Human GM-CSF: Molecular Cloning of the Complementary DNA and Purification of the Natural and Recombinant Proteins

Gordon G. Wong, JoAnn S. Witek, Patricia A. Temple
Kathleen M. Wilkens, Anne C. Leary, Deborah P. Luxenberg
Simon S. Jones, Eugene L. Brown, Robert M. Kay
Elizabeth C. Orr, Charles Shoemaker, David W. Golde
Randal J. Kaufman, Rodney M. Hewick
Elizabeth A. Wang, Steven C. Clark

Abstract. *Clones of complementary DNA encoding the human lymphokine known as granulocyte-macrophage colony-stimulating factor (GM-CSF) were isolated by means of a mammalian cell (monkey COS cell) expression screening system. One of these clones was used to produce recombinant GM-CSF in mammalian cells. The recombinant hematopoietin was similar to the natural product that was purified to apparent homogeneity from medium conditioned by a human T-cell line. The human T-cell GM-CSF was found to be 60 percent homologous with the GM-CSF recently cloned from murine lung messenger RNA.*

Proteins necessary for the survival, proliferation, and differentiation of hematopoietic progenitor cells are produced by a variety of murine and human cell types (*1*). The biological activity of these factors, known as colony-stimulating factors or CSF's, is measured by their ability to stimulate hematopoietic progenitor cells to form colonies in semisolid medium. The CSF's are classified by the types of mature blood cells found in the resulting colonies. Thus multi-CSF, also known as interleukin-3 (IL-3) (*2*), stimulates the progenitor cells for most of the hematopoietic cell lineages. Other CSF's have been identified which specifically stimulate committed progenitor cells of the granulocyte-monocyte lineage: macrophage-CSF (M-CSF; also known as CSF-1) (*3*) and granulocyte-CSF (G-CSF) (*4*) stimulate the proliferation of progenitors committed to the macrophage or granulocyte lineages, respectively, while granulocyte-macrophage CSF (GM-CSF; also known as CSF-2) (*1, 5*) stimulates the proliferation of cells of both lineages.

Results of experiments in vitro suggest a primary role for the various CSF's in the regulation of hematopoiesis (*1, 6*), but the paucity of highly purified material has impeded their study and virtually precluded analysis of their functions in vivo. Thus it has not been possible to evaluate the potential clinical utility of natural CSF in the treatment or prevention of human disease. The ability of GM-CSF to stimulate both neutrophilic granulocyte and macrophage production implies that this factor could be clinically useful in situations where increased production of these cells may be desirable, for example, in immuno-compromised patients or those about to undergo irradiation or chemotherapeutic treatment of cancer (*7*).

We now report the isolation of complementary DNA (cDNA) clones that express biologically active human GM-CSF. The cloning was accomplished by a novel direct method involving the construction of cDNA libraries in an expression vector and the screening of plasmid pools by transient expression in COS monkey cells. We also describe the purification of natural human GM-CSF to homogeneity. This protein served as a standard to demonstrate the authenticity of the CSF we further purified from medium conditioned by monkey cells expressing the cloned sequence. The recent publication (*8*) of the nucleotide sequence of a murine lung GM-CSF cDNA has allowed us to compare the sequences of the human and murine proteins.

Identification and isolation of a human GM-CSF cDNA clone. A number of human cell lines that produce GM-CSF have been described (*9*). The development of methods for transforming normal human T cells in vitro with the use of human T-cell leukemia virus (HTLV) has provided a routine means of generating continuous T-cell lines (*10*), many of which could serve as useful sources of the hematopoietin. We have used a naturally arising HTLV-transformed T-lymphoblast cell line designated Mo (*11*) as a starting point both for the purification of analytical amounts of the GM-CSF protein and for the isolation of GM-CSF cDNA clones. Like other HTLV transformed T-cell lines, the Mo cell line produces several regulatory factors including GM-CSF (*9, 12*). GM-CSF protein has been partially purified from the Mo-conditioned medium (*13*) and the GM-CSF mRNA has been identified by translation of Mo cell messenger RNA (mRNA) in *Xenopus* oocytes (*14*). However, the difficulty we encountered in obtaining sufficient sequence information from the very small quantities of GM-CSF that could be purified to homogeneity and the unreliability of the oocyte assay for the GM-CSF mRNA prompted us to develop a novel strategy to directly identify lymphokine cDNA clones by transient expression in COS-1 (*15*) cells.

Our cloning procedure exploits the efficient protein expression obtained after transfection of COS-1 monkey cells with the cDNA cloning vector (*16*) illustrated in Fig. 1. Control experiments demonstrated that p91203(B)-derived constructs synthesized proteins such as γ-interferon or interleukin-2 (IL-2) sufficiently well to allow detection after 500- to 1000-fold dilution with other plasmid DNA. It was therefore possible to construct cDNA libraries directly in p91203(B) and to screen pools of 300 or more recombinants for the ability to induce GM-CSF secretion by transfected COS-1 cells (see Table 1). The library, derived from membrane-bound mRNA of lectin-stimulated Mo cells, consisted of 60,000 independent clones. Two hundred pools, each containing 200 to 500 colonies, were screened initially, colony formation with the KG-1 human myeloid

All but one of the authors are members of the Genetics Institute, 225 Longwood Avenue, Boston, Massachusetts 02115. David W. Golde is in the Division of Hematology-Oncology, Department of Medicine, UCLA School of Medicine, Los Angeles 90024.

leukemia cell line being used as an indicator of CSF activity (*17*). Six pools induced the transfected COS cells to secrete GM-CSF activity; the same six pools were also positive when reassayed with fresh human bone marrow in the colony assay (Table 1). Standard sib-selection methods were used to identify one single GM-CSF cDNA clone present in each of three of the original six positive pools (Table 1). Hybridization and DNA sequence analysis proved that the three independent plasmids pCSF-1, -2, and -3 contain the same cDNA sequence. We subsequently identified, by colony hybridization, the same cDNA sequence in each of the three remaining positive plasmid pools. To confirm that this sequence encoded a GM-CSF, we examined several colonies from the bone marrow CSF assay of the transfected COS cell–conditioned medium and found that the majority of the cells were either granulocytes or macrophages (Fig. 2).

