Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 6250–6254, September 1985
Immunology

Cloning, sequence, and expression of a human granulocyte/macrophage colony-stimulating factor

(bone marrow colony assay/cDNA cloning/yeast expression)

MICHAEL A. CANTRELL*, DIRK ANDERSON*, DOUGLAS PAT CERRETTI*, VIRGINIA PRICE*, KATE MCKEREGHAN*, ROBERT J. TUSHINSKI†, DIANE Y. MOCHIZUKI†, ALF LARSEN*, KENNETH GRABSTEIN‡, STEVEN GILLIS, AND DAVID COSMAN*

Departments of *Molecular Biology, †Cellular Biochemistry, and ‡Cellular Immunology, Immunex Corporation, 51 University Street, Seattle, WA 98101

Communicated by Lewis H. Sarett, May 15, 1985

ABSTRACT **Human granulocyte/macrophage colony-stimulating factor (GM-CSF) is a glycoprotein that is essential for the *in vitro* proliferation and differentiation of precursor cells into mature granulocytes and macrophages. In this report we have used a mouse GM-CSF cDNA clone to isolate human GM-CSF clones from libraries made from HUT-102 messenger RNA and mitogen-stimulated T-lymphocyte messenger RNA. The human cDNA clones contained a single open-reading frame encoding a protein of 144 amino acids with a predicted molecular mass of 16,293 daltons and showed 69% nucleotide homology and 54% amino acid homology to mouse GM-CSF. One of these cDNA clones was shown to direct the synthesis of biologically active GM-CSF using a yeast expression system. The gene for human GM-CSF appears to exist as a single-copy gene.**

The growth and differentiation of hematopoietic cells is mediated by a number of glycoproteins, collectively known as colony-stimulating factors (CSFs), including granulocyte/macrophage colony-stimulating factor (GM-CSF), granulocyte-CSF (G-CSF), macrophage-CSF (M-CSF), and multi-CSF (interleukin 3) (1, 2). There has been considerable difficulty in defining these regulatory proteins, partly because they have been obtained from a variety of sources and characterized by their capacity to generate hematopoietic cells of various lineages *in vitro*. An understanding of the biology and biochemistry of these CSFs is further complicated by their existence in small quantities. Molecular cloning and expression of these factors should lead to a better definition of their structure and biological activities and exploration of their possible therapeutic uses.

GM-CSF is a glycoprotein that is required for the production of granulocytes and macrophages from normal bone marrow and appears to regulate the activity of mature, differentiated granulocytes and macrophages (1, 3, 4). Human GM-CSF, which has been isolated recently from the Mo T-lymphoblast cell line, has been shown to modulate the activities of mature neutrophilic granulocytes and appears to be identical to a neutrophil migration-inhibition factor from T lymphocytes (NIF-T) (5). A murine GM-CSF cDNA clone has been obtained following the determination of a partial amino-terminal sequence for the protein purified from mouse lung-conditioned medium (6, 7).

We report here the molecular cloning of human GM-CSF from cDNA libraries prepared from the HUT-102 cell line and from mitogen-stimulated T lymphocytes. The human sequences were isolated by hybridization with a mouse GM-CSF cDNA probe and were shown to direct the synthesis of biologically active GM-CSF using a yeast expression system.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Construction and Analysis of Libraries. Polyadenylylated messenger RNA was isolated from human peripheral blood T lymphocytes stimulated with concanavalin A (Con A) and phorbol myristate acetate (PMA) at 20 μg/ml and 10 ng/ml, respectively, and from the mouse T-lymphoma cell line LBRM-33-5A4 (ATCC-CRL-8080) stimulated with 1% phytohemagglutinin M (PHA-M). Procedures for RNA purification and cDNA library construction have been described (8).

Small-scale plasmid DNA preparations from pools representing approximately 1×10^5 transformants were digested with *Pst* I, electrophoresed on 0.8% agarose gels, blotted onto nitrocellulose filters, and hybridized either with cDNA probes (^{32}P-labeled by nick-translation) or with oligonucleotide probes (labeled with T4 polynucleotide kinase). Hybridizations were for 16 hr at 55°C in 6× NaCl/Cit (1× NaCl/Cit = 0.15 M NaCl/0.015 M sodium citrate, pH 7) containing 0.1% sarcosyl, 5× Denhardt's solution (1× = 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin), 0.5% Nonidet P-40, 100 μg of denatured salmon sperm DNA per ml, and probe at 10^6 cpm/ml. Filters were washed extensively in 6× NaCl/Cit at room temperature, and then washed for 1 hr at 42°C and for 1.5 hr at 55°C before autoradiography. Positive pools containing the largest hybridizing cDNAs were subdivided, and the process was repeated until pools of 2,000–10,000 transformants were obtained. Then these pools were used in colony filter hybridization experiments (9) to identify transformants that hybridized strongly with the probes.

