The EMBO Journal vol.6 no.9 pp.2693–2698, 1987

Human CSF-1: gene structure and alternative splicing of mRNA precursors

Martha B.Ladner, George A.Martin, Janelle A.Noble, D.Michele Nikoloff, Rony Tal[1], Ernest S.Kawasaki and Thomas J.White

Departments of Molecular Biology and [1]Microbial Genetics, Cetus Corporation, 1400 Fifty-Third Street, Emeryville, CA 94608, USA

Communicated by J.E.Davies

Bone marrow progenitor cells differentiate into mononuclear phagocytes in the presence of colony stimulating factor-1 (CSF-1). Characterization of the human CSF-1 gene shows that it contains 10 exons and 9 introns, which span 20 kb. Analysis of multiple CSF-1 transcripts demonstrates that alternate use of exon 6 splice acceptor sites and 3′ noncoding sequence exons occurs. These alternatively spliced transcripts can encode either a 224 or a 522 amino acid CSF-1. Implications of differential splicing for the production and function of CSF-1 are discussed.

Key words: alternative splicing/cDNA sequence/lymphokine/promoter sequence

Introduction

Production of hematopoietic cells is regulated by a group of proteins known collectively as colony stimulating factors (CSF) (Metcalf, 1986). cDNAs encoding four distinct human CSF proteins have been cloned. These include granulocyte-CSF (G-CSF) (Nagata *et al.*, 1986) granulocyte-macrophage CSF (GM-CSF) (Wong *et al.*, 1985) macrophage CSF (CSF-1) (Kawasaki *et al.*, 1985) and multi-CSF (IL-3) (Yang *et al.*, 1986).

CSF-1 supports the growth and proliferation of mononuclear progenitor cells and promotes the proliferation of mature macrophages. Several different CSF-1 transcripts are produced by a human pancreatic carcinoma cell line, (Das and Stanley, 1982; Das *et al.*, 1981) MIA PaCa-2, after PMA induction under serum free conditions (Kawasaki *et al.*, 1985). We initially isolated a 1600 bp CSF-1 cDNA (pcCSF-17) from these cells (Kawasaki *et al.*, 1985). Recently, a second CSF-1 cDNA (p3ACSF-69) was isolated from a human SV40 transformed trophoblast cell line which codes for a 522 amino acid form of CSF-1 (Wong *et al.*, 1987). This longer CSF-1 is homologous to the CSF-1 encoded by cDNA clones we have isolated from murine L929 cells (Ladner *et al.*, in preparation). Our present results show that Mia PaCa-2 cells also produce CSF-1 transcripts encoding the 522 amino acid form of CSF-1, and that a cDNA clone corresponding to this human mRNA expresses active CSF-1 protein in COS cells.

Characterization of the genomic clone and Northern blot analysis of the CSF-1 transcripts by oligonucleotide hybridization have enabled us to analyze the CSF-1 mRNAs produced in MIA PaCa cells. This report describes the organization of the human CSF-1 gene and shows that the diverse sequences described for CSF-1 mRNAs arise from differential splicing of a large primary transcript.

Results

Genomic organization

The human CSF-1 gene (λC and λK in Figure 1) was isolated from a genomic library (Maniatis *et al.*, 1978) using a degenerate, 35-base oligonucleotide probe complementary to the amino terminal protein sequence of human urinary CSF-1 (Kawasaki *et al.*, 1985). An additional clone (λ11), which overlaps λC, was subsequently isolated using a cDNA probe. Comparison of the known CSF-1 cDNA sequence with the sequence of genomic subclones enabled us to elucidate the structure of the CSF-1 gene.

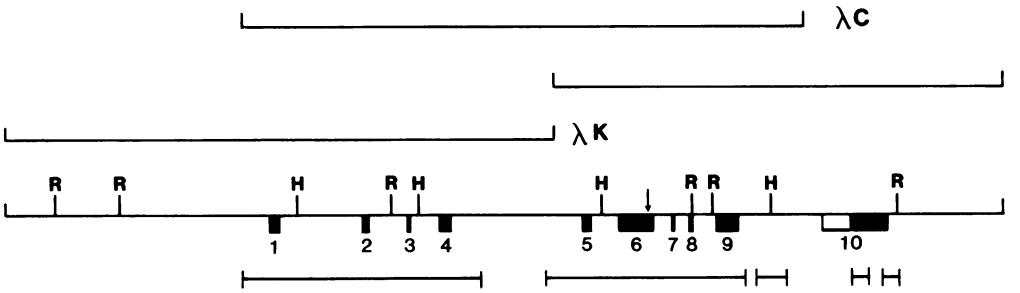
The CSF-1 gene is divided into at least 10 exons (Figure 1). The first exon contains the 5′ untranslated (UT) sequence and the first 13 amino acid residues of the leader sequence. The remaining 19 amino acid residues of the leader sequence and the first 22 amino acids of the mature protein are contained within the second exon; the remaining coding sequence is divided into six exons. Exon 6 is variable in length, depending upon which of two splice acceptor sites is used. The translation stop signal (TAG) lies 16 bp upstream of the end of exon 8. Exon 9 contains the remaining 679 bp of 3′ UT sequence contained in pcCSF-17. We have determined that there is at least one additional exon ~2.5 kb downstream of exon 9 which codes for an alternate 3′ untranslated sequence.