Sequence of human GM-CSF. The DNA sequence of the first GM-CSF cDNA (pCSF-1), shown in Fig. 3, contains a single long open reading frame of 432 nucleotides encoding 144 amino acids. Comparison of this sequence with the NH_2-terminal sequence from the purified Mo protein suggested that GM-CSF, like other secreted proteins, is synthesized as a precursor that is cleaved after residue 17 to yield a mature protein of 127 residues. Two potential NH_2-glycosylation sites (Asn-X-Thr/Ser) (*18*) were noted in the sequence (residues 44 to 46 and 54 to 56; Fig. 3). The complete nucleotide sequence of pCSF-1 consists of 754 nucleotides (exclusive of the polyadenylated stretch) including eight nucleotides of 5′ and 314 nucleotides of 3′ untranslated sequence.

Determination of the DNA sequence of four other, independently isolated, GM-CSF clones confirmed the sequence of pCSF-1. All of the clones contain the complete coding region identified in Fig. 3. In three of the clones, however, a single base substitution (thymine for cytosine) was noted at the nucleotide numbered 358. This was the only base change detected among all of the clones. As multiple independent clones were obtained having either sequence, we believe this difference probably results from a polymorphism among the GM-CSF genes in Mo cell DNA. Restriction enzyme analysis of the CSF genomic sequences in Mo cell DNA and in human liver DNA suggested that there is a single human gene encoding GM-CSF that has one or more introns (Fig. 4B). The sequence polymorphism found in Mo cell cDNA's was not detected at the level of restriction endonuclease mapping of the genomic DNA. We infer therefore that Mo has two different alleles of a common GM-CSF gene. Our results imply that there are no other genes with a DNA sequence very closely related to T cell GM-CSF, but do not rule out the possibility that there are other CSF genes that have little or no sequence homology.

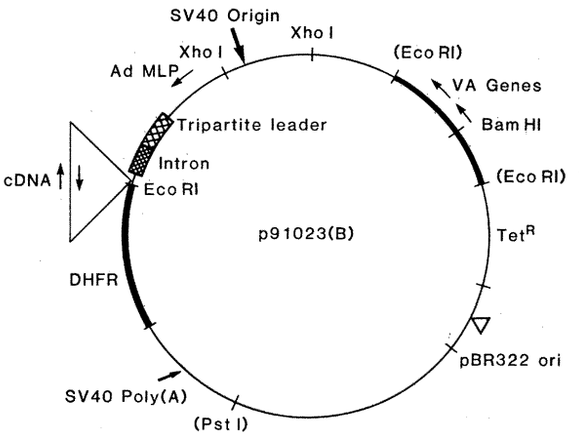


Fig. 1. Structure of expression vector p91023(B). The basic features of the expression system used have been described (*16*). The expression vector p91023(B) was derived from pQ2 (*16*) by standard DNA manipulations. The plasmid has the pBR322 origin of replication and tetracycline resistance gene for propagation in *Escherichia coli*, and lacks the bacterial sequences which inhibit DNA replication in COS monkey cells (*32*). It contains, in addition, eukaryotic regulatory elements from several different sources: (i) an SV40 origin and SV40 enhancer segment, which together allow the plasmid to replicate to a very high copy number in COS cells (*15*) and to efficiently transcribe inserted cDNA genes (*33*); (ii) the adenovirus major late promoter (*34*) coupled to a cDNA copy of the adenovirus tripartite leader (*16*); (iii) a hybrid intron consisting of a 5′ splice site from the first exon of the tripartite leader and a 3′ splice site from a mouse immunoglobulin gene (*16*); (iv) the SV40 early polyadenylation signal; and (v) the adenovirus VA I and VA II gene region (*35*). Plasmid p91023(B) also contains the mouse dihydrofolate reductase (DHFR) coding sequence downstream from the unique Eco RI cloning site which is positioned just after the 3′ splice site. cDNA's inserted at this Eco RI site can be transiently expressed at high levels in COS cells. The inserted cDNA is transcribed into a hybrid mRNA such that the cDNA sequence is flanked by the adenovirus tripartite leader at the 5′ side and by the mouse DHFR sequence serving as a nontranslated sequence at the 3′ side. The adenovirus tripartite leader and the VA RNA's increase the translatability (*16, 36*) while the DHFR sequence appears to enhance the stability of the hybrid mRNA (R. Kaufman, unpublished results).

Table 1. CSF activity in supernatants from COS cells transfected with DNA's from the Mo cell cDNA expression library. The Mo cell cDNA expression library in p91023 (*13*) was prepared by standard methods (*30*). Sixty thousand bacterial colonies from the library were replica-plated onto nitrocellulose filters. The colonies from each filter were scraped into L-broth and plasmid DNA was isolated as described (*31*). Each DNA sample was prepared from a pool of 200 to 500 independent clones. We used 5 μg of each DNA sample to transfect a 10-cm dish of COS-1 cells by DEAE-dextran–mediated DNA transfection with the addition of chloroquine treatment (*16*). The transfected COS cell–conditioned medium was harvested 72 hours after DNA transfection and activity was measured in KG-1 CSF assay (*17*). From the primary screen of 200 DNA samples, six pools (A) were found to produce CSF activity in COS cells. These pools were also assayed with the use of 10^4 light-density nonadherent human bone marrow cells as targets (*17*). The bacterial colonies from each of the first three positive pools were picked and placed in square matrices (pool M99 had approximately 340 clones, pool M181 had 370 clones, and pool M195 had 470 clones). DNA was prepared from each horizontal row and vertical column of each matrix, and transfected to produce COS cell–conditioned media samples which were assayed for CSF by the KG-1 cell assay. Two positive samples (B) were obtained from each set of transfections corresponding to each matrix, thereby unambiguously identifying the position of the CSF clone in the matrix. These cloned DNA's induced the expression of GM-CSF when transfected into COS-1 cells (C). NT, samples that were not tested in the bone marrow assay.

