acquiring pocket calculators and then personal computers can be appreciated only by those old enough to still remember owning mechanical slide rules.

Precisely the same transition is at hand in chemical and biological laboratories, where miniaturization and machine intelligence will return measurement power to the end user, while freeing him from becoming either a specialist in a different field or an almost-full-time money-raiser to do research in his own field. This open-ended task will occupy the coming decade and more, but will be indeed worth the effort.

References and Notes

1. S. Heller and R. Poterzone, Eds., *Computer Applications in Chemistry* (Elsevier, New York, 1983).
2. P. Lykos, *Personal Computers in Chemistry* (Wiley-Interscience, New York, 1981).
3. T. Hirschfeld, in *Fourier Transform Infrared Spectroscopy*, J. Ferraro, Ed. (Academic Press, New York, 1979), p. 193.
4. E. R. Malinovski and D. G. Howery, *Factor Analysis in Chemistry* (Wiley-Interscience, New York, 1980).
5. D. L. Massart and L. Kaufman, *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis* (Wiley, New York, 1983).
6. W. E. Biles, *Optimization and Industry Experimentation* (Wiley-Interscience, New York, 1980).
7. B. Kowalski, *Chemometrics: Theory and Application* (American Chemical Society, Washington, D.C., 1977).
8. M. Delaney, *Anal. Chem.* **56**, 261 R (1984).
9. T. Hirschfeld and E. Stark, in *Food Analysis*, G. Charambolous, Ed. (Wiley-Interscience, New York, 1984), p. 505.
10. T. Hirschfeld, paper presented at the American Society for Testing Materials Symposium on Computerized Infrared, Philadelphia (1984).
11. G. A. Kentgen, *Anal. Chem.* **56**, 69 R (1984).
12. T. Hirschfeld, *ibid.* **52**, 297 A (1980).
13. P. C. Jurs and T. L. Isenhour, *Chemical Applications of Pattern Recognition* (Wiley-Interscience, New York, 1975).
14. K. Eckschlager and V. Stepanek, *Information Theory as Applied to Chemical Analysis* (Wiley-Interscience, New York, 1984).
15. T. Hirschfeld, J. Callis, B. Kowalski, *Science* **226**, 312 (1984).
16. T. Hirschfeld, *Energy Technolgoy Review* (Lawrence Livermore National Laboratory, Livermore, Calif., February 1984), pp. 16–23.

17. This work was performed under the auspices of the Department of Energy by the Lawrence Livermore National Laboratory under contract W-7405-ENG-48.

RESEARCH ARTICLE

Molecular Cloning of a Complementary DNA Encoding Human Macrophage-Specific Colony-Stimulating Factor (CSF-1)

Ernest S. Kawasaki, Martha B. Ladner, Alice M. Wang
Janelle Van Arsdell, M. Kim Warren, Mazie Y. Coyne
Vicki L. Schweickart, Mei-Ting Lee, Kenneth J. Wilson
Albert Boosman, E. Richard Stanley, Peter Ralph, David F. Mark

The generation of granulocytes and macrophages from immature hematopoietic progenitor cells depends on the presence of several hormone-like, glycoprotein growth factors, the colony-stimulating factors or CSF's (*1*). There are several subclasses of CSF's, and they can be distinguished by the kind of mature cells that they produce in semisolid culture media.

One subclass, CSF-1, stimulates monocyte and macrophage production predominantly or exclusively (*2*). Other human CSF's stimulate the proliferation of progenitors committed to neutrophil and macrophage, and neutrophil or eosinophil lineages (*3*). Interleukin-3 (IL-3) (*4, 5*) in the murine system and, perhaps, a similar pluripotent CSF in the human (*6*) can stimulate the proliferation of monocytic, granulocytic, erythroid, megakaryocytic, and mast cell colonies. Although all these CSF's, with the exception of the eosinophilic type, have been purified, the small quantities available from natural sources have hindered their biochemical and biological characterization. The molecular cloning of two of the above factors, murine IL-3 (*7, 8*), and murine and human granulocyte-macrophage CSF (GM-CSF) (*9, 10*), should aid in elucidating the mechanisms by which these hematopoietic factors act.

Native human CSF-1 is a heavily glycosylated homodimer with a molecular size of ~45,000 daltons (*11*). It can be clearly distinguished from the other subclasses by specific radioimmuno- and radioreceptor assays (*2*). The proliferative effects of CSF-1 are restricted solely to the mononuclear phagocytic lineage (*12*), and the specific cell receptors that mediate the biological effects of CSF-1 apparently occur exclusively on cells of the same lineage (*13*). In the murine system, L cell–conditioned medium is a convenient source of CSF-1 (*14*), and amino-terminal and internal amino acid sequences have been determined (*15*).

Isolation of CSF-1 genomic clones. We now describe the isolation of CSF-1 genomic clones using oligonucleotide probes derived from the amino terminal sequence of human CSF-1. Human urine was used as a source of CSF-1 protein (*2*). CSF-1 was purified as described (see legend to Fig. 1), and the NH_2-terminal sequence was determined (*16*). The 12–amino acid NH_2-terminal sequence obtained from the human urinary CSF-1 (Fig. 1A) is highly homologous to the NH_2-terminus of murine L-cell CSF-1 (see Fig. 2B). The finding that the NH_2-termini were similar in sequence indicated that we were studying the same protein, even though the CSF-1 was prepared from widely different sources. Two oligonucleotide probes were derived from the NH_2-terminal sequence (Fig. 1B). A 16-fold degenerate, 35-base oligonucleotide, complementary to the NH_2-terminus, was constructed according to preferred mammalian codon usage (*17*). An 18-base, 64-fold degenerate oligonucleotide derived from amino acids 5 through 10 was also made to aid in screening.

