptors for these molecules reveals that, while the ligands are quite dissimilar, a high degree of sequence identity occurs at the receptor level (Gerard & Gerard, 1991). These similarities have led us to investigate the genetic organization and control of these peptidergic receptors. Recently, we demonstrated that the human C5a (C5aR) and formyl peptide receptor (FPR) genes map to chromosome 19 (Lu et al., 1992). In the course of characterizing genomic clones encoding the formyl peptide receptor, we isolated two closely related genes, identified as homologues 1 and 2 (FPRH1 and FPRH2); neither one binds formyl peptides at physiologic concentrations when expressed in transfected cells. These genes are also located on chromosome 19 (Lu et al., 1992). In the present study, we demonstrate that all four of these genes in fact map to a similar locus at chromosome 19q13.3. In an effort to begin fine structure analysis of this chemotactic receptor cluster,

† This work was supported in part by NIH Grant HL 36162. C. Gerard is a recipient of a Faculty Scholar's Award from R. J. R. Nabisco.

‡ Children's Hospital, Beth Israel and Brigham and Women's Hospital, Center for Blood Research, and Thorndike Laboratory of Harvard Medical School.

§ Roswell Park Memorial Institute.

[1] Abbreviations: PAF, platelet-activating factor; LTB_4, leukotriene B_4; FPR, formyl peptide receptor; FPRH1, formyl peptide receptor homologue 1; FPRH2, formyl peptide receptor homologue 2; PCR, polymerase chain reaction; PMA, phorbol myristate acetate; PBS, phosphate-buffered saline; CAT, chloramphenicol acetyl transferase; TLC, thin-layer chromatography; cAMP, adenosine 3′–5′-cyclic monophosphate; AMV, avian myeloblastosis virus; EDTA, ethylenediaminetetraacetic acid; DMEM, Dulbecco's-modified essential medium; HEPES, *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfate; SDS, sodium dodecyl sulfate.

0006-2960/93/0432-1243$04.00/0

we have characterized the genomic organization for the C5a receptor. Additionally, a system has been found where expression of these myeloid receptor genes can be studied by transfection.

EXPERIMENTAL PROCEDURES

Materials. All reagents and supplies, unless specifically indicated, were as previously described for the characterization of the genes for human NK-1 and NK-2 receptors (Gerard et al., 1990, 1991).

Fluorescence in Situ Analysis. In situ hybridization to metaphase chromosomes was performed with biotinylated probes as described by Cherif et al. (1989), and Fan et al. (1990). Briefly, the ~1-kb cDNA encoding the protein sequence for the human C5a receptor or the FPRH1 homologue was subcloned into pBluescript vector, and biotinylated probe was prepared as described (Cherif et al., 1989; Fan et al., 1990). The FPRH1 cDNA hybridizes strongly to both FPR and FPRH2 as well as FPRH1 under conditions of high stringency. Both C5a and FPR mRNAs contain Alu repeat sequences in their 3′ untranslated regions, necessitating deletion of these sequences in order to achieve probe specificity. Metaphase chromosomes were prepared from 5-bromodeoxyuridine-synchronized lymphocyte cultures. Biotinylated receptor probes were hybridized to the chromosome spreads and detected by fluorescein-conjugated avidin (Vector Labs). Q and R banding, by counterstaining with 4′,6-diamidine-2-phenylindole dihydrochloride (DAPI) and propidium iodide, respectively, were used to confirm the identity of the chromosome. Slides were evaluated under fluorescence microscopy (Nikon).

PCR Amplifications. One to two micrograms of polyadenylated RNA from U937 human lymphoma cells, undifferentiated or differentiated with dibutyryl-cAMP as previously described (Gerard & Gerard, 1991), a mixture of 86% human peripheral blood eosinophils and ~14% neutrophils, or human lung parenchyma were each converted to cDNA (Boehringer Mannheim); 1–10% of the cDNA obtained was subjected to PCR using primers specific for C5aR, FPR, FPRH1, or FPRH2. Primer pairs were synthesized as previously described for C5aR (Gerard & Gerard, 1991). The formyl peptide receptor PCR primers corresponded to nucleotides 1–19 (sense: ATG GAG ACA AAT TCC TCT C) and 1037–1053 (antisense: TCA CTT TGC CTG TAA CGC), according to the numbering system described by Boulay et al. (1990b). Sense primers unique to FPRH1 and FPRH2 were synthesized from nucleotides 215–232 using the nomenclature of Lu et al. (1992) for the corresponding clones, M76672 and M76673 (Gen Bank accession numbers) (TTT CTT TCA CGG CCA CAT and TCT CTT TCA GTG CCA TCC for FPRH1 and FPRH2, respectively). Antisense probes corresponding to nucleotides 985–1012 (FPRH1; CTC ATT AGT TGG GCT GAG T) and 1043–1062 (FPRH2; TCA CAT TGC TTG TAA CTC C) were similarly prepared. For PCR, annealing was carried out for 2 min at 40 °C, extension for 3 min at 72 °C, and denaturation for 1 min at 94 °C. Typically, 25 cycles were performed. Aliquots of the PCR products were subjected to analytical agarose gel electrophoresis, and the specificity of the reactions was determined with isolated genes for the receptors and by Southern blotting.

Isolation of Genomic Clones. A library of human placental DNA cloned in EMBL3 was the generous gift of Dr. Stuart Orkin (Harvard Medical School). The library was screened as previously described (Gerard et al., 1990, 1991) using the ~1-kb cDNAs corresponding to the coding sequences for the human C5aR or FPR. High stringency washing conditions were employed, with 2× SCC, 0.1% SDS at 60 °C for 30–60 min followed by a final wash at room temperature using 0.2× SCC, 0.1% SDS. Positive clones were plaque-purified and subjected to restriction analysis to ensure fidelity with genomic DNA.

Genomic DNA Analysis. Aliquots of ~10 μg of human genomic DNA were subjected to restriction analysis. Digests were electrophoresed on 0.9% agarose gels, denatured with alkaline buffer and transferred to GeneScreen Plus membranes (New England Nuclear) by capillary blotting. Membranes were probed using cDNAs corresponding to the protein coding regions of the C5aR or FPR which were labeled with [^{32}P]-dATP by random priming (Boehringer Mannheim). The exon/intron structure of the C5a receptor gene was determined by restriction analyses and hybridization of Southern blots with ^{32}P-labeled cDNA or oligonucleotide probes.