A			B			C		
DNA sample (primary pools)	CSF Activity (U/ml)		DNA sample (secondary pools)	CSF activity (U/ml)		DNA sample (isolated clone)	CSF activity (U/ml)	
	KG-1	Bone marrow		KG-1	Bone marrow		KG-1	Bone marrow
M99	40	500	M99-4	>300	1620	pCSF-1	7200	9700
			M99-35	>300	925			
M181	50	450	M181-1	>300	NT	pCSF-2	2400	NT
			M181-3	270	NT			
M195	60	540	M195-58	>300	NT	pCSF-3	2600	NT
			M195-30	300	NT			
M267	50	180						
M286	150	210						
M293	60	300						

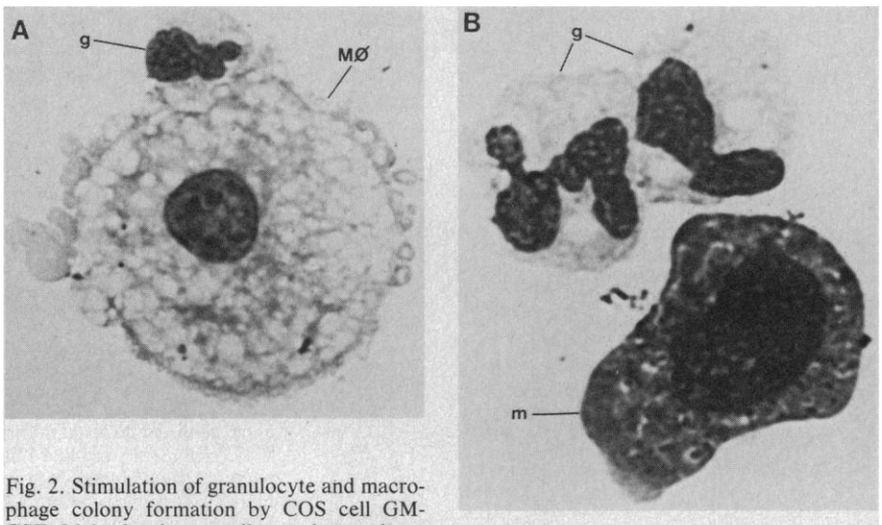
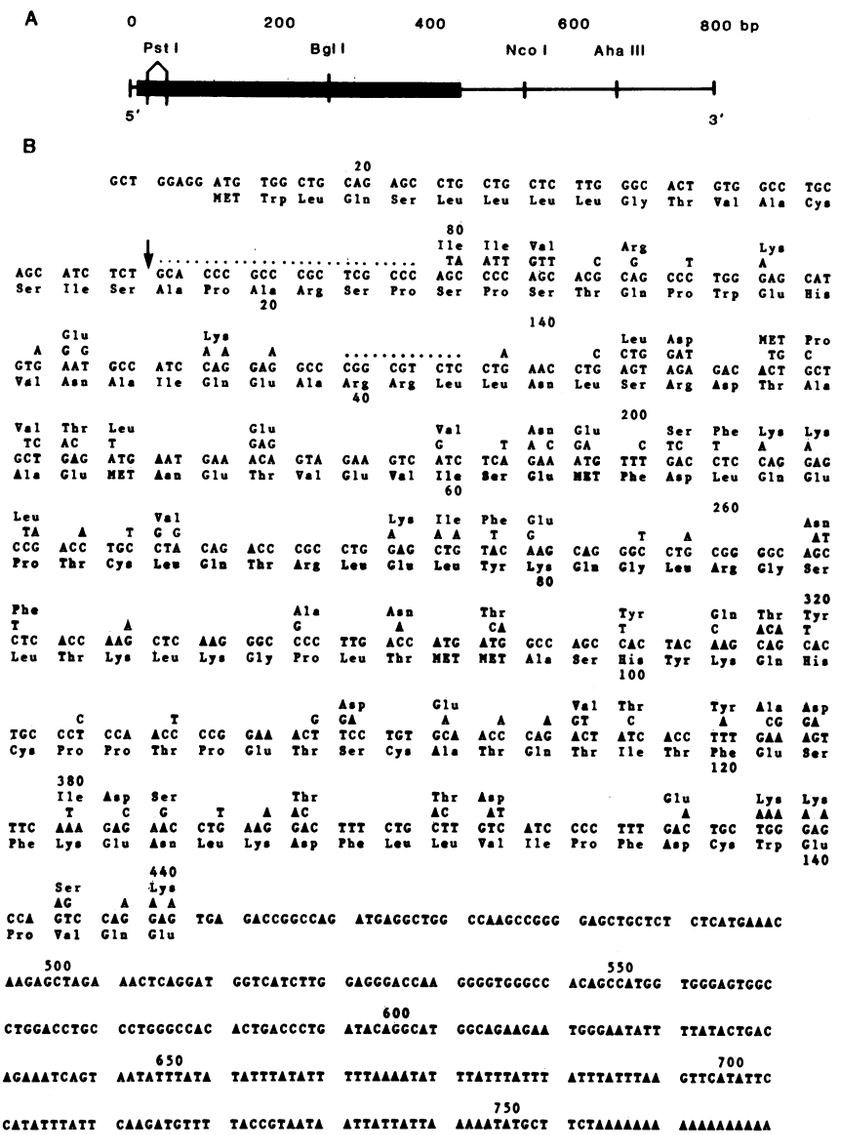


Fig. 2. Stimulation of granulocyte and macrophage colony formation by COS cell GM-CSF. Light-density nonadherent human bone marrow cells were plated (2×10^5 cells per milliliter) with 10 μl of conditioned medium from COS cells transfected with pCSF-1 in 35-mm dishes as described (*17*) except 0.87 percent methylcellulose was used instead of soft agar. More than 25 colonies were stimulated to grow; no colonies were observed in negative control dishes. A number of colonies were picked and the cells were dispersed onto slides and treated with Wright's stain. (A) Field showing a single macrophage (Mϕ) and a mature granulocyte (g); (B) field showing several mature granulocytes (g) and a single myelocyte, a granulocyte precursor (m).