DNA from the cDNA clones was subcloned into M13 mp18 and mp19 (10) and sequenced by the chain-termination method (11) as described (12, 13).

Expression of Human GM-CSF in Yeast. An expression vector (Fig. 1) was constructed which included pBR322 sequences, the *TRP1* gene of yeast for tryptophan selection, the yeast 2-μm origin of replication, and the yeast α-factor promoter and leader sequences sufficient to direct the synthesis and secretion of foreign proteins (14). The *Eco*RI to *Hin*dIII α-factor fragment was obtained by cloning the α-factor gene using as a probe a synthetic oligonucleotide derived from published sequences (15). The *Sfa*NI–*Nco* I fragment of pHG23 (see Fig. 2*A*) was fused in-frame to the α-factor signal sequence by use of a synthetic oligonucleotide (Fig. 1). This oligonucleotide linker joins the *Hin*dIII site of α factor to the *Sfa*NI site of GM-CSF, adding a second α-factor processing site to obtain complete processing of the signal (14) for secretion of the mature form of GM-CSF.

Abbreviations: CSF, colony-stimulating factor; GM-CSF, granulocyte/macrophage CSF; Con A, concanavalin A; PMA, phorbol myristate acetate; bp, base pair(s).

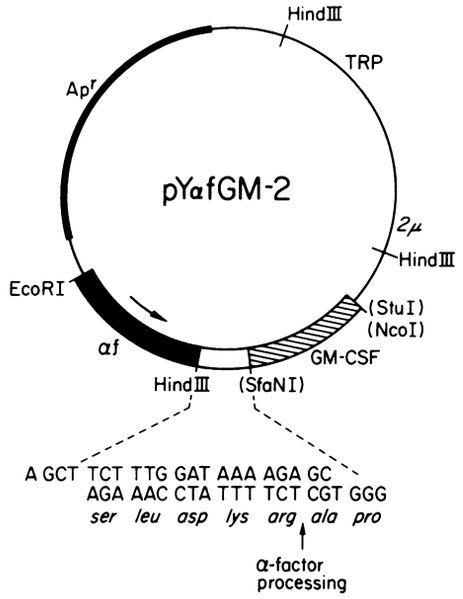

FIG. 1. Structure of the yeast expression plasmid. The plasmid, pYαfGM-2, contains sequences derived from pBR322 (thick line) containing the origin of replication and ampicillin resistance gene (Ap^r) and sequences from yeast (thin line) including the *TRP-1* gene as a selectable marker and the 2-μm origin of replication. The solid box indicates the α-factor (αf) promoter and leader sequences used to direct transcription and secretion of GM-CSF. The synthetic oligonucleotides used to fuse the α-factor sequences with GM-CSF sequences (hatched box) are shown.

The expression plasmid was transformed into yeast strain 79 [α, *trp1-1*, *leu2-1*] selecting for Trp^+ transformants. Cultures to be assayed for biological activity were grown in 20–50 ml of rich media (1% yeast extract/2% peptone/2% glucose) at 30°C to stationary phase. The protease inhibitors phenylmethylsulfonyl fluoride and pepstatin A were added at the time of harvest to a final concentration of 1 mM and 10 μM, respectively. Cells were then removed by centrifugation, and the medium was filtered through a 0.45-μm cellulose acetate filter.

Preparation of Human Bone Marrow Cells. Human bone marrow from the iliac crest of healthy donors was collected in a heparinized syringe. The marrow was diluted 1:3 with phosphate-buffered saline at room temperature and layered onto a solution of 54% Percoll (Pharmacia). After centrifugation at 500 × *g* at room temperature for 20 min, the interface was collected and washed with 20 vol of phosphate-buffered saline. The suspension was then centrifuged at 250 × *g* for 10 min. Cells were resuspended in the desired volume of alpha minimal essential medium (MEM α medium) with nucleotides (GIBCO) for cell counting and viability determination. Serum was then added, and the cell suspension was stored on ice until assay.