DNA Sequencing. Genomic clones were sequenced extensively in the case of the C5a receptor, and in a limited fashion for the FPR receptor family, following subcloning of restriction fragments into pBluescript. Double-stranded dideoxy sequencing reactions were performed on cesium chloride-banded plasmids using Sequenase 2 and procedures recommended by the manufacturer (U.S. Biochemical Corp). Potential regulatory sequences in the C5a receptor 5′ flanking region were identified by computer analysis (MacVector 3.5, IBI).

Primer Extension. Primer extension was carried out as previously described for the human NK-2 receptor (Gerard et al., 1990). An antisense oligodeoxynucleotide corresponding to nucleotides 40–70 in the C5a receptor coding sequence was labeled with [γ-^{32}P]ATP and 2×10^5 cpm were annealed to 2 μg of cAMP-induced U937 cell poly(A+) RNA for 16 h at 65 °C in 30 μL of 1 M NaCl containing 167 mM HEPES, pH 7.5, and 0.3 mM EDTA. The annealed RNA/primer mixture was precipitated with ethanol from 0.3 M sodium acetate, redissolved in reverse transcriptase buffer (RT1, Boehringer Mannheim), containing 70 units of RNasin and 20 mM dNTPs, and extended with 5 units of AMV reverse transcriptase (Boehringer Mannheim) for 90 min at 42 °C. The remaining RNA was digested with DNase-free RNase A, extracted with phenol/chloroform, ethanol precipitated, and redissolved in 5 mM Tris-HCl, pH 7.5, containing 48% formamide, 10 mM EDTA, 0.025% bromophenol blue, and 0.025% xylene cyanol FF. The reaction products were analyzed on an 8% polyacrylamide 4 M urea gel and sized by co-electrophoresis with a DNA sequencing reaction.

Transfection and Promoter Analysis. Expression plasmids were constructed with the cDNAs corresponding to the coding regions in the vector pCDNA (InVitrogen). Flanking regions of the C5a receptor gene were cloned into the polylinker region of the reporter gene construct, pBLCAT3 (generous gift of Dr. M. Green, Harvard Medical School) (Luckow & Shutz, 1987). Clones were transfected in RBL-1 cells by electroporation. Cells were suspended at 1×10^7/mL in serum-free DMEM, and 0.4 mL was mixed with 10 μg of plasmid DNA in a 4-mm cuvette (Bio-Rad) and incubated for 10 min at 22 °C. A pulse of 200 V at 960 μF was applied (Bio-Rad Gene Pulser apparatus), and cells were plated after an additional 10 min at 22 °C in DMEM containing 10% fetal calf serum and returned to the incubator. After 48 h, PMA was added to a final concentration of 1 μM to some cells transfected with the C5aR 5′ flanking region. At 72 h after transfection, the cells were harvested for the chloramphenicol acetyl transferase (CAT) assay. Briefly, they were washed with PBS, released

from the dishes by scraping in TEN buffer (40 mM Tris-HCl, pH 7.5, containing 1 mM EDTA and 150 mM NaCl), transferred to 0.25 M Tris-HCl, pH 8.0, and lysed by three cycles of freezing and thawing. Mammalian acetylase activity was inactivated by heating for 10 min at 65 °C. CAT activity was determined by incubating aliquots of the cell lysates with 0.04 μg/mL *n*-butyryl-coenzyme A (Sigma) and 0.25 μCi of [^{14}C]chloramphenicol (NEC-408A, New England Nuclear) in a final volume of 125 μL of 0.25 M Tris-HCl, pH 8, at 37 °C for 16 h. Ethyl acetate extracts of the reaction mixtures were applied to silica gel TLC plates (Whatman 60A LK6D) and developed with chloroform/methanol 97:3 (v/v) to 60–75% of the length of the plate. Autoradiographs were prepared using Kodak X-Omat film for 12–24 h at room temperature and the extent of [^{14}C]chloramphenicol acetylated was quantitated by liquid scintillation counting of relevant areas of the plate. Positive (thymidine kinase promoter–enhancer/CAT, pBLCAT2) and negative (pBLCAT3) controls were included in each experiment.

Reporter constructs were also transfected into the human neuroblastoma SK-N-SH cell line using cationic liposomes as described by Rose et al. (1991). Liposomes consisted of a 1:0.4 (w/w) mixture of phosphatidylethanolamine and dimethyldioctadecylammonium bromide (Sigma) suspended by sonication in water. For each 10-cm dish of cells at approximately 70% confluence, 15 μg of plasmid DNA and 63 μg of liposomes (total lipid) were mixed with 1 mL of serum-free DMEM, allowed to coprecipitate at room temperature for 10–20 min, and then added to the cells in DMEM containing 10% fetal calf serum. In some experiments, the plasmid pRSV β-Gal was contransfected to control for transfection efficiency. After 48 h, the cells were harvested and assayed as described above for RBL-1 cells, but enzyme reactions were terminated after 2 h so that enzyme activity was maintained in the linear range. β-Galactosidase activity was assayed as previously described (Rosenthal, 1987).

Constructs with progressive deletions at the 5′ end of the C5a receptor promoter fragment were generated by first digesting the plasmid with *Hin*dIII, filling in with α-phosphorothiolate nucleotides to protect the vector, and then cutting with *Pst*I. DNA was ethanol precipitated, redissolved in 66 mM Tris-HCl, pH 8.0, containing 0.66 mM $MgCl_2$, and incubated with 500 units of exonuclease III (Promega) at 22 °C, removing aliquots at 30-s intervals from 1 to 12 min. Samples were treated with 20 units of S1 nuclease (Promega) in 40 mM potassium acetate, pH 4.6, containing 338 mM NaCl, 1.35 mM $ZnSO_4$, and 7% glycerol for 30 min at room temperature. Digestion was terminated by addition of 0.3 M Tris base, 0.05 M EDTA, and heating at 70 °C for 10 min. Samples were blunted by treatment with Klenow enzyme for 5 min at 37 °C, ligated with T4 DNA ligase for 1 h at room temperature, and transformed into JM109, selecting on LB agar containing 50 μg/mL ampicillin. Clones containing suitable deletions were identified by DNA sequence analysis, and large-scale plasmid preparations were purified by alkaline lysis and banding on CsCl gradients.