Expression of GM-CSF mRNA. We examined mRNA from several cellular sources known to produce proteins with CSF activity. RNA blot analysis (Fig. 4A) revealed that two mature T-cell lines (Mo and UCD-144-MLA) (*34*), as well as lectin-stimulated peripheral blood lymphocytes (PBL's), synthesized readily detectable levels of a 1-kb mRNA that hybridized with the GM-CSF clone. The messenger could not be detected in RNA samples from the immature T-lymphoblastoid cell line CCRF-CEM (*19*) or from the B-lymphoblastoid cell line Daudi (*20*), neither of which produces measurable CSF (data not shown). Although Mo cells constitutively produce the transcript, treatment with phytohemagglutinin (PHA) and phorbol myristate acetate (PMA) increased the abundance of the message two- to fourfold whereas treatment with the glucocorticoid dexamethasone decreased the level of the CSF mRNA by a similar amount. Analysis of a second HTLV-transformed T-cell line, C10-MJ2 (*21*), confirmed that the CSF gene could be activated by lectin stimulation. In contrast to Mo cells, the level of GM-CSF mRNA detected in RNA obtained from the C10-MJ2 cells was very low in the absence of stimulation but could be increased more than tenfold by treatment of the cells with PHA and PMA. Isolation of further GM-CSF clones from a cDNA library prepared from lectin-stimulated C10-MJ2 cells suggested that the abundance of the CSF mRNA in these cells is 0.05 to 0.1 percent of the total mRNA.

Fig. 3. (A) Restriction map of human GM-CSF and (B) comparison of the sequences of the human and the murine GM-CSF's. (A) The cDNA insert of pCSF-1 is bounded by Eco RI linkers. Other restriction sites found in the sequence are as indicated. The heavy line indicates the position of the open reading frame. (B) The complete sequence of the cDNA insert of pCSF-1 was determined by the dideoxynucleotide chain termination method following subcloning of fragments into M13 vectors (*37*). The predicted amino acid sequence of the single long open reading frame is indicated below the nucleotide sequence. The position of cleavage of the signal peptide, indicated with an arrow, was determined by amino terminal sequencing of GM-CSF isolated from Mo cell–conditioned medium and recombinant GM-CSF from medium conditioned by COS cells transfected with pCSF-1 (*23*). The nucleotide sequence analysis of five independent CSF clones (all from the Mo cell cDNA library) revealed sequence heterogeneity at one position (two clones had the sequence shown while three clones had a T at position 358 instead of the C indicated here). The nucleotide sequence of the coding region of murine lung GM-CSF (*8*) along with predicted amino acid changes is indicated above the human sequence. The sequence as indicated (·····) is deleted in the mouse sequence relative to the human.

Purification of the natural and recombinant CSF's. The purification of human GM-CSF from T cell–conditioned medium proved to be difficult for several reasons. First, the time (10 to 14 days) required for colonies to grow in the bioassay made it difficult to follow the activity through sequential fractionation steps. Second, regardless of the source, GM-CSF is always present as a minor component of the total protein in conditioned medium and extensive purification is necessary to achieve homogeneity. Finally, the molecular heterogeneity of the GM-CSF protein, which is evident as heterogeneity in size, charge, and hydrophobicity [see also (*13*)] made fractionations based on these physical properties inefficient. Nevertheless, using conventional column chromatography and two reversed-phase high-performance liquid chromatography (HPLC) steps, we were able to extensively purify the T cell–derived GM-CSF (*22*). In a typical preparation, 40 liters of starting conditioned medium yielded 4 to 6 μg (about 200 to 400 picomoles) of the natural hematopoietin. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) of the protein in the most active fraction from the final purification step revealed that the CSF protein migrated as a single diffuse band with an apparent molecular mass of 18 to 22 kilodaltons (Fig. 5, lane 2). This highly purified GM-CSF had a specific activity of 1×10^7 to 4×10^7 units per milligram when assayed with human bone marrow target cells and was active at 4 to 20 picomolar concentrations in the KG-1 cell assay.

To produce quantities of recombinant GM-CSF, we used DNA transfection to introduce pCSF-1 into large numbers of COS-1 cells (see Fig. 5) which were subsequently allowed to condition medium in the absence of serum. By this procedure we could readily prepare 4- to 5-liter batches of serum-free conditioned medium with CSF activity 10 to 30 times higher than the activity in the T-cell conditioned medium. The recombinant GM-CSF had fractionation properties virtually identical to those of the natural protein as described above. This similarity plus the higher levels of activity generated by COS cells greatly facilitated the isolation of the recombinant GM-CSF (*23*). In fact, gel filtration followed by reversed-phase HPLC yielded 200 μg of recombinant CSF from 4 liters of conditioned medium (Fig. 5, lane 1). One of the most highly purified fractions, when analyzed by SDS-PAGE, migrated as a heterogeneous band with an apparent molecular mass of 18 to 24 kD, a result similar to that obtained with the natural T cell–derived protein (Fig. 5, lane 1). The specific activity of the recombinant CSF was indistinguishable from that of the native protein isolated from Mo cell–conditioned medium. We have been able to prepare milligram quantities of such highly purified material from the transiently expressing COS monkey cells.