Using the two oligonucleotides described above, we screened a human genomic library (*18*), and several CSF-1 genomic clones were isolated. Restriction enzyme and Southern gel (*19*) analysis of nine separate phage isolates showed that a common 3.8-kilobase (kb) Hind III fragment hybridized strongly to both of the oligonucleotide probes. This Hind III fragment was subcloned into M13mp19 (*20*) for further analysis (Fig. 2A). DNA sequence analysis (Fig. 2B)

E. S. Kawasaki, M. B. Ladner, A. M. Wang, M. Y. Coyne, V. L. Schweickart, and D. F. Mark are in the Department of Molecular Biology; J. Van Arsdell is in the Department of Microbial Genetics; M. K. Warren, M-T. Lee, and P. Ralph are in the Department of Cell Biology; and K. J. Wilson and A. Boosman are in the Department of Protein Chemistry at Cetus Corporation, 1400 Fifty-Third Street, Emeryville, California 94608. E. R. Stanley is in the Departments of Microbiology and Immunology, and of Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461. Send correspondence to E.S.K. The present address for K.J.W. is Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404.

(*21*) of the subclone indicated the presence of at least two exons and one intron. The assignment of coding regions downstream from the human 12–amino acid NH_2-terminus was accomplished by comparison with a longer 30–amino acid NH_2-terminal sequence obtained from (*23*). To partially enrich for CSF-1 RNA sequences, mRNA isolated from PMA-induced cells was fractionated on a sucrose gradient. Samples (50 nl) from the gradient fractions were injected into *Xenopus* oocytes (*24*), and the supernatants were tested in a mouse bone marrow proliferation assay (*25, 26*). Also, a sample (5 μl) from each of the fractions was dot blotted (*27*) and probed with the exon 2 oligonucleotide. A peak that contained bone marrow proliferative activity was found at ~14*S* to 16*S* in the gradient (Fig. 3A). However, the fractions that hybridized most intensely to the probe were larger than 18*S* (Fig. 3B). Since Northern blot analysis (see Fig. 6A) with the second exon probe indicated that there were multiple species of CSF-1 mRNA, ranging from ~1.5 to 4.5 kb, we decided to pool the activity peak fractions 8 and 9, plus the strongly hybridizing fraction 11, for use in constructing a cDNA bank by the Okayama-Berg method (*28*). The larger molecular size fractions 14 to 19, larger than 18*S*, were not used because mRNA isolated from MIA-PaCa cells, which were negative in CSF-1 production, also showed hybridization in this area (see Fig. 6A, lane 2).

Abstract. *Complementary DNA (cDNA) clones encoding human macrophage-specific colony-stimulating factor (CSF-1) were isolated. One cDNA clone codes for a mature polypeptide of 224 amino acids and a putative leader of 32 amino acids. This cDNA, which was cloned in the Okayama-Berg expression vector, specifies the synthesis of biologically active CSF-1 in COS cells, as determined by a specific radioreceptor assay, macrophage bone marrow colony formation, and antibody neutralization. Most of the cDNA isolates contain part of an intron sequence that changes the reading frame, resulting in an abrupt termination of translation; these cDNA's were inactive in COS cells. The CSF-1 appears to be encoded by a single-copy gene, but its expression results in the synthesis of several messenger RNA species, ranging in size from about 1.5 to 4.5 kilobases.*

murine L cell CSF-1 (Fig. 2B). The overlined region in the intron defines the sequence found in one of the CSF-1 complementary DNA (cDNA) clones (see Fig. 4). The underlined region in the second exon indicates the sequence from which an exact match complementary 32-base oligonucleotide probe (exon 2 probe) was derived for screening a cDNA bank.

Isolation of CSF-1 cDNA clones. Messenger RNA (mRNA) isolated from the pancreatic tumor cell line, MIA-PaCa-2 (*22*), was used to prepare a cDNA bank. This line constitutively synthesizes moderate levels of CSF-1 in serum-containing but not in serum-free media. When MIA-PaCa cells are incubated with phorbol myristic acetate (PMA) in serum-free conditions, CSF-1 production resumes

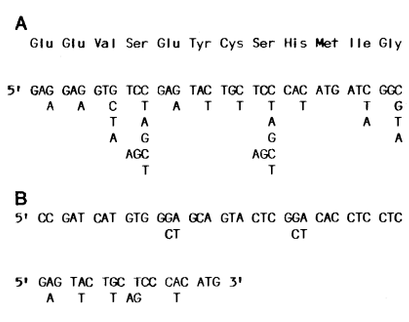


Fig. 1. NH_2-terminal sequence of human urinary CSF-1 and oligonucleotide probes derived from the sequence. (A) NH_2-terminal 12 residues of CSF-1 and possible coding sequence. (B) Oligonucleotide probes: 35-base consensus sequence complementary to the entire amino terminus and an 18-base 64-fold degenerate oligonucleotide corresponding to the coding sequence of amino acids 5 through 10. Stage V human urinary CSF-1, 10^6 units, 2 × 10^6 units per milligram of protein (*2*), was further purified by immunoaffinity chromatography on a rat monoclonal antibody to murine CSF-1 (*50*) attached to Sepharose B column (*51*). One-third of the preparation was applied to a Brownlee RP-300 high-performance liquid chromatography column which was eluted with a gradient from 30 to 60 percent acetonitrile in 0.1 percent trifluoroacetic acid. Fractions (2 ml) were collected into siliconized tubes containing 100 μl of 1*M* NH_4HCO_3, lyophilized, resuspended in H_2O, and assayed by RRA. The fractions with peak CSF-1 activity, spanning two protein peaks, were pooled for sequencing. The specific activity of this pool (~8 × 10^7 units per milligram of protein) agreed closely with the specific activity of purified murine CSF-1 (~8 × 10^7 unit/mg) (*51*) and the estimated specific activity of purified human CSF-1 (~5 × 10^7 unit/mg) (*2*). The pooled material (~100 pmol) was subjected to NH_2-terminal sequence analysis (*16*) (Applied Biosystems gas-phase sequenator). The Cys at residue 7 was inferred by homology with the mouse sequence (*15*). Oligonucleotides were synthesized by the phosphoramidite method (*52*) on a Biosearch SAM-1 DNA synthesizer. The 35 bases of the consensus sequence were chosen based on mammalian preferred codon usage (*17*).