Cell Culture. U937 cells were cultured as previously described (Gerard & Gerard, 1990) in RPMI containing 10% fetal calf serum. Differentiation to a monocytic phenotype was accomplished by growing in 1 mM dibutyryl-cAMP (Sigma) for 3 days. TSA cells, a derivative of the 293 human embryonic kidney cell line, were obtained from Dr. John Didsbury, Duke University Medical Center. SK-N-SH cells, a human metastatic neuroblastoma cell line, were obtained from the ATCC. RBL-1 rat basophilic leukemia cells were the gift of Dr. Hava Avraham, New England Deaconess Hospital. The latter three cell lines were cultured in Dulbecco's-modified Eagle's medium containing 10% fetal calf serum.

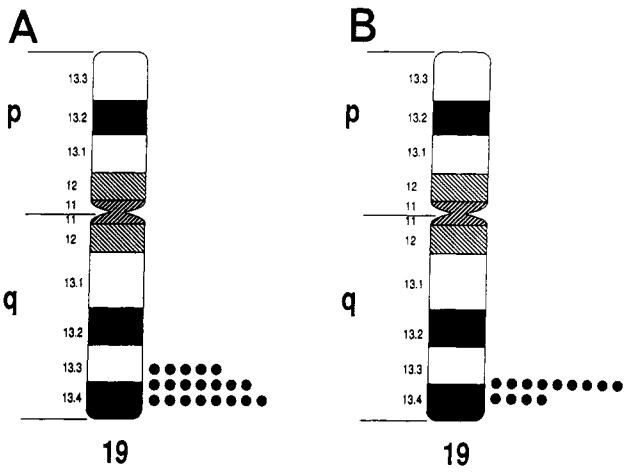

FIGURE 1: (A) Idiogram of the distribution of signals for FPR, FPRH1, and FPRH2 on both chromatids of chromosome 19. 67% of the cells examined (20/30) had signals on both chromatids of a single chromosome 19 at q13.3–q13.4. (B) Idiogram of the distribution of signals for C5aR on both chromatids of chromosome 19. 59% of the cells examined (13/22) had signals on both chromatids of a single chromosome 19 at q13.3–13.4.

RESULTS

Genes for C5aR and the FPR Map to Distal 19q13.3. Thirty metaphase chromosome preparations were examined following hybridization with the formyl peptide receptor homologue 1 (FPRH1) probe, which also hybridizes with the FPR and FPRH2 genes. A fluorescent signal was detected at the 19q13.3–19q13.4 interface on both chromatids of a single chromosome 19 in 20 out of 30 cell metaphases. Six of these 20 had an additional signal on one chromatid of the other chromosome 19 at the 19q13.3–13.4 interface. Eight of the remaining 10 other cells had a signal on one chromatid of both chromosomes 19 at the identical interface. The idiogram shown in Figure 1A demonstrates the localization of the formyl peptide receptor gene and its homologues.

In a similar fashion, the biotinylated C5a receptor probe was examined with 22 metaphase chromosomes. A fluorescent signal was detected at 19q13.3–19q13.4 on both chromatids of a single chromosome 19 in 13 out of the 22 cells. One of these 13 cells had an additional single signal on one chromatid of the other chromosome 19 at the 13.3–13.4 interface. Of the remaining nine other cells, five each had a signal on one chromatid of both chromosomes 19 and four had a signal on one chromosome 19 in the identical position. As shown in Figure 1B, the idiogram demonstrates that the C5a receptor also localizes to the 19q13.3–13.4 interface. Thus, by fluorescence in situ hybridization, the genes for the formyl peptide receptor family and the C5a receptor are clustered at a single locus on chromosome 19 at the interface of band 13.3 and 13.4.

Genes for the Formyl Peptide Receptor Homologues Are Transcribed. While the C5a and formyl peptide receptors are well-characterized, the formyl peptide receptor structural homologues, FPRH1 and FPRH2 (Lu et al., 1992), are presently "orphan receptors" whose ligands have not yet been identified. In order to determine whether these genes are expressed as cDNAs, we amplified several mRNA populations using reverse transcriptase–polymerase chain reaction. The

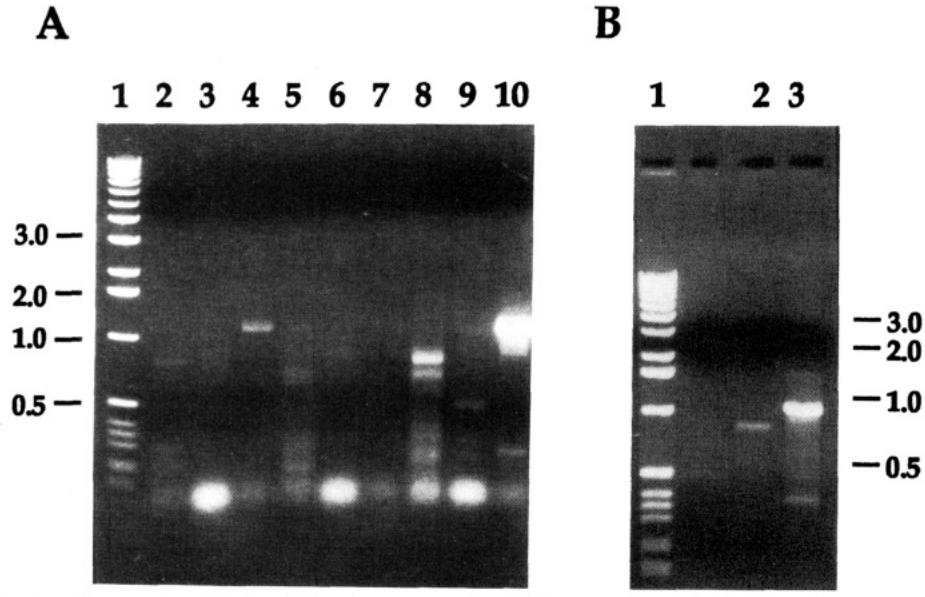

FIGURE 2: (A) Distribution of expression of FPR and homologues 1 and 2. PCR was performed using cDNA from cAMP-differentiated U937 cells (lanes 2–4), undifferentiated U937 (lanes 5–7), or a mixture of human peripheral blood eosinophils and neutrophils (lanes 8–10). The primers used were specific for FPRH1 (798 bp), lanes 2, 5, and 8, FPRH2 (848 bp), lanes 3, 6, and 9, or FPR primers (1053 bp), lanes 4, 7, and 10. Lane 1 represents DNA standards with sizes in kilobases and positions as indicated. (B) PCR using primers for FPRH2 (lane 2, 848 bp) and FPR (lane 3, 1053 bp) with human lung cDNA. Lane 1 shows DNA standards. Details are described in Experimental Procedures.