5′ flanking sequence

The putative promoter region of the CSF-1 gene has several elements which may be involved in regulating CSF-1 transcription (Figure 2). The sequence CATAAA, which has homology with the TATA box (Goldberg and Hogness box) (Breathnach and Chambon, 1981), lies 54 bases upstream from the first nucleotide of our most 5′ cDNA. Although the initial T in the TATA box is strongly conserved in many genes, a compilation of 60 genes analyzed for sequence homology upstream of the mRNA start site shows that C is the second most common initial nucleotide (Breathnach and Chambon, 1981). Another possible TATA box, TTAAAA, 26 bp upstream from the cDNA start site in pcCSF-17, is similar to the sequence TTTAAA, which is present 28 bp upstream from the putative initiation site of RNA transcription for mGK-1 in the murine kallikrein gene (Mason *et al.*, 1983).

Sp1, the transcription factor required for SV40 transcription, has been shown to bind to GC boxes (GGGCGG) in both the SV40 and the mouse DHFR promoter. Sp1 can stimulate transcription from the DHFR promoter (Dynan *et al.*, 1986). Identical GC sequences have been found recently in several eucaryotic genes, including the epidermal growth factor receptor (Ishii *et al.*, 1985a), human adenosine deaminase (Valerio *et al.*, 1985) and the Harvey *ras* 1 proto-oncogene (Ishii *et al.*, 1985b). The hamster gene for HMG CoA reductase, a gene which is controlled by negative feedback, has both GGGCGG and CCGCCC sequences in its 5′ flanking region (Reynolds *et al.*, 1984).

The hexanucleotide, GGCGGG, is present twice in the 5′ flanking region of CSF-1 (positions −159 and −177). This same sequence is present five times between positions −167 and −214



EXON	SIZE (bp)	INTRON	SIZE (Kb)
1	217	I	3.1
2	123	II	1.3
3	63	III	1.3
4	171	IV	4.6
5	148	V	1.2 or 2.1
6	1025 or 131	VI	0.6
7	53	VII	0.3
8	56	VIII	0.8
9	679	IX	2.7
10	2Kb		

Fig. 1. Organization of the human CSF-1 gene. Three overlapping genomic clones containing the entire gene are indicated at the top of the figure. The ten exons are shown as closed boxes; the open box indicates the presumed location of the first half of exon 10. The location of the exon 6 5′ splice site used by pCSF-17-type transcripts is shown by an arrow. The map indicates *Eco*RI (R) and *Hin*dIII (H) restriction sites, and sequenced regions of the gene are delineated below.

```
-570 ctgcagaggaagaaggggctgccggcaaacctgctgactcaggctccacgagggagcaagtaacactggactcctttcggcactccgagaatggggtgggggcgtcttcaaaggatttc

-450 cctcccttcccagtgcttgtccctgctctcggtccgttttctgctaagatttggggattttcagggcctggagggaaagtcccttgggacgatcatagagcgctagcactgaatcagcct

-330 ggagagcgcggaaggaaagggtcggtccgcagagggcgcggggaaggcagggtggggacgcggtggagcccgcgctcgtttgctgaaggcttggaagtgcagcgcagaagacagaggtg

-210 actaggaagacgcgcgaacggggctggccggccggcgggtgggggaggggaggcgggggaaggcggctgagtgggcctctggag[tgtgtgtgtctgtgtcagtgtgtgtgtgtgtgtg

                                                                                                                    -1
 -90 tatgtgtgtg]tctggcgcctggccagggtgatttcccataaaccacatgccccccagtcctctcttaaaaggctgtgccgagggctggccAGTGAGGCTCGGCCCGGGGAAAGTGAAAGT

     TTGCCTGGGTCCTCTCGGCGCCAGAGCCGCTCTCCGCATCCCAGGACAGCGGTGCGGCCCTCGGCCGGGGCGCCCACTCCGCAGCAGCCAGCGAGCGAGCGAGCGAGCGAGGGCGGCCGA

     CGCGCCCGGCCCGGGACCCAGCTGCCCGTATGACCGCGCCGGGCGCCGCCGGGCGCTGCCCTCCCACGgtaag
                                  METThrAlaProGlyAlaAlaGlyArgCysProProThr
```