The quantities of natural GM-CSF purified from the continuous T-cell line and the larger amounts of recombinant GM-CSF purified from the pCSF-1 transfected COS cells were sufficient for microsequencing of the NH_2-terminus (*24*). The sequence of the first six residues of natural GM-CSF was found to be Ala-Pro-Ala-Arg-Ser-Pro, which is identical to the sequence predicted from the cDNA sequence (Fig. 3) for residues 18 to 23. This confirms our identification of the cDNA as that for GM-CSF and indicates that the signal sequence is removed by cleavage after serine residue 17. With the recombinant GM-CSF we determined the sequence of the 16 residues (Ala-Pro-Ala-Arg-Ser-Pro-Ser-Pro-Ser-Thr-Gln-Pro-Trp-Glu-His-Val) corresponding to the 18th through the 32nd amino acids predicted by the DNA sequence. The fact that the same NH_2-terminal sequence was obtained with both natural and recombinant CSF provides evidence that the observed molecular heterogeneity of the purified proteins is not a result of differential processing at the NH_2-terminus but most likely is a result of heterogeneity in carbohydrate modification of the polypeptide. We cannot rule out the alternative possibility of COOH-terminal processing or other forms of posttranslational modification. In either case, our results demonstrated that the expression of a unique GM-CSF gene in COS-1 cells can produce protein with substantial molecular heterogeneity similar to that observed

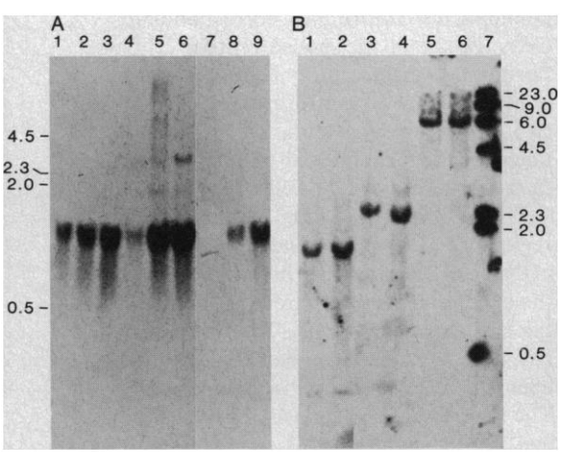


Fig. 4. Nitrocellulose blot analysis of (A) the CSF mRNA and (B) restriction fragments of the human CSF gene region. (A) Polyadenylated cytoplasmic RNA (5 μg), isolated as described (*16*), from Mo cells cultured for 20 hours in the presence of 10 n*M* dexamethasone (lane 1); from Mo cells (lane 2); from Mo cells cultured for 20 hours in the presence of PHA and PMA (see legend to Fig. 1) (lane 3); from C10 MJ2 cells (lane 4); from C10 MJ2 cells cultured for 20 hours in the presence of PHA and TPA (lane 5); from UCD-144-MLA cells cultured for 20 hours in the presence of 5 n*M* PMA (lane 6); from Ficoll-separated peripheral blood lymphocytes (PBL's) (lane 7); and from two different samples of PBL's pooled from four donors and stimulated for 20 hours with PHA and PMA (lanes 8 and 9). The samples were fractionated on a 1 percent agarose gel in the presence of 6 percent formaldehyde and transferred to nitrocellulose as described by Kaufman and Sharp (*16*). The nitrocellulose filter was hybridized with the GM-CSF cDNA clone labeled with ^{32}P to about 5×10^7 cpm/μg with T4 DNA polymerase replacement synthesis (*38*). The hybridization was performed for 16 hours at 68°C in a mixture containing 4× SSC (SSC contains 0.15*M* NaCl, 0.015*M* sodium citrate, *p*H 7.4), 5× PM (PM contains 0.02 percent each of bovine serum albumin, polyvinylpyrrolidone, and Ficoll 400), denatured herring sperm DNA (100 μg/ml), and the labeled probe (10^6 cpm/ml). After hybridization, the filter was washed at 68°C for several hours in 2× SSC with 0.1 percent SDS and for 30 minutes in 0.2× SSC with 0.1 percent SDS. The size markers were denatured λ Hind III fragments. (B) High molecular weight DNA (20 μg) from human liver (kindly provided by J. Toole) and from Mo cells (20 μg) was digested to completion with Eco RI (lane 5, Mo cell DNA; lane 6, human liver DNA), Pst I (lane 3, Mo cell DNA; lane 4, human liver DNA), and with both enzymes (lane 1, Mo cell DNA; lane 2, human liver DNA), fractionated by agarose gel electrophoresis (0.7 percent gel), and transferred to nitrocellulose (*39*). The resulting filter was probed as described above except 1×10^7 cpm of labeled probe per milliliter was used. The size markers are end-labeled fragments of a Hind III digest of λ DNA. Digestion of the human DNA's with Eco RI, which does not cleave within the cDNA sequence, gave a single hybridizing fragment of approximately 6000 base pairs (lanes 5 and 6). Cleavage with Pst I, which cleaves the cDNA twice within a 25 nucleotide sequence near the 5′ end of the clone, generates two hybridizing fragments of 3000 and 300 nucleotides, respectively (lanes 3 and 4). Since the ^{32}P-labeled cDNA probe only extends 25 nucleotides upstream from the first Pst I site, this probe will not anneal with the genomic Pst I fragment containing the first 25 nucleotides of the cDNA sequences. Therefore, the two hybridizing fragments probably result from an additional Pst I site contained within an intron in the genomic sequence. Cleavage of the genomic DNA's with both Pst I and Eco RI yields two hybridizing fragments, one 2000 base pairs and the other 300 base pairs in length (lanes 1 and 2). From these data we conclude that under the hybridization stringency used here, there is a single human GM-CSF gene that contains at least one intron.

with the natural purified T-cell protein.

Discussion. The isolation by expression screening of six virtually identical independent CSF clones from a library of 60,000 clones suggests that the GM-CSF reported here is the major species of CSF produced by Mo cells. Alternatively, CSF's encoded by other mRNA's may be more difficult to convert to full-length cDNA's or more difficult to express in the COS cell expression system and therefore could have been missed in the expression screen. Antibodies to the recombinant GM-CSF protein would help clarify the issue of whether or not there are other human proteins with GM-CSF activity.