primer sequences used were designed to provide specificity in discriminating the FPRH1 and FPRH2 from the FPR. Using the genomic clones as control templates, no amplification was detected using the FPR template and the primers corresponding to FPRH1 or FPRH2. The FPR primers amplified both of the orphan homologues. As shown in Figure 2A, primers for both FPR and FPRH1 amplified cDNAs in dibutyryl-cAMP-differentiated U937 lymphoma cells and in human eosinophils, yielding the anticipated 1053- and 798-bp products, respectively. Undifferentiated U937 cells did not yield product with either set of primers, consistent with previous reports of formyl peptide receptor upregulation by differentiation of these cells (Harris & Ralph, 1985). The primers specific for FPRH2 did not amplify a cDNA in any of these three message populations. Since lung tissue contains a wide variety of cell types, including virtually every member of the lymphoid and myeloid series, as well as epithelial, neural, and mesenchymal cells, cDNA generated from lung tissue was amplified with both the FPR and FPRH2 primers to give the expected products of 1053 and 848 bp, respectively (Figure 2B), demonstrating transcription of both the formyl peptide receptor and the second orphan homologue, FPRH2 from this cDNA source. The identity of the PCR products was confirmed by Southern hybridization with the FPR cDNA probe. Thus, both FPR homologues are expressed as messenger RNAs, with the FPRH1 analogue apparently present in several cell types which express FPR, and FPRH2 species notably absent from several myeloid cell types. The C5a receptor is observed in every cell type and under the same conditions as the formyl peptide receptor and FPRH1 (Gerard & Gerard, 1991).

Characterization of the Chemotactic Receptor Genes. Southern analysis of genomic DNA digested with multiple enzymes reveals two or three hybridizing species when FPR is used as a probe (Boulay et al., 1990a). We previously demonstrated that the three fragments observed following *Eco*RI digestion correspond to FPR and two homologous genes, FPRH1 and FPRH2 (Lu et al., 1992), as opposed to the alternate possibility that introns interrupt the cDNA sequence. This finding was confirmed by PCR analysis of the four

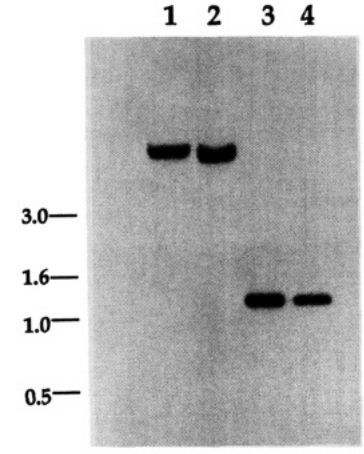

FIGURE 3: Human genomic DNA digested with *Eco*RI (lane 1), *Bam*HI (lane 2), or *Pst*I (lanes 3 and 4), Southern blotted, and hybridized with the 700-bp *Pst* fragment of the C5aR cDNA encompassing nucleotides 329–1054. The DNA in lane 4 was from a second individual. Positions of DNA standards are indicated at the left.

genomic clones constituting the members of the 19q13.3–q13.4 chemotactic receptor gene cluster including the C5a receptor, the formyl peptide receptor, and the homologues, FPRH1 and FPRH2. Amplification using oligonucleotide primers corresponding to the 5′ and 3′ ends of the coding sequences and the isolated genes indicates that each has the receptor-encoding portion in a single exon (not shown).

Southern analysis of human genomic DNA digested with *Eco*RI, *Bam*HI, or *Pst*I and probed with the C5a receptor cDNA encompassing nucleotides 329–1054 is presented in Figure 3. In contrast to the pattern observed with the FPR probe, the C5aR probe shows a simple restriction pattern consistent with a single-copy gene. A highly homologous gene with identical restriction patterns, however, would behave similarly due to comigration of the DNA fragments. In the case of restriction digests with *Pst*I, the major band at ~1.5 kb is accompanied by a minor band at ~2.5 kb which was observed in the DNA from two unrelated individuals. This

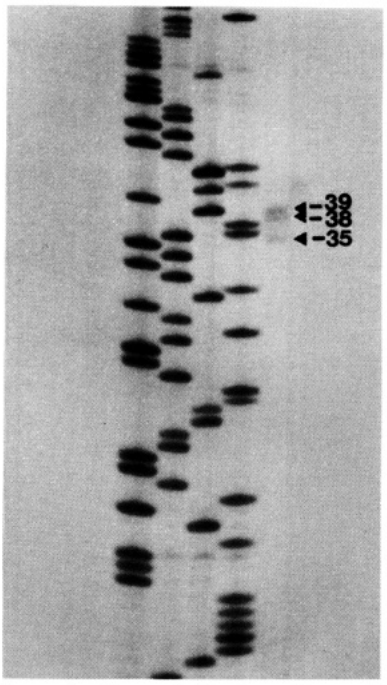

FIGURE 4: Primer extension analysis. An antisense oligonucleotide corresponding to nucleotides 40–70 of the C5a receptor coding sequence was labeled with ^{32}P using polynucleotide kinase. The labeled primer was hybridized with ~2 μg of dibutyryl-cAMP-differentiated U937 cell poly(A+) RNA and extended with AMV reverse transcriptase as described in Experimental Procedures. The extended fragments of 75, 78, and 79 bp relative to a calibrated sequence ladder indicate that initiation of transcription occurs at –39 relative to the translational initiation site.

weakly hybridizing fragment likely represents another gene with limited homology to this region of the C5a receptor.