Fig. 2. Nucleotide sequence of the 5′ flanking region and the first exon of the human CSF-1 gene. cDNA sequence is indicated by capital letters. Two potential TATA sequences are boxed and the two GGCGGG hexanucleotide sequences are underlined. The sequences similar to the enhancer core consensus sequence are overlined. The long series of repeating thymine and guanine bases is enclosed in brackets.

in the human HPRT gene (Kim *et al.*, 1986). Because of the similarity to the canonical GC box, these hexanucleotides may also interact with transcription factors.

Two areas in the CSF-1 genomic sequence are homologous to enhancer core sequences. The consensus sequence, TGGAAAG, was derived from 14 known and putative enhancer sequences and compared with the core element of the SV40 72-bp repeat (Laimins *et al.*, 1983). Similar sequences (GGGAAAG and AGGAAAG) are located at positions −377 and −317 in the 5′ flanking region of the human CSF-1 gene.

Between positions −85 and −127 there is a series of alternating guanine and thymine bases which has the potential of forming Z-DNA. Rich *et al.* (1984) postulate that Z-DNA can have either positive or negative effects upon transcription, depending upon its location and surrounding sequences. Determination of which elements contribute to the expression of the CSF-1 gene requires further study.

Isolation of a new CSF-1 cDNA (CSF4)

pcCSF-17, the first CSF-1 cDNA isolated from MIA PaCa cells, encodes a 224-amino acid residue CSF-1 protein; however, the CSF-1 cDNA from murine L cells codes for a 520 amino acid residue protein (Ladner *et al.*, in preparation). The size difference between these clones results from 295 additional internal amino acids in the murine coding sequence beginning at the position corresponding to amino acid residue 150 of pcCSF-17 (Kawasaki *et al.*, 1985). Examination of our human CSF-1 genomic clone revealed an 894 bp DNA sequence immediately preceding the exon 6 acceptor site in pcCSF-17 which is 78% homologous to the additional murine cDNA sequence. Translation of this human sequence revealed an open reading frame with 80% homology to the 295 additional amino acid residues coded for by the murine sequence.

A 40 bp oligonucleotide probe (GM11), specific for this human genomic sequence, hybridized to induced MIA PaCa mRNA on