Approximately 300,000 cDNA clones were screened with the exon 2 probe. Ten clones that were hybridization-positive were isolated, and preparative amounts of plasmid DNA were prepared. The purified DNA's were screened by a transient expression assay in COS-7 cells (*29*). Only one clone (pcCSF-17) expressed CSF-1 activity in this system (Table 1). This clone and an inactive clone (pcCSF-12) were picked for further analysis. A partial restriction map of the two clones is shown (Fig. 4) with their coding regions indicated by the boxed areas. Sequence analysis (Fig. 5) of pcCSF-17 indicated that there was only one methionine codon upstream of the start of the presumptive mature protein, yielding a putative leader sequence of 32 amino acids. The sequence following the leader matches exactly the amino acid sequence obtained from the NH_2-terminus of human urinary CSF-1. Including the first amino acid of the NH_2-terminus, the sequence extends for 224 amino acids, which gives a molecular size of about 26,000 daltons for the mature protein. There are two potential *N*-linked glycosylation sites of the Asn-X-Ser/Thr variety in the mature protein, and these are underlined in Fig. 5.

Analysis of the inactive clone, pcCSF-12, revealed that the clone is 102 bases shorter at the 5′ end and 424 bases shorter at the 3′ end. The truncation at the 3′ end is due to the fact that the polyadenylation signal for pcCSF-12 is different from that for pcCSF-17. However, neither clone contains an exact match to the consensus sequence AATAAA (*30*), preceding the polyadenylation site. More interestingly, clone pcCSF-12 contains 115 bases of the 3′ end of the intron that separates the two exons in the CSF-1 genomic clone described in Fig. 2B. The 5′ end of this intron sequence is preceded by an acceptor splice sequence in the genomic clone (Fig. 2B), and therefore it appears that the mRNA counterparts of clones pcCSF-12 and pcCSF-17 were derived from mRNA precursors that were differentially spliced. Clone pcCSF-12 contains the entire coding sequence for the mature CSF-1 protein, but is inactive in the COS cell assay because there is a stop codon immediately after the beginning of the intron sequence. This results in a severely truncated protein of only 23 amino acids excluding the leader peptide. Using an oligonucleotide probe derived from the intron sequence present in pcCSF-12, we have found that the re-

maining eight clones that were inactive in the COS cell assay also contained the intron sequence (*31*). Clone pcCSF-12 also differs from pcCSF-17 in one amino acid, at position 59 of the mature protein (asterisk in Fig. 5) where Asp was found instead of Tyr. Since Asp was also found in the coding sequence of the genomic clone (*32*), the Tyr found in pcCSF-17 may be due to a reverse transcriptase error or a natural polymorphism. In any case, the single amino acid change was not critical to the function of the protein, since pcCSF-17 expressed biologically active CSF-1 in the transient assay system.

The bone marrow proliferation assay used to test the sucrose gradient fractionated mRNA is not specific for CSF-1. The radioreceptor and bone marrow colony formation assays were not used for this purpose because we could not detect CSF-1 activity from oocytes with either of these two methods. Thus, the activity found in the MIA-PaCa mRNA may have been due to GM-CSF as well as CSF-1, since both factors are produced by this cell line (*22*). Therefore, we could not be sure whether the CSF-1 cDNA clones originated from mRNA in

Table 1. Expression of CSF-1 cDNA in COS-7 cells. Plasmid DNA's were purified by CsCl banding and transfected into COS-7 cells by a modification of the calcium phosphate coprecipitation technique (*48, 49*). Medium from the transfected cells was removed after 72 hours of incubation and tested for CSF-1 in the RRA. Pooled culture fluids from three transfected cultures were used as a source for further CSF-1 characterization. The RRA was performed, except for slight modification, as described (*2*). The incorporation of [^{3}H]thymidine by mouse bone marrow cells was assayed as described (*25, 26*). Bone marrow colony formation (*12*) was measured (one unit = one colony) with mononuclear bone marrow cells from BALB/c mice. The cells were incubated for 7 days in Alpha medium containing 0.3 percent agar (10^5 cells in 1 ml plus 0.1 ml of CSF, in six well plates). Colonies containing more than 50 cells were counted, removed from the agar individually, cytocentrifuged, and stained. A neutralizing antiserum (R52) to human urinary CSF-1 (*2*) was used to test the CSF made by the COS cells. The antiserum (ammonium sulfate purified IgG) was mixed with samples and incubated at 37°C for 1 hour before the cells were added. The final concentration of the IgG fraction in the assays was 5 percent. Partially purified human urinary and MIA-PaCa CSF-1, and a source of predominantly GM-CSF (GCT-Gibco), were used as standards in the above assays. R52 antibody had little effect on GCT GM-CSF activity. Abbreviations: RRA, radioreceptor assay; BM, bone marrow; M, macrophage; GM, granulocyte-macrophage; NRG, normal rabbit IgG; R52, IgG antibody to human urinary CSF-1; GCT, giant cell tumor tissue culture medium (Gibco), which contained the GM-CSF; N.D., not determined.