Genomic clones for the C5a receptor isolated by hybridization with the *Pst*I restriction fragment of the cDNA encompassing nucleotides 1–329 were further characterized. Digestion of the genomic clones with *Pst*I generated two fragments hybridizing with the coding sequence cDNA which were subcloned to pBluescript. The genomic DNA fragment hybridizing with the 329-bp 5′ portion of the C5a receptor cDNA was sequenced on both strands and revealed an intron interrupting the C5aR coding sequence between the initiating methionine and the asparagine codons, between nucleotides 3 and 4. PCR primers based on the coding sequence 1–20 (sense) and 1043–1063 (antisense) efficiently amplify expressable C5a receptor using genomic DNA or the isolated genomic clone (Gerard & Gerard 1991; Lu et al., 1992). Therefore, while the 5′ untranslated region and initiating methionine codon must reside in a different DNA fragment from the remainder of the receptor, the coding sequence is devoid of introns. An oligonucleotide probe corresponding to –1 to –18 in the 5′ untranslated region identified a strongly hybridizing *Pst*I fragment of the genomic clone of ~1 kb. Sequence analysis of this DNA revealed that it contains the 22-bp 5′ untranslated region isolated from multiple cDNA libraries. Primer extension analysis using poly(A+) RNA isolated from dibutyryl-cAMP-differentiated U937 cells an an antisense oligonucleotide primer corresponding to nucleotides 40–70 in the C5a receptor coding sequence suggests a transcriptional initiation site at –39 bp relative to the translational initiation site (Figure 4), some 17 bp longer than the longest cDNA clones isolated. Extension products are also observed at –35 and –38 bp and are interpreted as pause sites for the reverse transcriptase, possibly due to neighboring G- and C-containing sequences. A restriction map of the C5a receptor gene is presented in Figure 5. As seen from this map, the 5′ untranslated region, ~39 bp in length, and the codon for the initiating methionyl residue are located approximately 9 kb away from exon 2 which contains the coding sequence for the receptor.

The sequences of the flanking regions for exons 1 and 2 of the C5a receptor gene are presented in Figure 6. The 39-bp 5′ untranslated region is located ~20 nucleotides downstream from a potential modified TATAA box. Computer analysis of this sequence indicates a potential CCAAT motif in the intron just 3′ to the end of exon 1. Also recognized in this region are candidate AP-1 and SP-1 sites and a second CCAAT motif at –280 nucleotides. The splice junction sequence at the initiating ATG residue is consistent with the most frequently observed consensus sequence. At the downstream site, the flanking region before exon 2 shows a classic polypyrimidine tract upstream from a consensus splice acceptor site.

Expression of Promoter Activity by the C5aR 5′ Flanking Region. The ~800-bp *Pst*I–*Xba*I fragment containing the 42-bp exon 1 (see Figure 5) was examined for promoter activity in several cell lines following subcloning into the reporter construct pBLCAT3. Transfection into U937 cells using either electroporation or cationic liposomes was irreproducible with control CAT constructs because of extremely low efficiency of transfection (data not shown). The basophilic leukemia cell line RBL-1, when transfected with this C5aR gene fragment, displayed promoter activity, increasing the expression of CAT by ~100% above baseline (Figure 7). The low transfection efficiency of these cells made the use of β-galactosidase as an internal control impractical; however, numerous experiments using several test constructs yielded similar activity. Because candidate AP1 sites were recognized in the sequence of this fragment (Figure 6), cells transfected with this plasmid were also stimulated with phorbol myristate acetate. This treatment resulted in a 2-fold increase in expression of CAT activity relative to untreated cells transfected with the same plasmid construct in RBL-1 cells.

Transfection of the same plasmids into human neuroblastoma SK-N-SH cells, which do not express endogenous C5a receptor, revealed no intrinsic promoter activity for the 5′ flanking fragment (Figure 7). A similar result was observed with human embryonic kidney TSA cells (not shown). Transfection efficiencies were internally controlled by cotransfecting with pRSV β-Gal in some experiments, and the β-galactosidase activity was found not to vary significantly among the test constructs used (data not shown). In TSA and SK-N-SH cells, PMA had no significant effect on CAT expression with the 800-bp C5aR gene fragment. The results of transfecting constructs with sequential 5′ deletions of the C5a receptor flanking sequence in SK-N-SH cells are also shown in Figure 7. Removal of 121 bp, including the CAAT motif at –280, reveals strong promoter activity in this portion of DNA, producing CAT activity 20-fold greater than either the promoterless plasmid or the plasmid containing the CAAT motif. Additionally, this promoter activity was almost 3-fold greater than that generated by plasmid containing the thymidine kinase promoter. Deletion of sequences to –82 bp relative to the translational initiation site did not diminish the C5a receptor promoter activity. However, when a plasmid which was deleted to –49 bp, removing the modified TATAA sequence, was transfected in these cells, CAT expression was reduced to baseline. Assays were carried out under conditions of linear enzyme activity, as indicated by the conversion of ≤50% of the substrate (Figure 8). These data are consistent

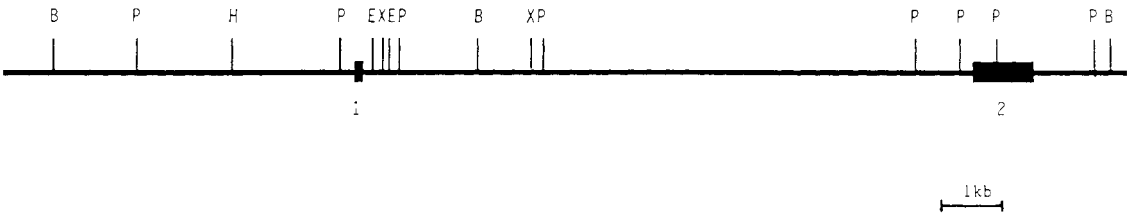

FIGURE 5: Restriction map of the human C5aR gene indicating the relative positions of exons 1 and 2 (heavy lines) and restriction sites for *Bam*HI (B), *Hin*dIII (H), *Eco*RI (E), *Xba*I (X), and *Pst*I (P). Additional *Pst*I sites exist but have not been mapped. Approximately 9 kb separate exons 1 and 2.