Colony Assay. The presence of human GM-CSF was determined by its ability to stimulate the growth of colonies in agar. Bone marrow cells were added at a final concentration of 1×10^5 per ml to incubation medium consisting of 3 parts of a 1.4% bacto-agar solution (Difco) and 7 parts of a solution containing 28.1% fetal calf serum, 70 μM mercaptoethanol, 0.12 mg of asparagine and 0.7 mg of glutamine per ml, 150 units of penicillin G, 150 units of streptomycin, 1.1× MEM α medium with nucleotides and 2.2× vitamins (GIBCO). The cultures were incubated in humidified 5% CO_2/95% air at 37°C and examined after 7 and 14 days to determine the number and type of colonies (granulocyte, macrophage, and mixed granulocyte/macrophage).

RNA Analysis. Primer extension experiments were performed as described (8) except that electrophoresis was in a 6% polyacrylamide/8 M urea gel.

Total RNA for Northern blots was isolated by the guanidinium thiocyanate/cesium chloride method (16). Samples were then electrophoresed in 1.1% agarose gels containing formaldehyde, transferred to nitrocellulose filters (16), and hybridized with a ^{32}P-labeled RNA probe transcribed by SP6 polymerase (17). The ^{32}P-RNA probe was synthesized from the 600-base-pair (bp) *Pst* I to *Nco* I fragment of pHG23, which was inserted into pSP64 (Promega Biotec, Madison, WI). Hybridization and washing of blots was as described (8).

RESULTS

Cloning and Sequencing of Mouse and Human GM-CSF DNA. A cDNA library was constructed with polyadenylylated messenger RNA from the mouse T-cell lymphoma line LBRM-33-5A4 stimulated with phytohemagglutinin. An oligonucleotide complementary to the 40 nucleotides at the 5′ end of the published sequence for mouse GM-CSF (6) was synthesized, labeled with ^{32}P, and used as a hybridization probe to screen the library. One hybridizing clone, TH1, was isolated and sequenced (Fig. 2*B*). This sequence contained a single, long, open-reading frame beginning 27 nucleotides 5′ of the previously published sequence for mouse GM-CSF and largely identical to it except for a few nucleotide changes. The two sequences will be compared elsewhere (unpublished data). The cDNA of TH1 has been shown to direct production of biologically active GM-CSF using a yeast expression system (unpublished results).

A 372-bp *Hae* III fragment from TH1, extending from nucleotide positions 42 to 414 (Fig. 2*B*) was isolated, nick-translated, and used under conditions of low stringency to probe two human cDNA libraries that were constructed from mRNA isolated from either the human cell line HUT-102 or from mitogen-stimulated human peripheral blood T lymphocytes. A number of cDNA clones that contained sequences complementary to the mouse GM-CSF probe were isolated from these libraries and shown by cross-hybridization and restriction mapping to be related. The cDNA from two of these HUT-102 clones (pHG23 and pHG25) is shown in Fig. 2*A*. Sequence analysis of these clones showed an open-reading frame encoding a protein of 144 amino acids with a predicted molecular mass of 16,293 daltons. The cDNA includes the predicted amino acid sequence Ala-Pro-Ala-Arg-Ser-Pro-Ser-Pro-Ser, which is identical to the reported amino terminus for mature human GM-CSF protein (5). The 5′ end of the open-reading frame encodes a region of 17 amino acids, with a potential initiator methionine and many of the characteristics expected of a hydrophobic signal peptide for a secreted protein (18).

Synthesis of Recombinant GM-CSF in Yeast. To show that the cDNA in pHG23 encoded a protein with GM-CSF activity, we constructed a plasmid designed to direct synthesis and secretion of the mature form of GM-CSF from yeast (see Fig. 1). Secretion of the mature protein is expected to result in glycosylation of the protein. The construction contains the yeast α-factor promoter plus leader sequences sufficient to direct the secretion of GM-CSF, followed immediately by the sequence for mature GM-CSF beginning with the alanine at amino acid position 1 in Fig. 2*B*. Supernatants from yeast transformed with the GM-CSF expression plasmid or from the control plasmid, pYαf, lacking the GM-CSF sequences of Fig. 1, or from human placental cells cultured at 1.2×10^7 per ml for 6 days in the presence of 5% fetal bovine serum were assayed for the ability to direct the formation of mixed granulocytic and macrophage-type colonies from human bone marrow cells as described. pYαfGM-2 directed synthesis of high levels of GM-CSF activity, where-