Sample	RRA (unit/ml)	BM assay (unit/ml)		Colony morphology (%)	
		Proliferation	Colony	M	GM
pcCSF-17	2433	2358	4287	85	15
pcCSF-17 + NRG	N.D.	3693	4920		
pcCSF-17 + R52	N.D.	424	0		
pcCSF-12	<10	<15	N.D.		
MIA-PaCa	2000	2000	2562	94	6
GCT	350	N.D.	1480	15	85
NRG	N.D.	<15	2		
R52	N.D.	30	0		

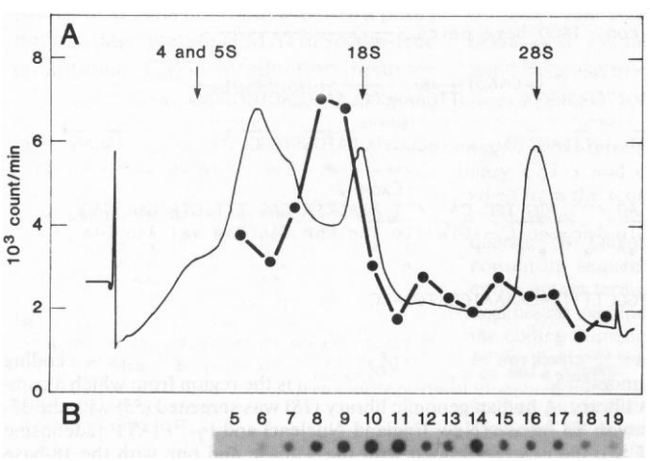
Fig. 2. Isolation and partial characterization of human CSF-1 genomic clone. (A) Restriction map of Hind III subclone. The location of the NH_2-terminus of the mature protein is indicated (B, Bam HI; E, Eco RI; H, Hind III; P, Pst I). (B) Partial sequence analysis of the genomic clone. The assignment of the coding regions was chosen by comparison with the NH_2-terminal sequence of murine and human CSF-1, and confirmed by analysis of the CSF-1 cDNA clone. The murine NH_2-terminus is shown below the human sequence, with identical residues indicated by the asterisks. The intron-exon boundaries are indicated by arrows. The overlined region in the intron defines the 115 bp of sequence interrupting the normal coding sequence of one of the CSF-1 cDNA clones, pcCSF-12 (Figs. 4 and 5). The underlined portion in the second exon is the region from which a complementary 32-base probe (exon 2 probe) was made for screening the cDNA library. A human genomic library (*18*) was screened (*53*) with the 35-base and 18-base oligomers (Fig. 1B). The oligomers were 5′ end-labeled with T4 kinase (New England Nuclear) and [γ-^{32}P]ATP (adenosine triphosphate) (New England Nuclear). Plates were lifted in triplicate, and two filters were probed with the 35-base and one with the 18-base probe. Hybridization conditions for the 35-base probe were 5× SSC (standard saline citrate), 5× Denhardt's (*54*), 15 percent formamide, 10 m*M* sodium phosphate (*p*H 6.5), 2 m*M* EDTA, sheared, denatured salmon sperm DNA (100 μg/ml), and 0.2 percent sodium dodecyl sulfate (SDS). The filters were hybridized for 20 hours at 42°C, the probe contained 1×10^6 cpm/ml. The 18-base probe hybridization conditions were the same, except that the formamide was omitted. The filters were washed 3× in 2× SSC and 0.1 percent SDS at 50°C for the 35-base and at 42°C for the 18-base probe. The filters were then autoradiographed at −70°C (Kodak XAR film and DuPont Quanta III intensifying screens). DNA was isolated from plaque-purified phage that hybridized to both probes and subjected to restriction enzyme digestion followed by Southern blot (*19*) analysis with the same oligonucleotides as probes. A 3.8-kb Hind III fragment that hybridized strongly to the 35-base oligomer was subcloned into M13mp19 (*20*), and sequenced by the dideoxy chain-termination method (*21*), with the 35-base probe as a primer. To obtain sequence through the 35-base probe hybridization site, an oligonucleotide primer was synthesized according to the upstream sequence. This enabled us to determine the nucleotide sequence encoding the NH_2-terminus. Analysis of Pst I fragments, subcloned into M13mp19, was used to confirm the sequence through this area. Murine L-cell CSF-1 was purified (*51*), and 1 nmole of the purified protein was sequenced (Beckman model 890C spinning cup sequencer); 30 amino acid residues were identified. The Cys residue at position 7 was identified by labeling of the protein by carboxymethylation with [^{14}C]iodoacetic acid, followed by analysis (Applied Biosystems model 470A gas-phase sequenator) (*15*).

the strongly hybridizing fraction, or the peak activity fractions which hybridized more weakly to the oligonucleotide probe. In order to distinguish between the two possibilities, we used the technique of oligonucleotide mRNA hybrid-arrest in *Xenopus* oocytes (*33*) to try to inhibit the activity in the peak activity fractions. Two complementary oligonucleotides, specific for CSF-1 and GM-CSF (*10*), were coinjected separately or in combination into oocytes with the mRNA from these fractions. Since both oligonucleotides failed to completely block the biological expression of the mRNA (*34*), it appears that neither CSF-1 nor GM-CSF mRNA is responsible for the total biological activity in these fractions.

Characterization of CSF-1 produced in COS-7 cells. Although pcCSF-17 contained what appeared to be the entire CSF-1 coding sequence, it was necessary to demonstrate that the sequence, when expressed, would exhibit the biological activities characteristic of authentic CSF-1 protein. To this end, we used several criteria to characterize the protein synthesized during transient expression in the COS cell system. (i) The protein must compete with purified, labeled murine CSF-1 in the specific radioreceptor assay (RRA) (*2*). (ii) The protein must stimulate bone marrow proliferation (*2, 25, 26*), and the colonies formed must be predominantly of monocytic lineage. (iii) The biological activity of the protein must be inhibited by neutralizing antiserum to human urinary CSF-1 (*2*). Without exception, all of the above criteria were satisfied (Table 1). In both the RRA and bone marrow proliferation assays, more than 2000 units per milliliter were produced by the COS cells when transfected with pcCSF-17 plasmid DNA. In contrast, pcCSF-12 gave less than 10 unit/ml and less than 15 unit/ml in the RRA and bone marrow assays, respectively. Examination of the colonies formed by the pcCSF-17 transfection showed that ~85 percent were of mononuclear morphology, with the remainder being of the mixed granulocyte-macrophage type. Finally, the immunoglobulin G (IgG) fraction of rabbit antiserum (R52) to human urinary CSF-1 completely blocked colony formation, whereas the control normal rabbit IgG (NRG) had no effect. As controls, the supernatants from MIA-PaCa and GCT (giant cell tumor; Gibco) tissue cultures were used to demonstrate the difference in types of colonies formed by CSF-1 and GM-CSF. Taken together, these data provide convincing evidence that the cloned sequence encodes authentic CSF-1.