We have compared the sequence of human GM-CSF with all the sequences recorded in Genbank including those of oncogenes and such growth factors as γ-interferon, IL-2, and murine IL-3, and with the recently published sequence for murine GM-CSF. The only significant homology that we detected was with the murine GM-CSF, which was recently cloned from murine lung mRNA. Maximum homology of the coding regions of the human and mouse cDNA's was obtained by deleting the codons for amino acids 18 to 23 and 40 to 42 from the human sequences as indicated in Fig. 3. The amino acid sequences are then 60 percent homologous while the nucleotide sequences are approximately 70 percent conserved. The sequence of the murine signal peptide was not reported, but it is interesting that the processing of the human GM-CSF apparently occurs six amino acids before the corresponding cleavage site of the mouse protein. The positioning of the four cysteine residues appears to have been preserved in the two proteins, suggesting some importance for the location of disulfide bridges in the structure of GM-CSF.

The sequences of the 3′ noncoding regions of the murine and human CSF cDNA clones show little homology, with the exception of a long stretch of nucleotides consisting of T's and A's beginning roughly 110 nucleotides before the polyadenylic acid sequence in each clone. The mouse sequence consisting of 59 T's and A's [nucleotides 542 to 620 in (*8*)] contains within it an almost exact replica (93 percent homologous) of a 47-nucleotide sequence (nucleotides 646 to 692) in the human cDNA. Regions rich in A and T which share some homology with this sequence have been found in the 3′ noncoding regions of other cDNA's, notably the human and mouse sequences for interferon α2 (*25*). Other conserved sequences have been observed in 3′ noncoding regions, but the function of such sequence elements remains unclear (*26*).

Another factor produced by Mo cells, T-cell derived neutrophil migration inhibitory factor (NIF-T), has recently been purified to homogeneity (*27*). NIF-T was identified by its ability to inhibit the migration of human neutrophils in vitro, but the highly purified protein also proved to have GM-CSF activity (*27*). The partial NH_2-terminal sequence determined from the purified NIF-T exactly corresponded with the sequence of the GM-CSF protein reported here. The identity of NIF-T and GM-CSF was confirmed by the demonstration that our recombinant GM-CSF has high levels of NIF-T activity (*27*). Thus a single hematopoietin can not only stimulate the growth and differentiation of progenitor cells in bone marrow but can also activate the biological function of the resulting circulating mature blood cells. These results are consistent with the findings that the purified murine G-CSF and GM-CSF can activate murine neutrophil cytotoxic functions (*28*) and that semipurified human CSF from placental-conditioned medium can enhance the antibody-dependent cell-mediated cytotoxicity of human neutrophils (*29*). Thus GM-CSF and NIF-T may have several functions, including the stimulation of the production of effector cells and ultimately the activation of these cells in the periphery. With the availability of the human GM-CSF cDNA clone it should be possible to produce pure CSF in quantities sufficient to evaluate the relative importance of these activities in vivo and to test the clinical potential of CSF for the treatment and prevention of granulocytopenia and infection.

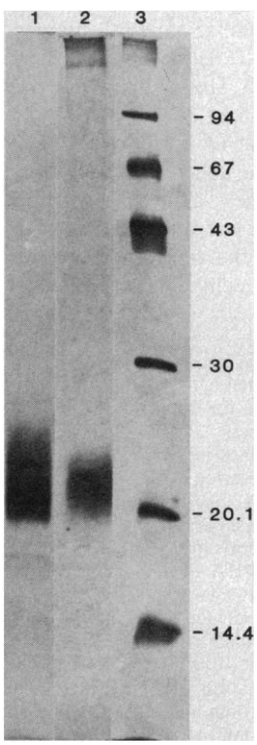


Fig. 5. Electrophoresis (SDS-PAGE) of purified Mo T cell and recombinant GM-CSF. Highly purified Mo GM-CSF (1 μg; lane 2) and recombinant GM-CSF (approximately 4 μg; lane 1) were subjected to electrophoresis (13.5 percent SDS-PAGE) in parallel with the indicated molecular size markers (lane 3) as described (*40*). The protein bands in the gel were visualized by silver staining. The Mo GM-CSF was isolated from 40 liters of serum-free Mo-conditioned medium (*22*). Recombinant CSF was isolated from 4 liters of serum-free conditioned medium from COS-1 cells transfected with pCSF-1 (*23*).