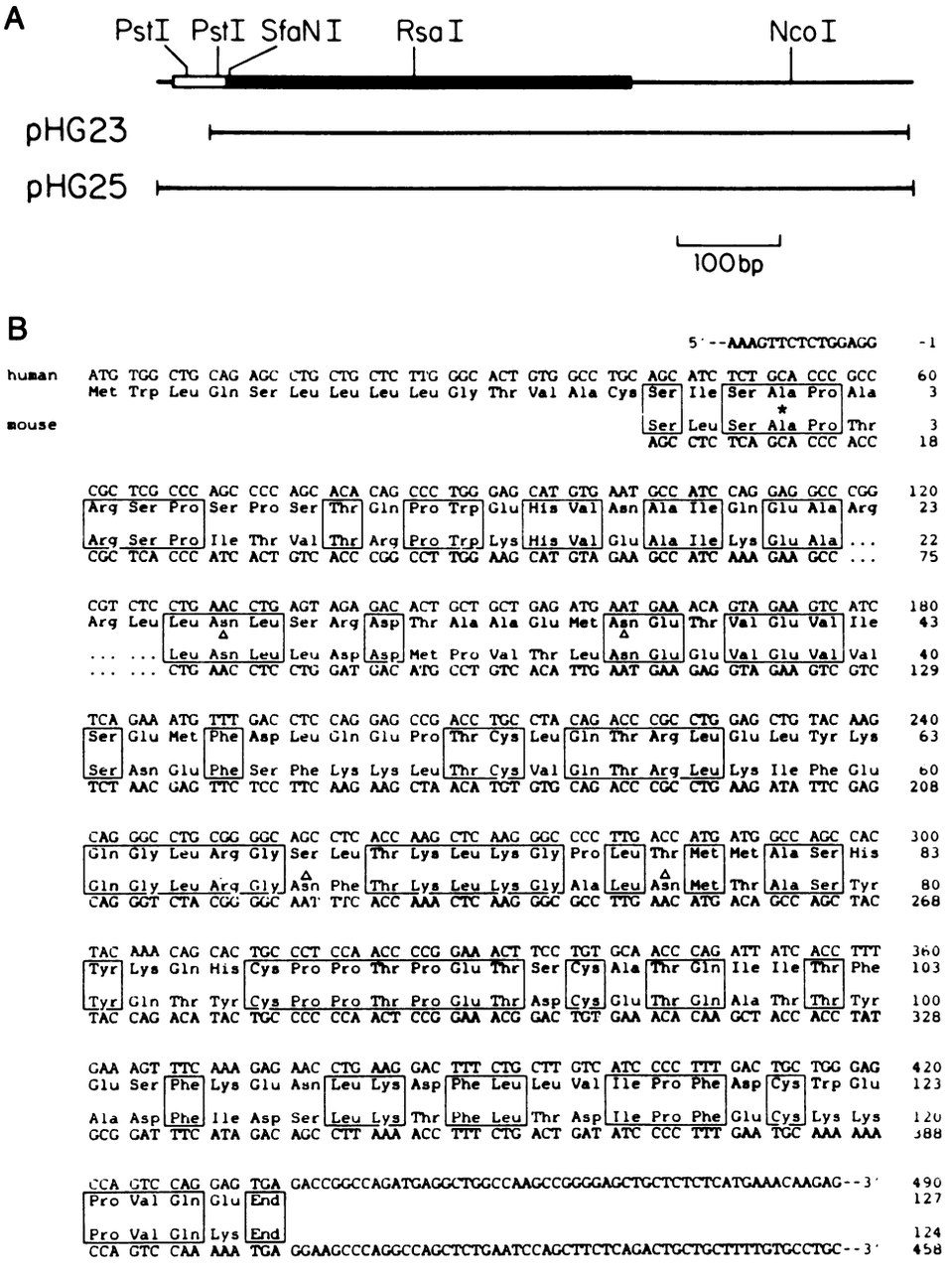

FIG. 2. Restriction map and nucleotide sequence of GM-CSF cDNA clones. (*A*) Partial restriction map of human GM-CSF cDNA. Coding sequences are boxed: the open box represents the proposed signal sequence and the shaded box represents the coding region for mature protein. The sizes of the cDNA inserts in pHG23 and pHG25 are indicated. (*B*) The nucleotide sequences and the predicted amino acid sequences for mouse and human GM-CSF. The cDNA sequences of the complete 3′ untranslated regions are not shown. The nucleotides are numbered from the presumed initiator methionine codon in the human GM-CSF and from the first nucleotide in the mouse GM-CSF. The amino acids are numbered from the amino terminus of the mature proteins, the alanine residues marked with a star. Boxed amino acid residues indicate regions of homology between the mouse and human proteins. Triangles represent possible N-linked glycosylation sites.