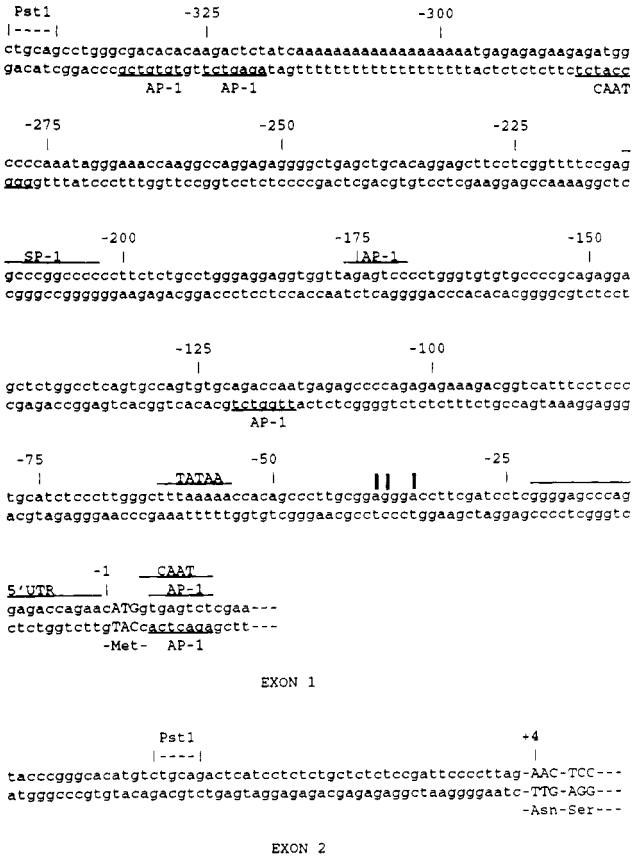

FIGURE 6: DNA sequences flanking exons 1 and 2 of the C5aR gene. Exon 1 consists of ~39 bp of 5′ untranslated sequence (5′ UTR) and the initiating methionine codon. Vertical lines indicate the putative transcriptional start sites determined by primer extension, while the longest isolated cDNA 5′ sequence is overlined. Exon 2 begins at nucleotide 4 and contains the remainder of the coding sequence. Putative regulatory elements are indicated by overlining on the positive strand and underlining on the negative strand. Numbering is relative to the translational initiation site.

with identification of the modified TATAA sequence as the functional C5a receptor promoter, and, by comparison with the data obtained using RBL-1 cells, suggest that cell-type specific expression of the C5a receptor may relate to suppressor sequences contained in the region from −346 to −225.

DISCUSSION

The G-protein coupled receptor gene family is rapidly expanding as more members are identified by PCR analysis and expression cloning. To date, over 100 unique receptors have been identified which can be broadly classified by sequence identities into subfamilies. The largest subfamily to date appears to encode olfactory receptors, followed by receptors encoding neurotransmitters. Other agonists such as lipids and peptides form distinct receptor subfamilies. In general, gene clustering has not been observed among related receptors. For example, we previously reported the mapping of genes for the highly conserved receptors for the neuropeptide tachykinins, substance P and neurokinin A, to chromosomes 2 and 10, respectively. Exceptions known to date are the β1 and α2 receptors on chromosome 10q24–26 and β2 and α1 receptors on 5q32–34 (Yang-Feng et al., 1990). One might anticipate clustering of the olfactory receptors by virtue of their high degree of sequence identity (Buck & Axel, 1991). The present data, indicating a gene cluster for the C5a and formyl peptide receptor group at 19q13.3–13.4, are of interest in part because the C5a and formyl peptide ligands are more divergent than the tachykinins, whose receptor genes are widely separated, or other adrenergic receptor genes which are also widely separated. Also, clustering per se does not follow function, because other receptors such as those for related catecholamines are separated. One reason for gene clustering might relate to coordinate regulation of gene transcription. Our data indicate that the FPR, FPRH1, and C5a receptors are always present in the same cells among those examined, and it is conceivable that these genes in the 19q13.3–13.4 cluster are regulated by a common element. While this paper was in preparation, two groups demonstrated that the IL-8 receptor forms A (Holmes et al., 1991) and B (Murphy & Tiffany, 1991) as well as a homologous pseudogene map to chromosome 8q34.35,[2,3] thus providing another example that clustering of chemotaxis receptor genes may serve some function. Fine mapping of these loci will focus on this possibility.

Molecular cloning of genomic DNA led to the identification of two homologues of the formyl peptide receptor (Lu et al., 1992). While neither homologue binds formyl peptide analogues at physiologic concentrations following transfection in mammalian cells, we anticipate that these orphan receptors transmit similar signals to the cells in which they are expressed by virtue of the strong sequence homologies in the regions predicted to interact with heterotrimeric G-proteins. The FPRH1 homologue is apparently expressed in each instance where the formyl peptide receptor is expressed (Figure 2), as is the C5a receptor. Other investigators have recently cloned this receptor analogue as an orphan receptor by homology with the FPR (Murphy et al., 1992).

The second FPR-related gene, FPRH2, was not found to be expressed coordinately with the C5a and formyl peptide receptors. The presence of this transcript in human lung cDNA indicates that the gene is potentially expressed as a protein in this organ. Candidate cell types present in lung which might express such a putative chemotactic receptor include lymphocytes, macrophages, mast cells, eosinophils, basophils, and neutrophils. Since we did not detect this receptor message in PMN neutrophils, eosinophils, or macrophage-like cell lines, it is possible that the FPRH2 molecule is expressed on lymphocytes or mast cells/basophils. Further investigation in this area is necessary, as is the identification

[2] D. Cerretti and M. P. Beckman, personal communication.
[3] P. Murphy, personal communication.