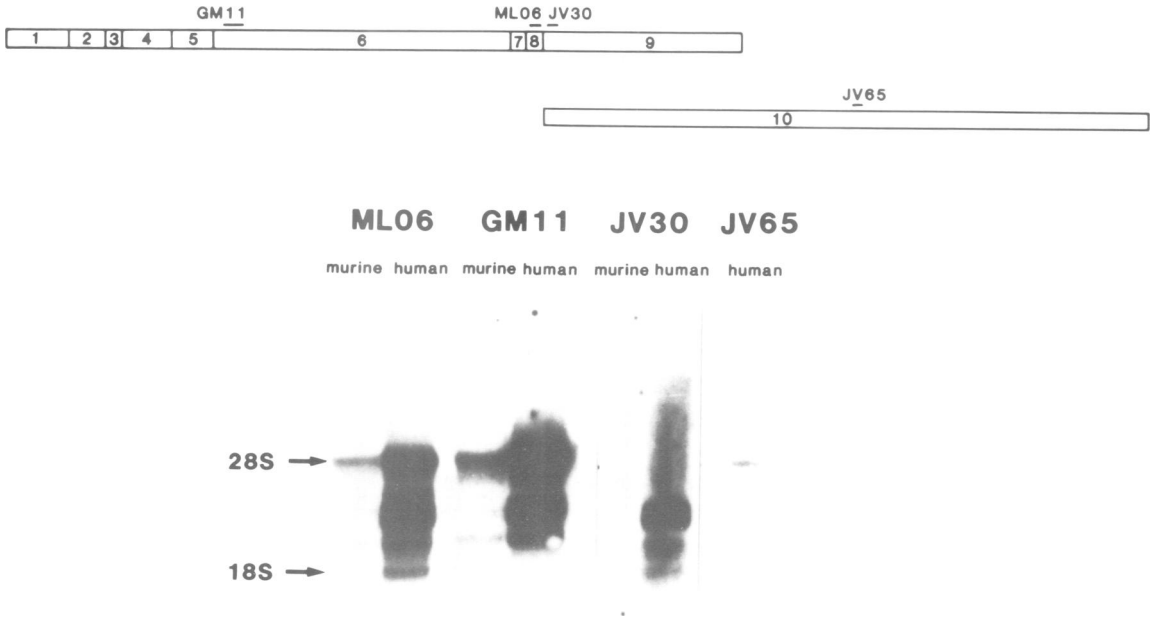


Fig. 3. Northern blots of murine L929 cell and PMA-induced MIA PaCa cell mRNA. ML06 is specific for exon 8; GM11 is specific for the long form of exon 6; JV30 is specific for exon 9 and JV65 is specific for the 3′ untranslated sequence in the murine 4 kb CSF-1 cDNA clone.

Northern blots (Figure 3), indicating that some CSF-1 transcripts contain sequence not present in pcCSF-17. GM11 was subsequently used to probe a λgt10 cDNA library made using total mRNA from induced MIA PaCa cells. Duplicate filters were hybridized with a pcCSF-17 cDNA probe. From the initial screening, 10 cDNA clones were isolated which hybridized to both the cDNA probe and GM11. The largest clone (CSF-4) which hybridized with GM11 was subcloned and sequenced. CSF-4 (Figure 4) begins at amino acid residue 19 of the pcCSF-17 leader sequence and codes for a 522 amino acid residue CSF-1 protein which contains a 298 amino acid residue insertion at amino acid residue 150 relative to pCSF-17. The protein has a predicted molecular weight of 60.5 kd and contains four asparagine-linked glycosylation signals.

Expression of CSF-4

A *Bst*XI to *Eco*RI restriction fragment from pcCSF-17 was replaced with the corresponding fragment from CSF-4 in the expression vector pcDB (Ladner *et al.*, in preparation). The new construct, pcDBCSF-4, which contains all of the leader sequence from pcCSF-17 and the long CSF-4 coding sequence, was used to transfect COS-A_2 cells. CSF-1 activity in the culture supernatants was measured by a human CSF-1 specific RIA (Das and Stanley, 1981) and a murine-bone-marrow colony assay (Kawasaki *et al.*, 1985). The results of these assays, shown in Table I, prove that both pcCSF-17 and pcDBCSF-4 encode active CSF-1.

CSF-1 mRNA heterogeneity in MIA PaCa cells

CSF-1 transcripts vary in size depending in part upon which form of exon 6 they contain. ML06, an exon 8-specific oligonucleotide probe, hybridizes to all CSF-1 transcripts on a Northern blot of mRNA from induced MIA PaCa cells (Figure 3). These transcripts range in size from 1.6 to 4.5 kb. Hybridization of the oligonucleotide GM11 (complementary to nucleotides 506 to 545 in CSF-4 in Figure 4) on the same Northern blot established that a subset of all CSF-1 transcripts contain the long form of exon 6. Clearly, GM11 hybridizes to the largest CSF-1 mRNA and to three mRNAs of intermediate size (~2.6–3.3 kb). The faint transcript (~3.6 kb) below the largest mRNA and the two smallest transcripts do not hybridize to GM11, indicating that these messages contain the short form of exon 6.

The two murine cDNA clones diverge 13 bp beyond the TAG stop codon and contain 3′ UT sequences that are vastly different in both size and composition (Ladner *et al.*, in preparation). The presence of an exon boundary 13 bp after the stop codon in the human CSF-1 sequence (Figure 4) suggested that the human CSF-1 mRNAs might also contain alternate 3′ UT sequences. Two lanes on the Northern blot in Figure 3 were probed with JV30, a 20-bp oligonucleotide which is specific for the pcCSF-17 3′ UT region. JV30 hybridizes to the five smaller human mRNAs and also to the small murine CSF-1 mRNA. It does not hybridize to the two largest human CSF-1 mRNAs or the larger of the two murine L929 cell mRNAs. These data confirm our hypothesis that human CSF-1 mRNAs contain different 3′ untranslated regions.

The 4.2-kb *Hin*dIII–*Eco*RI genomic fragment, which lies immediately downstream from exon 9 in λ11 (Figure 1), hybridizes to only the two largest CSF-1 transcripts (data not shown). This result established that at least a portion of the alternate 3′ sequence is conained in this *Hin*dIII–*Eco*RI fragment.