Analysis of CSF-1 mRNA and its gene. To determine the size of CSF-1 mRNA, several different MIA-PaCa mRNA preparations were fractionated on a denaturing formaldehyde gel (*35*), transferred to nitrocellulose (*36*), and probed with the 32-base exon 2 oligomer. Surprisingly, several distinct species of mRNA were detected (Fig. 6A), ranging in size from ~1.5 to 4.5 kb. Since the hybridization conditions were stringent [5× SSC (saline, sodium citrate), 35 percent formamide, 45°C], we do not believe the presence of multiple species was an artifact. The largest mRNA was detected in all samples, even when isolated from cells not producing any detectable CSF-1 (compare lane 1 with lane 2 in Fig. 6A).

Because there were so many species of CSF-1 mRNA's, there existed the possibility that they were transcribed from a family of CSF-1 genes. To test this possibility, nuclear DNA was isolated from MIA-PaCa cells, digested with several restriction endonucleases, fractionated on an agarose gel, transferred to nitrocellulose (*19*) and probed with nick-translated pcCSF-17. The results (Fig. 6B) are consistent with the interpretation that CSF-1 is encoded by a single-copy gene. The number and size of the bands hybridizing to the Bam HI, Eco RI, Hind III, and Pst I digests (lanes 1 to 4, respectively) correspond reasonably well with the data obtained from several bacteriophage lambda clones isolated from the human genomic library. Several faint bands are visible in the autoradiogram which may represent partial enzyme digestion or hybridization to distantly related sequences. However, if two or more closely related genes exist in the genome, their sequences would have to be highly homologous over many



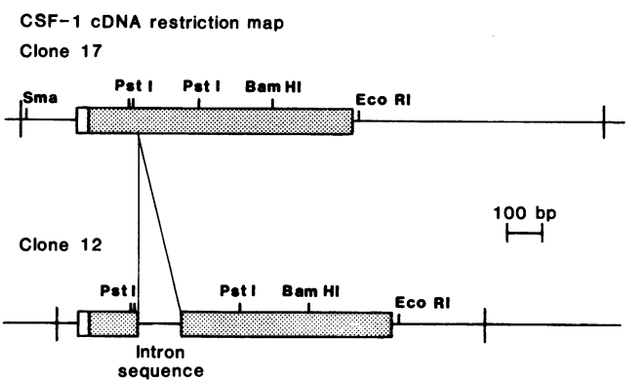


Fig. 3 (left). Sucrose gradient fractionation of MIA-PaCa mRNA. (A) The broken line indicates the bone marrow proliferation assay of supernatants from *Xenopus* oocytes injected with fractionated mRNA. (B) RNA dot blot of fractions. MIA-PaCa cells were grown to confluency in roller bottles in Dulbecco's modified Eagle's medium containing 10 percent fetal bovine serum (FBS). The medium was removed and replaced with fresh medium without FBS, but containing phorbol myristic acetate (PMA) (100 ng/ml). Cells were harvested for isolation of mRNA 24 to 48 hours later. Without PMA, there was no detectable CSF-1 in the medium during this incubation period, whereas PMA-induced cells produced 1000 to 2000 unit/ml. Total cytoplasmic RNA was isolated from MIA-PaCa cells by lysis of the cells in isotonic buffer with 0.5 percent NP-40 in the presence of ribonucleoside vanadyl complex (*55*), and subsequent extraction of the postnuclear fraction with phenol-chloroform and then ethanol precipitations. Polyadenylated RNA was isolated by oligo d(T) chromatography (*56*). The mRNA was fractionated on 5 to 20 percent (by weight) sucrose gradients at 27,000 rev/min for 17 hours at 20°C (Beckman SW40 rotor). Portions (50 nl) of the fractions were injected into *Xenopus* oocytes (*24*), and the medium was assayed for bone marrow proliferative activity 48 hours later essentially as described (*25, 26*). Portions (5 μl) were also dot blotted (*27*) onto nitrocellulose filters and probed with exon 2 oligonucleotide that had been labeled with ^{32}P by T4 kinase. The hybridization conditions were similar to those described in Fig. 2, except that the formamide concentration was 35 percent and the hybridization temperature was 45°C. Fig. 4 (right). Restriction map of CSF-1 clones pcCSF-17 and pcCSF-12. The limits of the coding sequences are defined by the boxed regions with the shaded portion representing the presumptive mature protein. The area labeled "intron sequence" in pcCSF-12 contains 115 bp of the CSF-1 genomic clone intron sequence as described in Fig. 2B.

kilobases to give the simple restriction pattern obtained here. In addition, the only restriction pattern difference found in the genomic clone isolates was in the position of the Eco RI cloning sites, giving further evidence for the existence of only one CSF-1 gene.