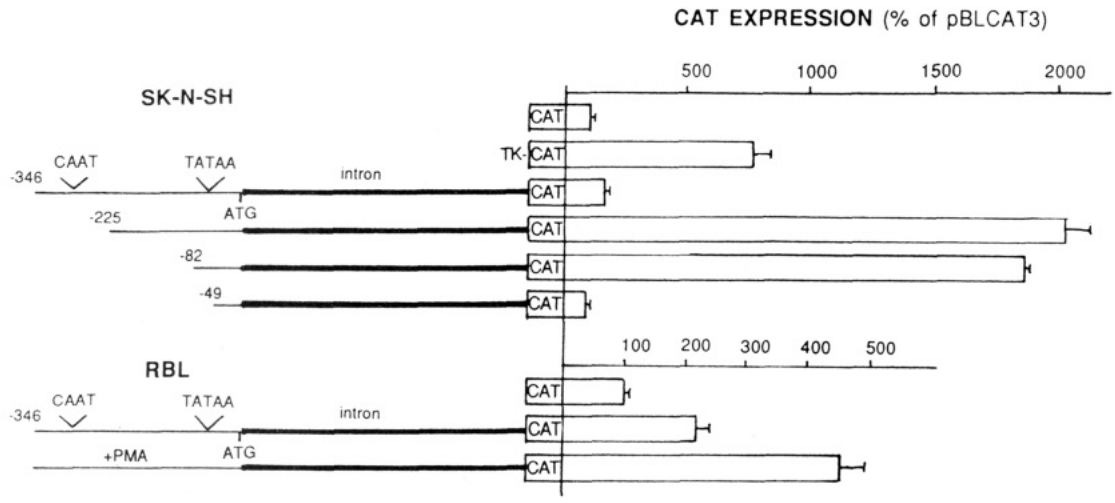

FIGURE 7: Relative CAT activity expressed in RBl-1 or SK-N-SH cells by C5aR promoter constructs. Control transfections in RBL-1 cells were performed with pBLCAT-3 to determine baseline activity or with the same vector containing the ~800-bp *Pst*I–*Xba*I fragment which includes exon 1. The heavy line indicates intronic sequence downstream from the initiating methionine. At 18–24 h before harvest, 1 μM PMA was added to some of the cells transfected with the C5aR promoter. Transfections in SK-N-SH cells were also performed with a control plasmid containing the thymidine kinase promoter (TK-CAT, pBLCAT2) as well as C5aR promoter deletion constructs as indicated at the left. CAT activity was determined as described in Experimental Procedures. The data shown are the mean of three separate experiments with transfections in duplicate for RBL-1 cells and two separate experiments with transfections in triplicate for SK-N-SH cells.

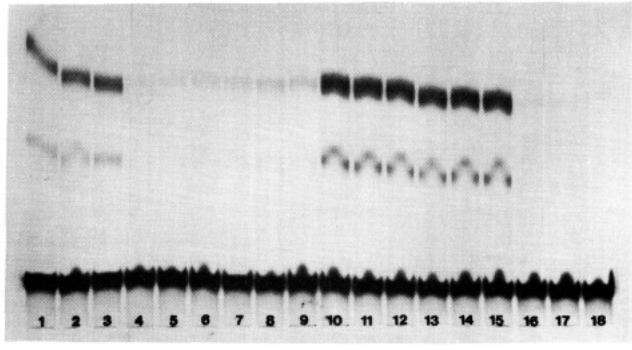

FIGURE 8: Autoradiogram of a SK-N-SH cell transfection indicating relative conversion of [^{14}C]chloramphenicol to its faster-migrating acetylated derivatives. Lanes 1–3, TK-CAT (pBLCAT2); lanes 4–6, promoterless pBLCAT3; lanes 7–9, the ~800-bp *Pst*I–*Xba*I C5aR gene fragment containing exon 1; lanes 10–18, C5aR gene fragments with 5′ deletions to –225 bp (lanes 10–12), to –28 bp (lanes 13–15), or to –49 bp (lanes 16–18).

of the natural ligand(s) for the formyl peptide receptor homologues. Interestingly, we have observed increased baseline phosphatidylinositol turnover in cells transfected with the formyl peptide receptor, suggesting that a factor present in cell culture media can activate the receptor (N. P. Gerard, unpublished observation).

Further characterization of the chemotactic receptor gene cluster on chromosome 19 will require detailed mapping of these genes and dissection of their regulatory elements. The present data provide initial characterization of the C5aR locus. We observe here the general tendency for members of G-protein coupled receptor family to have the protein coding sequence contained in a single exon. Interestingly, we find a large (~9-kb) intron separating the small 5′ untranslated sequence (39 bp) and initiating methionine codon from the receptor-containing exon. A similar case has recently been reported for the formyl peptide receptor gene where a single intron of ~5 kb separates the 66-bp 5′ untranslated sequence from the coding exon (Haviland et al., 1992).

Exon 1, containing the 39-bp 5′ untranslated sequence and initiating methionine codon, was isolated in an ~800-bp *Pst*I–*Xba*I fragment. This fragment contains 346-bp upstream from the initiating methionine and approximately 450 bp of intronic sequence downstream from exon 1. The promoter activity of the ~800-bp *Pst*I–*Xba*I fragment was examined with a CAT reporter gene in SK-N-SH human neuroblastoma cells and in RBL-1 cells. The rat basophilic leukemia cell line, which has additionally been recognized to have mast cell properties, is myeloid in origin while SK-N-SH is an epithelial cell.

In the SK-N-SH cell, which does not express C5a receptors, the ~800-bp *Pst*I–*Xba*I fragment displays no additional CAT activity relative to promoterless pBLCAT3. Deletion of the most 5′ 121-bp, yielding a construct with 225 bp upstream from the translational initiation site and ~450 bp of intron 1, generated strong promoter activity. Because the –225 construct contained both exon 1 and intronic sequence of ~450 bp 3′ to the initiating methionine codon, the DNA was progressively deleted around a putative modified TATAA motif at –62. The fragment deleted to –82 relative to the translational initiation site remained a strong promoter, while the fragment deleted to –49 returned to baseline promoterless activity (i.e., the same as pBLCAT3). Taken together, these data indicate that a transcriptional suppressor element between –225 and –346 masks a strong promoter activity present between –49 and –82. The –49 construct, containing exon 1 and ~450 bp of intronic sequence may contain additional suppressor or enhancer functions.

When the ~800-bp *Pst*I–*Xba*I fragment was examined with CAT reporter constructs in RBL-1 cells, a different result was obtained. In this myeloid-derived cell, weak promoter activity was observed instead of baseline (promoterless) activity. When the transfected RBL-1 cells were stimulated with PMA, modest (4–5-fold) promoter activity was observed. Thus, the intrinsic suppressor activity contained in –225 to –346 was not active in myeloid-origin cells compared with epithelial SK-N-SH cells. A similar result has been observed with the myeloid-specific gp91-Phox gene (Neufeld et al., 1992; Skalnick et al., 1991), a component of the NADPH oxidase. Myeloid-specific expression of gp91-Phox reporter constructs could be reversed by deletion of one of two CCAAT motifs located upstream from a relatively strong promoter. The mutation of this motif to CCGGT, as well as footprinting and gel-shift shift assays, established that repression of promoter activity involved the CCAAT displacement protein

(CDP). On this note, a modified CCAAT site is identified in the C5a receptor 5′ flanking region at −280, in the suppressor sequence region.

The present results clearly define model systems for analysis of the C5a receptor gene regulatory regions and potentially for the other genes located in the chromosome 19 chemotaxis receptor gene cluster.

ACKNOWLEDGMENT

We thank Olivia Orozco for expert technical assistance and Teisha Williams for preparation of the manuscript.