as the control plasmid, pYαf, produced supernatants with no activity. The titer [granulocytic progenitor cells (CFU-C) per ml] was calculated by multiplying by 50 the reciprocal of the dilution giving 50% of the maximum colony number: pYαfGm-2, 1.25×10^6 CFU-C per ml; pYαf, 0 CFU-C per ml; placental cells, 5×10^2 CFU-C per ml. (In seven experiments with different donors, the average maximum number of colonies from 10^5 marrow cells was 96 ± 29.) At 14 days, the recombinant GM-CSF and placental cell supernatants each produced well-defined colonies, of which approximately one-third was granulocyte/macrophage colonies, one-third was dispersed macrophage colonies, and one-third was tight granulocyte colonies. We have observed the same types of colonies with recombinant murine GM-CSF in murine bone marrow colony assays (unpublished data).

Analysis of mRNA. Since the availability of cells that synthesize GM-CSF is limited, we investigated the expression of GM-CSF mRNA from a number of cell types that have been reported previously to express other CSFs. Blots of RNA from these cells were analyzed by hybridization with a probe derived from pHG23. Fig. 3 *Left* shows strong hybridization to a band of approximately 900 nucleotides in RNA from peripheral blood T cells activated with PMA and Con A and from HUT-102 cells (lanes 3 and 1). A low level of hybridization was seen in RNA from a human bladder tumor cell line (lane 6) and no hybridization was seen in RNA from

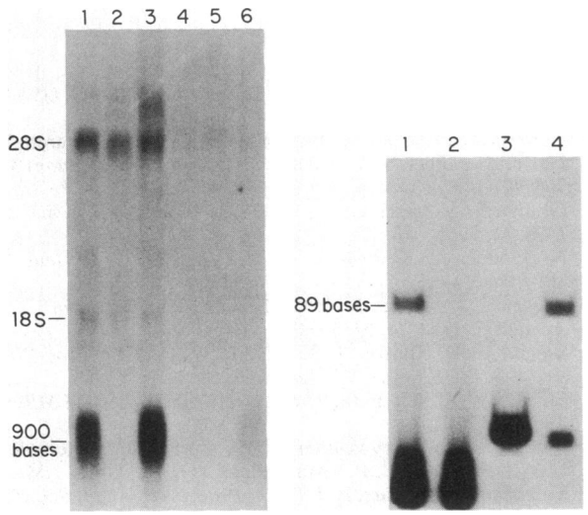

FIG. 3. Analysis of GM-CSF messenger RNA. (*Left*) Hybridization of GM-CSF probe to blots of RNA from human cells. Lanes: 1, 5 μg of total RNA from HUT-102 cells; 2, 5μg of total RNA from unstimulated peripheral blood T cells; 3, 5 μg of total RNA from peripheral blood T cells stimulated with Con A and PMA; 4, 1.5 μg of polyadenylylated RNA from peripheral blood macrophages stimulated with lipopolysaccharide; 5, 1.5 μg of polyadenylylated RNA from the pancreatic carcinoma cell line 1420; 6, 1.5 μg of polyadenylylated RNA from the bladder carcinoma cell line 5637. The positions of 18S and 28S rRNA bands are indicated. (*Right*) Primer extension analysis of GM-CSF RNA using a 5′ end-labeled probe complementary to the mRNA from nucleotide numbers 38–57 in Fig. 2*B*. Lanes: 1, analysis of 30 μg of total RNA from peripheral blood T cells stimulated with Con A and PMA; 2, analysis of 30 μg of total RNA from the Jurkat T-cell lymphoma cell line; 3, a 47-base oligonucleotide that was 5′ end-labeled with ^{32}P as a size marker; 4, ^{32}P-labeled size markers of 88 and 42 bp derived from an *Ava* II digest of pBR322.