Further analysis of CSF-1. We have identified a cDNA clone that expresses biologically active CSF-1 when transfected into COS cells. Sequence analysis showed that the mature protein is 224 amino acids long with a putative leader peptide of 32 amino acids. An exact match was found to the 12 amino acids of the NH_2-terminus of human urinary CSF-1 protein, and the predicted amino acid sequence was also highly homologous to the sequence of murine L-cell CSF-1 (Fig. 2B) (*15*). After this manuscript was submitted, the first 25 amino acids of the amino terminus of murine L-cell CSF-1 was reported (*37*), and, in agreement, this sequence is also quite similar to the human CSF-1 sequence.

It was of great interest to determine whether human CSF-1 contained any homology to other known proteins. For this purpose, the predicted amino acid sequence of CSF-1 was compared to all the sequences recorded in the National Protein Information Resource protein sequence database (version 5.0, released 17 May 1985). In addition, the CSF-1 sequence was compared in more detail to the recently published sequences for other hematopoietic factors, including human GM-CSF (*10*), erythropoietin (*38*), erythroid-potentiating activity factor (*39*), interleukin-1α and -1β (*40*), and murine interleukin-3 (*7, 8*). We found no significant homology to any protein in the database, including the ones noted above, when the SEARCH program (*41, 42*) was used.

During our study, we found several interesting features of the human CSF-1 system. Although CSF-1 appears to be encoded by a single gene, several CSF-1 mRNA transcripts of different sizes are synthesized by MIA-PaCa cells (Fig. 6A). Part of the mRNA size heterogeneity can be accounted for by the fact that at least two of the transcripts have different polyadenylation signals. This was shown by the analysis of the structures of pcCSF-17 and pcCSF-12 (Fig. 4) in

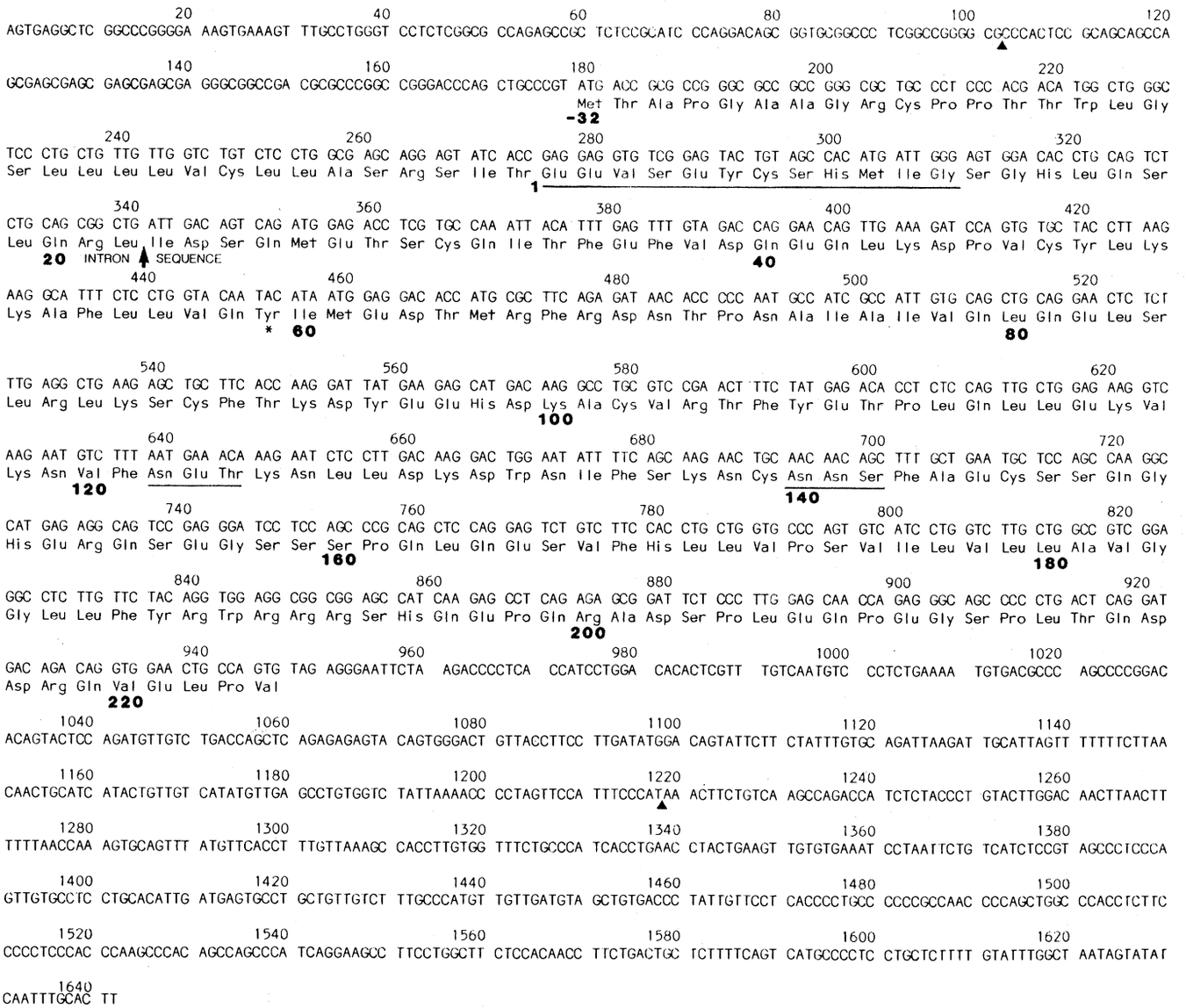


Fig. 5. DNA sequence and deduced amino acid sequence of CSF-1 clone pcCSF-17. The 12 amino acids found at the NH_2-terminus of urinary CSF-1 are underlined. The start of the mature protein is designated as amino acid number 1. Two potential *N*-linked glycosylation sites in the amino acid sequence are also underlined. The arrow labeled "intron sequence" indicates the position where 115 bases of intron sequence were found in pcCSF-12. The filled triangles delineate the boundaries of clone pcCSF-12 relative to pcCSF-17. Fractions from the sucrose gradient were used to synthesize a cDNA bank as described (*28, 49*). Candidate clones were obtained by screening colonies with the exon 2 probe described in Fig. 2B. The hybridization conditions were the same as described for the RNA dot blots in Fig. 3. CsCl purified plasmid DNA was prepared from the candidate clones, and restriction fragments were subcloned into M13mp18 (*20*) and sequenced (*21*).