References and Notes

1. D. Metcalf, *Hematopoeitic Colonies. In vitro Cloning of Normal and Leukemic Cells* (Springer-Verlag, New York, 1977); A. W. Burgess and D. Metcalf, *Blood* **56**, 947 (1980); M. A. S. Moore, *Clin. Hematol.* **8**, 287 (1979); N. A. Nicola and M. Vadas, *Immunol. Today* **5**, 76 (1984).
2. J. H. Ihle, L. Rebar, J. Keller, J. C. Lee, A. J. Hapel, *Immunol. Res.* **63**, 5 (1982). D. Metcalf, in *Normal and Neoplastic Haemopoiesis*, D. W. Golde and P. A. Marks, Eds. (Liss, New York, 1983), p. 141.
3. E. R. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 2969 (1979); R. K. Shadduck, G. Pigoli, A. Waheed, *J. Supramol. Struct. Suppl.* **4**, 116 (1980).
4. N. A. Nicola, D. Metcalf, M. Matsumoto, G. R. Johnson, *J. Biol. Chem.* **258**, 9017 (1983).
5. A. W. Burgess, J. Camakaris, D. Metcalf, *J. Biol. Chem.* **252**, 1998 (1977).
6. W. Hocking, J. Goodman, D. W. Golde, *Blood* **61**, 600 (1983).
7. J. C. Gasson, I. S. Y. Chen, C. A. Westbrook, D. W. Golde, in *Normal and Neoplastic Hematopoiesis*, D. W. Golde and P. A. Marks, Eds. (Liss, New York, 1983), p. 129.
8. N. M. Gough *et al.*, *Nature (London)* **309**, 763 (1984).
9. M.-C. Wu, J. K. Clini, A. A. Yunis, *J. Biol. Chem.* **254**, 6226 (1979); C. Tarella, F. W. Ruscetti, B. J. Poiesz, A. Woods, R. C. Gallo, *Blood* **59**, 1330 (1982).
10. M. Popovic *et al.*, *Science* **219**, 856 (1983); I. S. Y. Chen, S. G. Quan, D. W. Golde, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 7006 (1983).
11. A. J. Saxon, R. H. Stevens, D. W. Golde, *Ann. Intern. Med.* **88**, 323 (1978); V. S. Kalyanaraman *et al.*, *Science* **218**, 571 (1982).
12. D. W. Golde, S. G. Quan, M. J. Cline, *Blood* **52**, 1068 (1978); D. W. Golde, N. Bersch, S. G. Quan, A. J. Lusis, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 593 (1980).
13. A. J. Lusis, D. H. Quon, D. W. Golde, *Blood* **57**, 13 (1980).
14. A. J. Lusis, D. W. Golde, D. H. Quon, L. A. Lasky, *Nature (London)* **298**, 75 (1982).
15. Y. Gluzman, *Cell* **23**, 175 (1981).
16. R. J. Kaufman and P. A. Sharp, *Mol. Cell. Biol.* **2**, 1304 (1982); R. J. Kaufman, *Proc. Nat. Acad. Sci. U.S.A.*, in press.
17. H. P. Koeffler and D. W. Golde, *Science* **200**, 1153 (1978); A. J. Lusis and H. P. Koeffler, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5346 (1980). For the KG-1 CSF assay, 10 μl samples were added to microtiter wells containing 400 KG-1 cells in 140 μl of Iscoves medium with 0.3 percent agar, 20 percent fetal calf serum, and 10^{-4} *M* α-thioglycerol. The assays were scored after 10 to 14 days of incubation at 37°C by determining the increase in the number of KG-1 colonies over background. One unit of KG-1 CSF activity per milliliter is the concentration required to achieve a stimulation of 50 percent of the maximum response achieved when the CSF concentration is saturating. In the bone marrow CSF assay, 10^4 light density nonadherent bone marrow cells from normal human donors were substituted for the KG-1 cells. One unit of CSF in the bone marrow cell assay is that amount which will stimulate the formation of one colony per 10^4 cells above background when the CSF is below saturation levels.
18. R. J. Winzler, in *The Chemistry of Glycoproteins in Hormonal Proteins and Peptides*, C. H.