Several oligonucleotides specific for the 3′ UT sequence of the long murine cDNA were also tested in Northern analyses for hybridization to human CSF-1 mRNA. Many of these oligonucleotides do not hybridize to human mRNA. However, one of these murine-specific oligonucleotides, JV65, does hybridize faintly to the largest CSF-1 transcripts (Figure 3), suggesting that the human alternate 3′ UT sequence shows at least some homology with its murine analog. JV65 and its complement, JV66, were used successfully as sequencing primers on the 4.2 kb *Hin*dIII–*Eco*RI human genomic fragment described above. The

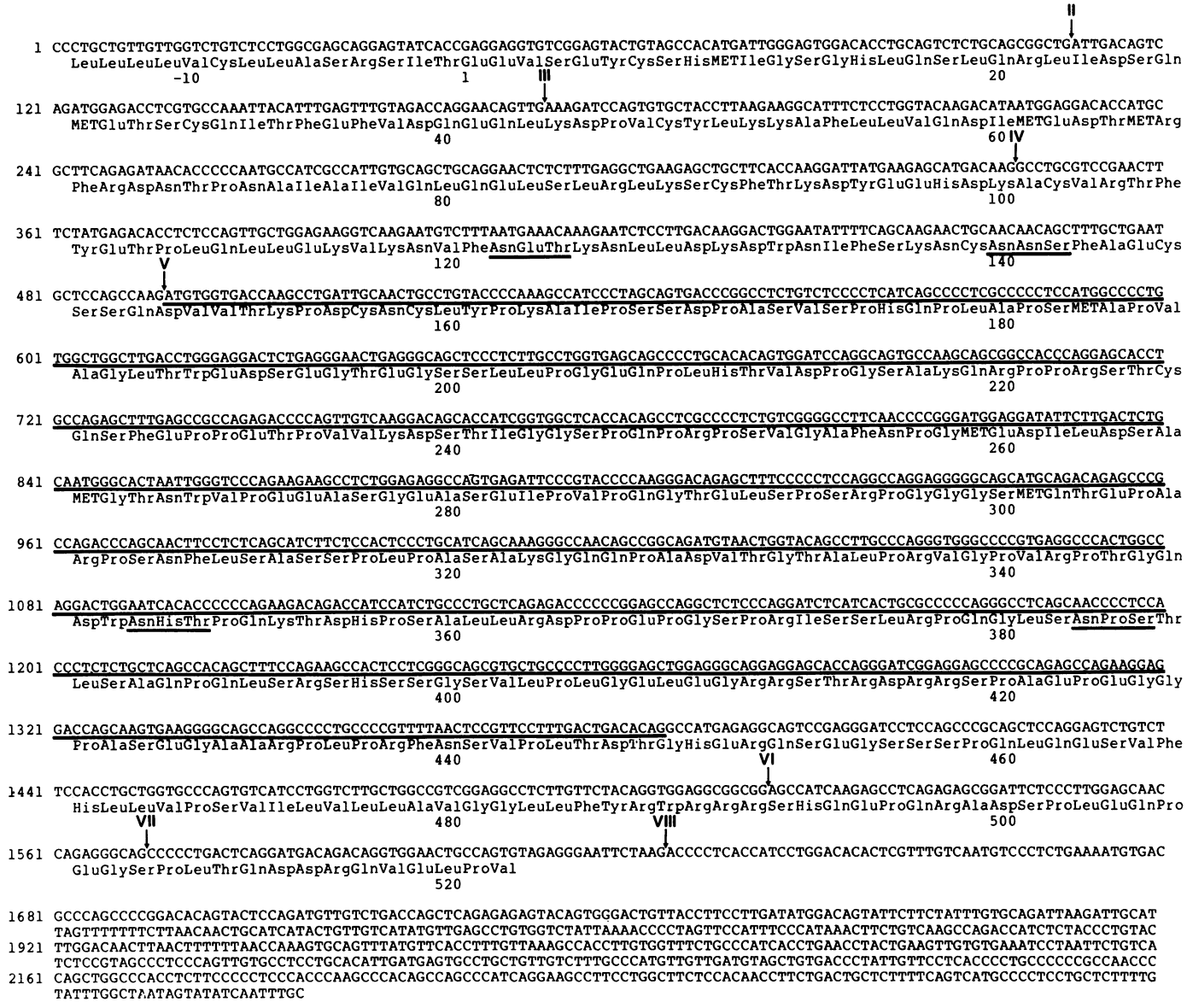


Fig. 4. DNA and amino acid sequence of CSF-4. The start of the mature protein is designated as amino acid 1. The 894 bp of sequence not present in pcCSF-17 are underlined. Arrows indicate intron/exon boundaries. The four asparagine-linked glycosylation signals in the amino acid sequence are underlined.