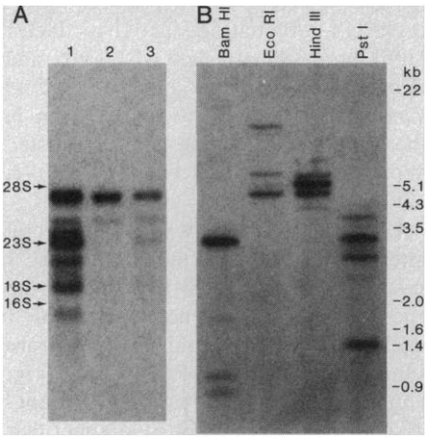


Fig. 6. Analyses of CSF-1 mRNA and DNA from MIA-PaCa cells. (A) Northern gel analysis of MIA-PaCa mRNA. (Lane 1) Messenger RNA from cells induced with PMA for 48 hours; (lane 2) mRNA from cells incubated for 48 hours without PMA; (lane 3) mRNA from cells incubated for 24 hours with PMA. Markers are mammalian and *Escherichia coli* ribosomal RNA's. (B) Southern gel analysis of MIA-PaCa DNA digested with restriction enzymes. (Lane 1) Bam HI digest; (lane 2) Eco RI; (lane 3) Hind III; (lane 4) Pst I. Markers are lambda DNA digested with Eco RI and Hind III. The mRNA was isolated from cells incubated without fetal bovine serum and with or without PMA as described in the legend of Fig. 3. Each mRNA sample (10 μg) was subjected to electrophoresis on a 1 percent agarose gel containing 2.2*M* formaldehyde (*35*) and transferred to a nitrocellulose filter (*36*). The blot was hybridized to the exon 2 probe described for the RNA dot blots in Fig. 3. High molecular weight DNA was isolated from MIA-PaCa cell nuclei essentially as described (*57*). Restriction enzyme digests of the genomic DNA (10 μg in each lane) were subjected to electrophoresis on a 1 percent agarose gel and transferred to nitrocellulose (*19*). The filter was hybridized with nick-translated pcCSF-17 (2×10^6 cpm/ml) at 42°C for 20 hours. The buffer was 5× SSC, 5× Denhardt's solution, 20 m*M* sodium phosphate (*p*H 7.0), sheared, denatured salmon sperm DNA (200 μg/ml), and 0.2 percent SDS. The filters were washed with 2× SSC containing 0.1 percent SDS at room temperature and 42°C. Final washings were done with 0.1× SSC containing 0.1 percent SDS at 50°C.

which the beginning of polyadenylation in clone 17 was 424 bases downstream from that found in clone 12. This phenomenon has been found in other systems such as mouse dihydrofolate reductase (*43*), but the biological significance of this occurrence remains unclear. A more intriguing finding was the presence of part of an intron sequence in nine out of the ten CSF-1 clones isolated. Since only part of an intron sequence was found, these aberrant sequences were probably due to incorrect splicing events and not simply to incomplete splicing. Also, because most of the clones contained the intron sequence, we do not believe their presence was due to an artifact such as nuclear leakage during cell lysis. A similar aberrant splicing event was reported for a cDNA clone of erythropoietin (*38*). The reason why a cell should synthesize such presumably inactive mRNA's is unknown, unless it is simply an artifact of PMA induction or an as yet unknown form of gene regulation.

The calculated molecular size of 26 kD for the mature CSF-1 protein is much larger than the reported 14.5 kD for the deglycosylated subunits of the 66-kD mouse or 45-kD human dimeric CSF-1 (*11*). Thus, it is possible that the mature protein may be formed by processing at both the NH_2-terminal and COOH-terminal ends, yielding a smaller protein than the one predicted from the DNA sequence. Of relevance to this possibility is the fact that the CSF-1 cDNA sequence predicts a very hydrophobic region of 23 amino acids (residues 166 to 188), followed by the amino acids Arg-Trp-Arg-Arg-Arg. This is a feature of many membrane proteins that have a transmembrane hydrophobic domain followed by three positively charged residues located on the cytoplasmic side (*44*). A proteolytic cleavage might process the protein to allow release of the secreted CSF-1 molecule. Previous studies have suggested that membrane and secretory forms of CSF-1 may exist (*45*, *46*).

Using the clones described above, we can now sort out these and other possibilities in the molecular and cell biology of CSF-1. In addition, large quantities of purified CSF-1 protein can now be produced for possible therapeutic use in cases of macrophage deficiencies. Finally, the recombinant CSF-1 protein can be used to study certain aspects of oncogenesis, in light of the recent finding that the product of the proto-oncogene, c-fms, may be identical to the CSF-1 receptor (*47*).