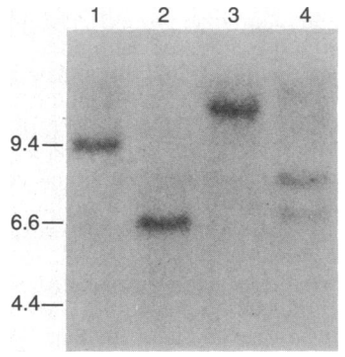

FIG. 4. Hybridization of GM-CSF cDNA to Southern blots of human genomic DNA. Genomic DNA (10 μg) was digested with *Hin*dIII (lane 1), *Eco*RI (lane 2), *Pst* I (lane 3), or *Bgl* II (lane 4), electrophoresed in a 0.7% agarose gel, blotted, and hybridized at high stringency to nick-translated ^{32}P-labeled GM-CSF cDNA by standard techniques. The molecular weight markers (in kilobase pairs) are from *Hin*dIII-digested bacteriophage λ DNA.

unstimulated T cells, lipopolysaccharide-stimulated macrophages, and a human pancreatic tumor cell line (lanes 2, 4, and 5). Other, higher molecular weight bands in Fig. 3 *Left* are due to hybridization of the probe to ribosomal RNA.

To determine the position of the 5′ end of the GM-CSF mRNA, primer extension experiments were performed by using an oligonucleotide complementary to the mRNA from nucleotide numbers 38–57 in Fig. 2*B*. A single primer extension product of 89 nucleotides was seen with RNA from mitogen-stimulated T cells (Fig. 3 *Right*), indicating that the cDNA of pHG25 is approximately 17 bp short of the 5′ end of the natural mRNA. Therefore, this mRNA must have an extremely short 5′ untranslated region. No primer extension product was seen with RNA from the T-cell lymphoma line Jurkat.

Analysis of Genomic Sequences. To determine the number of GM-CSF-related genes in human DNA, we hybridized a ^{32}P-labeled human GM-CSF probe to Southern blots of human DNA digested with a number of restriction enzymes expected to cut relatively infrequently. The results (Fig. 4) show that digestion with *Hin*dIII, *Eco*RI, or *Pst* I gave rise to a single band, while digestion with *Bgl* II gave rise to two bands. On the basis of these results, it appears probable that the gene for GM-CSF exists as a single-copy gene.

DISCUSSION

In this study we have used a murine GM-CSF cDNA probe to isolate a homologous sequence from human cDNA libraries. Two lines of evidence support the assumption that this sequence corresponds to the human GM-CSF gene. The amino acid sequence of the protein predicted from our nucleotide sequence includes nine amino acids that are identical to those reported by Gasson *et al.* (5) for the amino terminus of naturally occurring human GM-CSF. In addition, we have constructed an expression plasmid designed to direct the synthesis and secretion of this protein by yeast and have shown that it can stimulate the growth of granulocyte, macrophage, and granulocyte/macrophage colonies from human bone marrow.

In contrast, the secreted, recombinant human GM-CSF showed no activity in three biological assays that respond to recombinant murine GM-CSF. These are: a murine bone marrow colony assay, a murine bone marrow proliferation assay, and a murine bone marrow macrophage proliferation assay. Similarly, recombinant murine GM-CSF has shown no activity in human bone marrow assays (unpublished data). This lack of cross-species activity contrasts to that reported for human macrophage CSF, which can stimulate the growth of murine macrophages (19).

The human GM-CSF nucleotide sequence reported here predicts a molecular mass for the mature polypeptide of 14,476 daltons, starting at the position corresponding to the published amino terminus (position 1 in Fig. 2*B*). The naturally occurring glycoprotein purified from the Mo T-lymphoblast cell line is reported to have a molecular mass of 22,000 daltons (5), indicating that it could contain up to 34% carbohydrate by weight. Potential N-linked glycosylation sites exist at asparagine residues 27 and 37. Interestingly, the sequence of murine GM-CSF also contains two potential N-linked glycosylation sites, but in a different region of the molecule (asparagine residues 66 and 75). Comparison of the human and mouse sequences reveals substantial homology at both the nucleotide (69%) and amino acid (54%) levels throughout the coding regions. The amino terminus of murine GM-CSF has been reported to be Ile-7, whereas sequence comparison with human GM-CSF would suggest Ala-1 as a more likely possibility. The position of the carboxyl terminus of the two proteins is identical, as are the positions of the four cysteine residues, indicating the possible importance of disulfide bond formation in these molecules (5). The human sequence has an insertion of nine nucleotides relative to the mouse sequence, giving rise to three additional amino acids at positions 23–25.