DNA sequence generated with JV65 and JV66 shows homologies of 78 and 69% respectively to the 3′ UT sequence from the 4-kb murine cDNA and contains a unique *SalI* restriction site which allowed us to localize it ~3 kb downstream from exon 9. Analysis of the 3′ end of the 4.2-kb fragment revealed a sequence with 93% homology to the last 54 nucleotides of the 4-kb murine clone. The position of this region with respect to the position of the JV65/JV66 generated sequence suggests that at least 1 kb of alternate 3′ UT sequence is encoded in a single exon in the position indicated in Figure 1. Further sequence analysis is necessary to determine whether the entire alternate 3′ UT sequence is contained in a single exon.

Table I. Expression of CSF-4 and CSF-17 in COS cells

Experiment	Clone	RIA (units/ml)	Mouse BM colony assay (units/ml)
I	pcDBCSF-4	17 236	9200
	pcDBCSF-17	31 757	17 000
	MEDIUM	5	0
	VECTOR	100	20
	-DNA	100	
II	pcDBCSF-4	29 900	33 667
	pcCSF-17	26 400	24 500

The two CSF-1 coding sequences inserted in the Okayama Berg expression vector, pcDB, are designated pcDBCSF-4 and pcDBCSF-17. Vector is pcDB without insert. pcCSF-17 is the original human CSF-1 cDNA clone.

Discussion

Induction of the human cell line MIA PaCa-2, under serum free conditions, results in an increased production of heterogeneous CSF-1 transcripts. The same transcript heterogeneity is present in uninduced MIA PaCa-2 cells grown in fetal-calf serum (Ralph *et al.*, 1986). The heterogeneity results from differential gene splicing at two different locations. One of these locations is between exons 5 and 6. Two different splice acceptor sites are utilized in exon 6. Use of the downstream acceptor site creates CSF-1 mRNAs with 672 bp of coding sequence, as in pcCSF-17. When

Table II. Sequence at the intron/exon boundaries of the human CSF-1 gene

5′ donor boundary		3′ acceptor boundary		
Exon	Intron		Exon	
1 CACG	gtaagc . . . I . . .	$[t/c]_8$ gtcacag	ACATG	2
2 GCTG	gtgagt . . . II . . .	$[t/c]_9$ acag	ATTGA	3
3 GTTG	gtgagt . . . III . . .	$[t/c]_5$ tcag	AAAGA	4
4 CAAG	gtagga . . . IV . . .	$[t/c]_7$ ccag	GCCTG	5
5 CAAG	gtaagc . . . V . . .	$[t/c]_5$ tcag	ATGTG	6 (upstream site)
	or	$[t/c]_7$ gactgacacag	GCCAT	6 (downstream site)
6 GCGG	gtgagt . . . VI . . .	$[t/c]_9$ acctcag	AGCCA	7
7 GCAG	gtgagag . . VII . .	$[t/c]_3$ gcag	CCCCC	8
8 TAAG	gtaagg . . VIII . .	$[t]_4$ tcag	ACCCT	9

Consensus:

CAG gtaagt	$[t/c]_n$ ncag G
A g	t
Donor	Acceptor

Capital letters indicate exon sequence and small letters indicate intron sequence. The consensus sequence is from Mount (1982).

the upstream acceptor site is used, the resultant mRNAs contain 1563 bp of coding sequence, as in CSF-4.

As shown in Table II, the donor and acceptor sites resemble the consensus sequences (Mount, 1982); however, the acceptor site used to create pcCSF-17 contains a very poor pyrimidine stretch at the end of the intron. The acceptor site used by CSF-4 has a long pyrimidine tract. The upstream acceptor site is used by the majority of CSF-1 transcripts, as shown by the hybridization of GM11 to four of the most abundant CSF-1 mRNAs (Figure 3).

The choice of exons used for the 3′ untranslated regions also results in transcript heterogeneity. At least two different exons are used for the 3′ untranslated region, just as in the murine CSF-1 mRNAs (Ladner *et al.*, in preparation). The 2.3-kb cDNA (CSF-4) we have sequenced contains the same 3′ untranslated sequence contained in pcCSF-17. The recently published sequence of p3ACSF-69, which contains the same coding sequence as CSF-4, contains an alternate 3′ untranslated sequence (Wong *et al.*, 1987). This 3′ UT sequence is 2.2 kb in length and is homologous to the exon 10 sequence we have determined, except at the extreme 3′ end. The last 20 bp of the p3ACSF-69 sequence and the corresponding genomic sequence are:

TGCCGTTGAC	GTAGCTCGAG	p3ACSF-69 (Wong *et al.*, 1987)
TGCCATGTGA	GTCGTGGGAA	genomic CSF-1

The difference is probably not due to an intron in the genomic sequence because there is no recognizable donor splice site at the position where the sequences diverge.