References and Notes

1. D. Metcalf, *The Haemopoietic Colony Stimulating Factors* (Elsevier, Amsterdam, 1984).
2. S. K. Das, E. R. Stanley, L. J. Guilbert, L. W. Forman, *Blood* **58**, 630 (1981).
3. N. A. Nicola, D. Metcalf, G. R. Johnson, A. W. Burgess, *ibid.* **54**, 614 (1979).
4. J. N. Ihle, L. Rebar, J. Keller, J. C. Lee, A. J. Hapel, *Immunol. Rev.* **63**, 5 (1982).
5. D. Metcalf, in *Normal and Neoplastic Haematopoiesis*, D. W. Golde and P. A. Marks, Eds. (Liss, New York, 1983), pp. 141–156.
6. K. Welte *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1526 (1985).
7. M. C. Fung *et al.*, *Nature* (*London*) **307**, 233 (1984).
8. T. Yokota *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1070 (1984).
9. N. M. Gough *et al.*, *Nature* (*London*) **309**, 763 (1984).
10. G. G. Wong *et al.*, *Science* **228**, 810 (1985).
11. S. K. Das and E. R. Stanley, *J. Biol. Chem.* **257**, 13679 (1982).
12. E. R. Stanley, in *The Lymphokines*, W. E. Stewart II and J. W. Hadden, Eds. (Humana Press, Clifton, N.J., 1981), pp. 102–132.
13. P. V. Byrne, L. J. Guilbert, E. R. Stanley, *J. Cell Biol.* **91**, 848 (1981).
14. E. R. Stanley and P. M. Heard, *J. Biol. Chem.* **252**, 4305 (1977).
15. A. Boosman, J. E. Strickler, K. J. Wilson, E. R. Stanley, in preparation.
16. R. M. Hewick, M. E. Hunkapiller, L. E. Hood, W. J. Dreyer, *J. Biol. Chem.* **256**, 7990 (1981).
17. H. R. Chen *et al.*, *DNA* **1**, 365 (1982).
18. T. Maniatis *et al.*, *Cell* **15**, 687 (1978).
19. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
20. C. Yanisch-Perron, J. Vieira, J. Messing, *Gene* **33**, 103 (1985).
21. F. Sanger, S. Nicklen, A. R. Coulsen, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
22. M-C. Wu, J. K. Cini, A. A. Yunis, *J. Biol. Chem.* **254**, 6226 (1979).
23. P. Ralph *et al.*, in preparation.
24. J. B. Gurdon, C. D. Lane, H. R. Woodland, G. Marbaix, *Nature* (*London*) **233**, 177 (1971).
25. R. N. Moore and B. T. Rouse, *J. Immunol.* **131**, 2374 (1983).
26. M. B. Prystowsky, M. F. Naujokas, J. N. Ihle, E. Goldwasser, F. W. Fitch, *Am. J. Pathol.* **114**, 149 (1984).
27. B. A. White and B. C. Bancroft, *J. Biol. Chem.* **257**, 8569 (1982).
28. H. Okayama and P. Berg, *Mol. Cell. Biol.* **2**, 161 (1982).
29. Y. Gluzman, *Cell* **23**, 175 (1981).
30. N. J. Proudfoot and G. G. Brownlee, *Nature* (*London*) **263**, 211 (1976).
31. E. S. Kawasaki, unpublished results.
32. M. B. Ladner *et al.*, in preparation.
33. E. S. Kawasaki, *Nucleic Acids Res.* **13**, 4991 (1985).
34. ———, unpublished results.
35. H. Lehrach, D. Diamond, J. M. Wozney, H. Boedtker, *Biochemistry* **16**, 4743 (1977).
36. P. S. Thomas, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5201 (1980).
37. C. H. Ben-Avram *et al.*, *ibid.* **82**, 4486 (1985).
38. K. Jacobs *et al.*, *Nature* (*London*) **313**, 806 (1985).
39. J. C. Gasson *et al.*, *ibid.* **315**, 768 (1985).
40. C. J. March *et al.*, *ibid.*, p. 641.
41. R. M. Schwartz and M. O. Dayhoff, in *Protein Segment Dictionary 1978: Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, 1978), pp. 353–358.
42. M. O. Dayhoff, R. M. Schwartz, B. C. Orcutt, *ibid.* pp. 345–352.
43. D. R. Setzer, M. McGrogan, R. T. Schimke, *J. Biol. Chem.* **257**, 5143 (1982).
44. D. D. Sabatini, G. Kreibich, T. Morimoto, M. Adesnik, *J. Cell Biol.* **91**, 1 (1982).
45. M. Cifone and V. Defendi, *Nature (London)* **252**, 151 (1974).
46. E. R. Stanley, M. Cifone, P. M. Heard, V. Defendi, *J. Exp. Med.* **143**, 35 (1976).
47. C. J. Scherr *et al.*, *Cell* **41**, 665 (1985).
48. F. L. Graham and A. J. van der Eb, *J. Virol.* **52**, 456 (1973).
49. A. M. Wang *et al.*, *Science* **228**, 149 (1985).
50. S. Urieli-Shoral, M. Furth, E. R. Stanley, in preparation.
51. E. R. Stanley, *Methods Enzymol.* **116**, 564 (1985).
52. S. L. Beaucage and M. H. Caruthers, *Tetrahedron Lett.* **22**, 1859 (1981).
53. W. D. Benton and R. W. Davis, *Science* **196**, 180 (1977).
54. D. T. Denhardt, *Biochem. Biophys. Res. Commun.* **23**, 641 (1966).
55. S. L. Berger and C. S. Birkenmeier, *Biochemistry* **18**, 5143 (1979).
56. J. Aviv and P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1408 (1972).
57. N. Blin and D. W. Stafford, *Nucleic Acids Res.* **3**, 2303 (1976).

58. We thank M. Doyle, F. McCormick, and M. M. McGrogan for initial studies on CSF-1; A. Boosman and J. Strickler for additional amino acid sequence data; the Cetus DNA synthesis group for oligonucleotides; E. Aragon, J. Csejtey, R. Milley, M. Nikoloff, A. Sampson-Johannes, R. Tal, J. Weaver, F. Pellerito, and C. Saez for technical assistance; J. Larrick and T. White for advice; M. Innis for helpful discussion on the manuscript and advice on isolating genomic clones and DNA sequencing strategies; E. McCallan, S. Tickton, E. Ladner, S. Nilson, and T. Culp for help in manuscript preparation. This work was supported in part by NIH grant C32551 (E.R.S.).

22 July 1985; accepted 16 September 1985