Li, Ed. (Academic Press, New York, 1973), vol. 1, pp. 1–15.
19. G. E. Foley, *Cancer* **18**, 522 (1965).
20. E. Klein and G. Klein, *Cancer Res.* **28**, 1300 (1968).
21. S. K. Arya, F. Wong-Staal, R. C. Gallo, *Science* **223**, 1086 (1984).
22. The GM-CSF was isolated from 40 liters of heat-inactivated (30 minutes at 55°C) Mo cell–conditioned medium. The medium was concentrated by ultrafiltration, and the protein precipitated by addition of solid ammonium sulfate to 80 percent saturation. The precipitated protein (800 mg) was resuspended in 100 ml of 20 m*M* tris-HCl, *p*H 7.4, dialyzed, and then applied to a DEAE-Ultrogel column (2.5 by 10 cm). CSF activity, measured in the human bone marrow colony formation assay (*17*), trailed the protein peak that eluted from the column with 0.12*M* NaCl and 20 m*M* tris-HCl, *p*H 7.4. The active fractions (30-ml total volume) were pooled and concentrated by ultrafiltration and further fractionated by gel filtration (1.6 by 100 cm AcA44 Ultrogel column) in 20 m*M* Hepes, *p*H 7.4, 50 m*M* NaCl, and 0.01 percent polyethylene glycol (PEG-8000). The CSF activity emerged from the column with an apparent molecular weight of 30 kilodaltons. The pooled, active fractions were brought to 0.15 percent trifluoroacetic acid (TFA) and applied to a Vydac C4 reversed-phase column (1 by 25 cm) equilibrated in 0.1 percent TFA. The column was developed with a gradient of 0 to 9 percent acetonitrile in 0.1 percent TFA. The CSF activity eluted at approximately 47 percent acetonitrile. The pooled active fractions were brought to 0.05 percent heptafluorobutyric acid (HFBA) and applied to a Vydac C4 column (0.46 by 25 cm) equilibrated in 0.15 percent HFBA. The column was developed with a linear gradient of 0 to 90 percent acetonitrile in 0.15 percent HFBA. The CSF activity eluted at about 53 percent acetonitrile. The final yield was about 4 μg of protein in 1 ml. This sample had a specific activity of 1×10^7 to 4×10^7 U/mg on human bone marrow and gave half maximum stimulation of KG-1 colonies at about 1×10^{-11} to 5×10^{-11} *M*.
23. Recombinant CSF was isolated from 4 liters of conditioned medium from COS-1 cells transfected with pCSF-1 as described (see Table 1) except the transfections were performed with 1.2×10^8 COS cells in cell factories (Nunc) and the final medium fed to the cells was serum free. The conditioned medium was concentrated by ultrafiltration and the CSF activity fractionated by ammonium sulfate precipitation (the activity was recovered in the 30 to 80 percent fraction). The precipitated protein was resuspended in 20 m*M* sodium citrate, *p*H 6.1, containing 1*M* NaCl, and fractionated by gel filtration through a column (1.6 by 100 cm) of Ultrogel AcA54 equilibrated in the same buffer. Under these conditions, the CSF emerges from the column with an apparent molecular weight of 19 kD. The active fractions were pooled and brought to 0.15 percent TFA and applied to a Vydac C4 column (0.46 by 25 cm) equilibrated in 0.1 percent TFA. The column was developed with a 0 to 90 percent acetonitrile (1 ml/min) in 0.1 percent TFA. The CSF activity eluted between 39 and 43 percent acetonitrile. One of the peak fractions (fraction 19) contained approximately 40 μg of protein in 1 ml. Fraction 19 had a specific activity of 1×10^7 to 4×10^7 U/mg on human bone marrow and stimulated half colony number in the KG-1 cell assay at 1×10^{-11} to 5×10^{-11} *M*.
24. R. M. Hewick, M. W. Hunkapillar, L. E. Hood, W. J. Dreyer, *J. Biol. Chem.* **256**, 7990 (1981).
25. R. M. Lawn *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5435 (1981); G. D. Shaw *et al.*, *Nucleic Acids Res.* **11**, 555 (1983).
26. N. J. Cowan, P. R. Dobner, E. V. Fuchs, D. W. Cleveland, *Mol. Cell. Biol.* **3**, 1738 (1983); N. A. Tchurikov, A. K. Naumova, E. S. Zelentsova, G. P. Georgiev, *Cell* **28**, 365 (1982).
27. R. H. Weisbart, R. Billing, D. W. Golde, *J. Lab. Clin. Med.* **93**, 622 (1979); J. C. Gasson *et al.*, *Science* **226**, 1339 (1984).
28. A. F. Lopez *et al.*, *J. Immunol.* **131**, 2983 (1983).
29. M. A. Vadas, N. A. Nicola, D. Metcalf, *J. Immunol.* **131**, 795 (1983).
30. The Mo cell cDNA expression library was prepared beginning with membrane-bound mRNA [B. Mechler and T. H. Rabbitts, *J. Cell Biol.* **88**, 29 (1981)] from 2×10^9 Mo cells that had been stimulated for 16 hours with PHA (0.3 percent) and PMA (5 ng/ml) in Iscove's medium with 20 percent fetal calf serum (FCS) at 5×10^5 cells per milliliter. Double-stranded cDNA was prepared according to U. Gubler and B. J. Hoffman [*Gene* **25**, 263 (1983)] with ribonuclease H and DNA polymerase I in the second strand reaction. This cDNA was Eco RI methylated and ligated to Eco RI linkers as described [T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982]. The vector p91023(B) was linearized at the unique Eco RI site, treated with alkaline phosphatase, and ligated to equimolar amounts of cDNA at a final DNA concentration of approximately 100 μg/ml. The ligation reaction was extracted with phenol and chloroform, precipitated with ethanol, and used to transform *E. coli* strain MC1061.
31. J. A. Meyers, D. Sanchez, L. P. Elwell, F. Falkow, *J. Bacteriol.* **127**, 1529 (1976).
32. M. Lusky and M. Botchan, *Nature* (*London*) **293**, 79 (1981).
33. P. Mellon, V. Parker, Y. Gluzman, T. Maniatis, *Cell* **27**, 279 (1981).
34. E. Ziff and R. Evans, *ibid.* **15**, 1463 (1978).
35. M. B. Mathews, *ibid.* **6**, 223 (1975).
36. J. Logan and T. Shenk, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3655 (1984); R. J. Schneider, C. Weinberger, T. Shenk, *Cell* **37**, 291 (1984).
37. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977); J. Messing and J. Vieira, *Gene* **19**, 269 (1982).
38. P. O'Farrell, *Focus (Bethesda Research Laboratories)* **3** (No. 3), 1 (1981).
39. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
40. U. K. Laemmli, *Nature* (*London*) **227**, 680 (1970).
41. We thank J. Brown, P. Vanstone, M. Coe, L. Wasley for technical assistance, M. Erker and M. Richardson for help in preparing the manuscript, J. Gasson for discussions and help in identifying blood cell types, R. C. Gallo for cell lines, D. Stone for help with organizing the project, and R. Kamen and J. Lauer for review of the manuscript and discussions. This work was supported by Sandoz, Ltd., Basel, Switzerland.

7 November 1984; accepted 15 January 1985

RESEARCH ARTICLE

The LDL Receptor Gene: A Mosaic of Exons Shared with Different Proteins

Thomas C. Südhof, Joseph L. Goldstein
Michael S. Brown, David W. Russell

Cell surface receptors are multifunctional proteins with binding sites that face the external environment and effector sites that couple the binding to an intracellular event. Many receptors have an additional function: they transport bound ligands into cells (*1*). Such receptor-mediated endocytosis requires that the proteins have specific domains that allow them to cluster within clathrin-coated pits on the plasma membrane and in many cases to recycle to the cell surface after ligand delivery (*2*).

The multiple functions of coated pit receptors imply that they will have multiple domains, each with a single function. The structural features responsible for these functions are currently the subject of intense study. Recent insights have emerged from the complementary DNA (cDNA) cloning of the messenger RNA's (mRNA) for several receptors and the subsequent determination of their amino acid sequences. These studies have revealed surprising homologies between the primary structures of receptors and other proteins. For example, the receptor for plasma low-density lipoprotein (LDL), a cholesterol transport protein, contains one region that is homologous to the precursor of a peptide hormone, epidermal growth factor (EGF) (*3, 4*), and another region that is homologous to complement component C9, the terminal component of the complement cascade (*5*). The cell surface receptor for immunoglobulin A/immunoglobulin M is homologous to the immunoglobulins themselves (*6*). Finally, the receptor for EGF is homologous to a viral and cellular gene, *erb-B*, that produces a protein with tyrosine kinase activity (*7*).

These findings suggest that coated pit receptors share functional domains with other proteins. One likely mechanism for such sharing is through the duplication and migration of exons during evolution (*8*). Although the cDNA's for five coated pit receptors have been isolated and sequenced (*4, 6, 9*), the organizations of the genes encoding these proteins are not yet known. The elucidation of the gene structures of coated pit receptors should reveal the relationships between exons and protein domains and provide insight

Thomas C. Südhof is a postdoctoral fellow, Joseph L. Goldstein and Michael S. Brown are professors, and David W. Russell is an assistant professor in the Department of Molecular Genetics, University of Texas Health Science Center at Dallas, Southwestern Medical School, Dallas 75235.