The fact that two species as evolutionarily divergent as mouse and human product CSF-1 mRNAs with two 3′ UT sequences suggests these sequences are functionally significant. Examination of these sequences shows that the mouse and human shorter 3′ UT sequences (found in pCSF-17, CSF-4, and the murine 2-kb clone) are related to one another, and the longer 3′ UT sequences [p3ACSF-69 (Wong *et al.*, 1987) and the murine 4-kb clone] are similar. The degree of conservation also suggests that these sequences have a function. Between mouse and human, the shorter 3′ UT sequence homology is 90% and the longer sequence homology is 70%. 110 bp of sequence at the extreme 3′ end of the longer sequence is 95% homologous between the species and contains several ATTTA sequences. This motif may be a recognition signal for the rapid degradation of certain mRNAs (Shaw and Kamen, 1986). This motif is not found in the shorter 3′ UT sequence and suggests that CSF-1 mRNA stability could be determined by which 3′ UT sequence is used.

Since the utilization of exon 6 acceptor sites and alternate 3′ exons appear to be independent events, several different CSF-1 mRNAs can be generated. The splicing pattern of CSF-1 mRNA from induced MIA PaCa-2 cells is related to the state of these cells. Serum-free, uninduced MIA PaCa-2 cells contain the highest mol. wt CSF-1 mRNA (Kawasaki *et al.*, 1985). However, the cells contain no detectable lower mol. wt CSF-1 RNAs and very little CSF-1 protein. The appearance of several CSF-1 mRNAs is correlated to the induction of these cells with PMA (with or without retinoic acid). However, PMA does not induce the formation of heterogeneous CSF-1 mRNAs in monocytes. Only one CSF-1 mRNA, of high mol. wt, is detectable in human monocytes induced with 20 ng/ml PMA (Horiguchi *et al.*, 1986).

RNA coding for G-CSF, on the other hand, is not differentially spliced in the PMA induced MIA PaCa cells (Devlin *et al.*, in press). In another cell line, Chu-2, G-CSF transcripts are present in two alternatively spliced forms (Nagata *et al.*, 1986). These results indicate that both the cellular components and the sequence of the mRNA influence mRNA splicing. As discussed by Padgett *et al.* (1986), in their recent review on splicing, *trans*-acting factors (such as different levels of specific snRNPs) may influence the activity of different splice sites. Reed and Maniatis (1986) have shown that altering the concentration of splicing extracts can change the relative use of duplicated splice sites and suggest that alternative splicing can be controlled by variations in the amounts of splicing components available.

CSF-1 protein structure

Our data suggest that the amino-terminal portion of CSF-1 is responsible for the protein's biological function while the amino acids in the sequence unique to CSF-4 may impart a correct configuration for proteolytic processing of the membrane bound protein.

All CSF-1 cDNA clones (human and murine) code for a 23 amino acid residue hydrophobic sequence which begins 59 amino acids before the carboxy-terminus (60 amino acids before the carboxy-terminus in murine CSF-1). This hydrophobic sequence has the characteristics of a transmembrane domain. We have previously hypothesized that CSF-1 is initially a membrane bound protein (Kawasaki *et al.*, 1985). After removal of this transmembrane domain the proteins encoded by pcCSF-17 and CSF-4 would have substantially different carboxy-termini domains and the amino-terminal 149 amino acid residue sequence would be the only common amino acid sequence shared between the two proteins. This hypothesis clearly suggests that the active site of CSF-1 resides within the first 149 amino acids.

Analysis of the amino acid homology between murine and human CSF-1 (CSF-4) shows there is more homology between the first half of the two proteins than through the central region of the sequence unique to CSF-4. The 23 amino acid residues of the hydrophobic domain and the first 227 amino acid residues of the mature protein are highly conserved (87 and 82% respectively). Between amino acid residues 228 and 399, the amino acid sequence is diverging between the two species much more

rapidly (49% homology). However, there is a functional conservation in the amino acid changes in this area (71%). For example, neutral or weakly hydrophobic amino acids (proline, alanine, glycine, serine and threonine) are often substituted one for another. The amino acid divergence implies that the exact amino acid sequence downstream of amino acid 227 is not necessary. The functional conservation suggests that maintaining an appropriate 3-dimensional structure is important. Thus, only certain types of amino acid substitutions are allowed.