REFERENCES

Beaulieu, A. D., Paquin, R., Rathanaswami, P., & McColl, S. R. (1992) *J. Biol. Chem. 267*, 426–432.

Beckmann, M. P., Munger, W. E., Kozlosky, C., VandenBos, T., Price, V., Lyman, S., Gerard, N. P., Gerard, C., & Cerretti, D. P. (1991) *Biochem. Biophys. Res. Commun. 179*, 784–789.

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.

Boulay, F., Mery, L., Tardif, M., Brouchon, L., & Vignais, P. (1991) *Biochemistry 30*, 2993–2999.

Buck, L., & Axel, R. (1991) *Cell 65*, 175–187.

Cherif, D., Bernard, O., & Berger, R. (1989) *Hum. Genet. 81*, 358–362.

Connelly, J. C., Skidgel, R. A., Schulz, W. W., Johnson, A. R., & Erdos, E. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 8737–8741.

Dohlman, H. G., Thorner, J., Caron, M. G., & Lefkowitz, R. J. (1991) *Annu. Rev. Biochem. 60*, 653–688.

Fan, Y.-S., Davis, L. M., & Shows, T. B. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6223–6227.

Gerard, N. P., & Gerard, C. (1990) *Biochemistry 29*, 9274–9281.

Gerard, N. P., & Gerard, C. (1991) *Nature 349*, 614–617.

Gerard, N. P., Hodges, M. K., Drazen, J. M., Weller, P. F., & Gerard, C. (1989) *J. Biol. Chem. 264*, 1760–1766.25.

Gerard, N. P., Eddy, R. L., Jr., T. B., & Gerard, C. (1990) *J. Biol. Chem. 265*, 20455–20462.

Gerard, N. P., Garraway, L. A., Eddy, R. L., Jr., Shows, T. B., Iijima, H., Paquet, J. L., & Gerard, C. (1991) *Biochemistry 30*, 10640–10646.

Gilman, A. G. (1987) *Annu. Rev. Biochem. 56*, 615–649.

Goldman, D. W., & Goetzel, E. J. (1984) *J. Allergy Clin. Immunol. 74*, 373–377.

Goldman, D. W., Olson, D. M., Payan, D. G., Gifford, L. A., & Goetzel, E. J. (1986) *J. Immunol. 136*, 4631–4636.

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

Harvath, L. (1991) *Experientia Suppl. 59*, 35–52.

Haviland, D. L., Borel, A. C., Fleischer, J. C., Haviland, J. C., & Wetsel, R. A. (1992) *FASEB J. 6*, A621.

Holmes, W. E., Lee, J., Kuang, W. J., Rice, G. C., & Wood, W. I. (1991) *Science 253*, 1278–1280.

Hwang, S. B., Lam, M. H., & Chang, M. N. (1986) *J. Biol. Chem. 261*, 13720–13726.

Kunz, D., Gerard, N. P., & Gerard, C. (1992) *J. Biol. Chem. 267*, 9101–9106.

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

Luckow, B., & Shutz, G. (1987) *Nucleic Acids Res. 15*, 5490.

McLeish, K. R., Gierschik, P., Schepers, T., Sidiropoulos, D., & Jakobs, K. H. (1989) *Biochem. J. 260*, 427–434.

Murphy, P. M., & McDermott, D. (1991) *J. Biol. Chem. 266*, 12560–12567.

Murphy, P. M., & Tiffany, H. L. (1991) *Science 253*, 1280–1283.

Murphy, P. M., Ozcelik, T., Kenney, R. T., Tiffany, H. L., McDermott, D., & Francke, U. (1992) *J. Biol. Chem. 267*, 7637–43.

Nakamura, M., Honda, Z., Izumi, T., Sakanaka, C., Mutoh, H., Minami, M., Bito, H., Seyama, Y., Matsumoto, T., Noma, M., & Shimizu, T. (1991) *J. Biol. Chem. 266*, 20400–20405.

Nelson, R. D., Gracyk, J. M., Fiegel, V. D., Herron, M. J., & Chenoweth, D. E. (1981) *Blood 58*, 752–758.

Nelson, R. D., Chenoweth, D. E., Solomkin, J. S., & Solem, L. D. (1983) *Agents Actions Suppl. 12*, 274–289.

Neufeld, E. J., Skalnick, D. G., Lievens, P. M.-J., & Orkin, S. H. (1992) *Nature Genet. 1*, 50–56.

O'Flaherty, J. T., Surles, J. R., Redman, J., Jacobson, D., Piantadosi, C., & Wykle, R. L. (1986) *J. Clin. Invest. 78*, 381–388.

O'Flaherty, J. T., Redman, J. F., & Jacobson, D. P. (1990) *J. Immunol. 144*, 1909–1913.

O'Flaherty, J. T., Rossi, A. G., Redman, J. F., & Jacobson, D. P. (1991) *J. Immunol. 147*, 3842–3847.

Oppenheim, J. J., Zachariae, C. O., Mukaida, N., & Matsushima, K. (1991) *Annu. Rev. Immunol. 9*, 617–648.

Rose, J. K., Buonocore, L., & Whitt, M. A. (1991) *Biotechniques 10*, 520–525.

Rosenthal, N. (1987) *Methods Enzymol.* 152, 684.

Schepers, T. M., Brier, M. E., & McLeish, K. R. (1992) *J. Biol. Chem. 267*, 159–165.

Skalnik, D. G., Strauss, E. C., & Orkin, S. H. (1991) *J. Biol. Chem. 266*, 16736–16744.

Soberman, R. J., Sutyak, J. P., Okita, R. T., Wendelborn, D. F., Roberts, L. J., II, & Austen, K. F. (1988) *J. Biol. Chem. 263*, 7996–8002.

Tarbet, E. B., Stafforini, D. M., Elstad, M. R., Zimmerman, G. A., McIntyre, T. M., & Prescott, S. M. (1991) *J. Biol. Chem. 266*, 16667–16673.

Ward, P. A., & Ozols, J. (1976) *J. Clin. Invest. 58*, 123–129.

Yang-Feng, T. L., Xue, F. Y., Zhong, W. W., Cotecchia, S., Frielle, T., Caron, M. G., Lefkowitz, R. J., & Francke, U. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 1516–1520.

Ye, R. D., Prossnitz, E. R., Zou, A. H., & Cochrane, C. G. (1991) *Biochem. Biophys. Res. Commun. 180*, 105–111.