The assignment of the initiator methionine at position −17 must be considered as preliminary in the absence of any upstream, in-phase, termination codons or of the full-length cDNA sequence. However, as we have sequenced to within 17 nucleotides of the 5′ end of the mRNA, and the signal sequence predicted here conforms well to those reported for

other secreted and membrane proteins, it is probable that we have identified the true initiator methionine.

We have shown that HUT-102 cells and peripheral blood T cells activated with PMA and Con A contained relatively high amounts of GM-CSF mRNA (Fig. 3 *Left*). It has previously been reported that HUT-102 cells produce a CSF (20). Interestingly, a human bladder carcinoma cell line (5637), reported to make an early hematopoietic factor (21), showed an abundance of GM-CSF mRNA. However, analysis of mRNA from the human pancreatic cell line (1420), which has been reported to make human macrophage CSF and possibly GM-CSF (22), produced no detectable GM-CSF mRNA.

The availability of recombinant human GM-CSF should now allow unequivocal analysis of the physiological role of GM-CSF in hematopoiesis and immune functions.

We thank Judy Byce for preparation of the manuscript; Donna Birdsell for performing bone marrow colony assays; and Steve Gimpel, Noel Balantac, and Tami Parr for excellent technical assistance. This work was made possible by an ongoing collaboration between Immunex Corporation, Hoechst A.G., and Behringwerke A.G.

1. Metcalf, D. (1981) in *Tissue Growth Factors*, ed. Basenger, R. (Springer, New York), 343–384.
2. Dexter, T. M. (1984) *Nature (London)* **309,** 746–747.
3. Burgess, A. W., Camakaris, J. & Metcalf, D. (1977) *J. Biol. Chem.* **252,** 1998–2003.
4. Metcalf, D. & Burgess, A. W. (1982) *J. Cell. Physiol.* **111,** 275–283.
5. Gasson, J. C., Heisbart, R. H., Kaufman, S. E., Clark, S. C., Hewick, R. M., Wong, G. G. & Golde, D. W. (1984) *Science* **226,** 1339–1342.
6. Gough, N. M., Gough, J., Metcalf, D., Kelso, A., Grail, D., Nicola, N. A., Burgess, A. W. & Dunn, A. R. (1984) *Nature (London)* **309,** 763–767.
7. Sparrow, L. G., Metcalf, D., Hunkapiller, M. W., Hood, L. E. & Burgess, A. W. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 292–296.
8. Cosman, D., Cerretti, D. P., Larsen, A., Park, L., March, D., Dower, S., Gillis, S. & Urdal, D. (1984) *Nature (London)* **312,** 768–771.
9. Grunstein, M. & Hogness, D. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 3961–3965.
10. Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26,** 101–106.
11. Sanger, F., Nicklen, B. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
12. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3963–3965.
13. Shiba, K., Ito, K., Yura, T. & Cerretti, D. P. (1984) *EMBO J.* **3,** 631–635.
14. Brake, A. J., Merryweather, J. P., Cort, D. G., Heberlein, U. A., Masiary, T. P., Mullenback, G. T., Urdea, M. S., Valenzuela, P. & Barr, P. J. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 4642–4646.
15. Kurjan, J. & Herskowitz, I. (1982) *Cell* **30,** 933–943.
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
17. Green, M. R., Maniatis, T. & Melton, D. A. (1983) *Cell* **32,** 681–694.
18. Watson, M. E. E. (1984) *Nucleic Acids Res.* **12,** 5145–5164.
19. Stanley, E. R., Chen, D.-M. & Lin, H. S. (1978) *Nature (London)* **274,** 168–170.
20. Olsson, I. L., Sarngadharan, M. G., Breitman, T. R. & Gallo, R. C. (1984) *Blood* **63,** 510–517.
21. Bartelmey, S. H. & Stanley, E. R. (1985) *J. Cell. Physiol.* **122,** 370–378.
22. Wu, M. C., Cini, J. K. & Yunis, A. A. (1979) *J. Biol. Chem.* **254,** 6226–6228.