Because pcCSF-17 encodes a biologically active CSF-1 protein, we know that the additional coding sequence in CSF-4 is not necessary for the biological properties we have tested. However, whether the protein coded for by pcCSF-17 is enzymatically cleaved at a different position or at a different rate than the protein coded for by CSF-4 remains to be elucidated. Some differences in biological activity and pharmacokinetics between the two proteins may be found.

The isolation of two mRNA species that encode substantially different CSF-1 proteins raises many new questions. Among colony stimulating factors, the presence of a hydrophobic domain and two forms of protein encoded by two mRNAs are properties unique to CSF-1. These properties suggest that CSF-1 is functionally more diverse than the other colony stimulating factors. Studies designed to look at membane-bound forms versus serum forms will elucidate the reasons why this colony stimulating factor appears to have a transmembrane domain that can anchor it to the cell membrane. The hypothesis that the two forms of protein may have different half-lives either on cell membranes or in serum can be easily tested now that cDNA clones and large amounts of expressed protein are available.

Materials and methods

Isolation and sequencing of genomic CSF-1

Recombinant phage clones C and K were isolated as described (Kawasaki *et al.*, 1985). Phage 11 was obtained using a ^{32}P-labeled pcCSF-17 probe. Random shearing and specific restriction digestion were used to produce fragments of the CSF-1 genomic DNA. We subcloned these fragments into M13 (Yanisch-Perron *et al.*, 1985) and sequenced (Sanger *et al.*, 1977) a total of 16 kb. Approximately 40% of the sequence shown in Figure 1 was determined on both strands.

cDNA library construction

MIA PaCa-2 cells were induced with 10 ng/ml PMA and 10 μM retinoic acid as described (Ralph *et al.*, 1986). Total RNA was isolated by the method of Chirgwin *et al.* (1979). Poly (A)$^+$ RNA was twice selected on an oligo d(T) column and used to generate a cDNA library in λgt10 (Huynh *et al.*, 1985). The library was screened with GM11 (5'GCTAGGGATGGCTTTGGGGTACAGGCACTTGCAATCAGGC3') and pcCSF-17. Hybridization conditions for GM11 were 5 × SSC (standard saline citrate); 5 × Denhardt's, 10% formamide; 0.1% SDS; 25 mM $NaPO_4$; pH 7; 1 mM EDTA and 200 μg/ml yeast RNA at 42°C. Hybridization conditions for pcCSF-17 were the same except the formamide concentration was 50%. Insert cDNA from candidate clones was purified and ligated into M13mp18 and mp19 for sequencing.

Northern blot analysis

The murine and human mRNA (5 μg) were electrophoresed on a 1% agarose gel containing 2.2 M formaldehyde (Lehrach *et al.*, 1977) and transferred to nitrocellulose. Strips containing 1 lane each of murine mRNA and human mRNA were hybridized to oligonucleotide probes as indicated.

The oligonucleotides were end-labeled with T4 polynucleotide kinase and [λ-^{32}P]ATP. Hybridization conditions were essentially as described above for GM11. The hybridization and washing temperatures varied according to the T_m of the oligonucleotide. Final washes were in 1 × SSC; 0.1% SDS; 1 mM EDTA and 50 mN $NaPO_4$ at 5°C below the T_m of the oligonucleotide. The filters were autoradiographed at −70°C using Kodak XAR film and DuPont Quanta III intensifying screens.

Cos cell expression

COS-A_2 cells were transfected with CsCl gradient purified DNA using a DEAE dextran method (Sompayrac and Dana, 1981) with the addition of chloroquine (Luthman and Magnusson, 1983). Supernatants were harvested 72 h after transfection and assayed for CSF-1 activity by a human CSF-1 specific RIA (Das *et al.*, 1981) and a murine bone marrow colony assay (Kawasaki *et al.*, 1985).

Acknowledgements

We would like to thank M.Kim Warren, M.-T.Lee and the Cetus Assay Group for CSF-1 biological assays, Mazie Y.Coyne and Vicki Schwieckart for CSF-17 plasmids, Ken Myambo for assistance in DNA sequencing, Corey Levenson and Lauri Goda for oligonucleotide synthesis, Sharon Nilson and Eric A.Ladner for graphics and Deborah Jackson for typing the manuscript.

References

Breathnach,R. and Chambon,P. (1981) *Ann. Rev. Biochem.*, **50**, 349–383.
Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5294–5299.
Das,S.K., Stanley,E.R., Guilbert,L.J. and Forman,L.W. (1981) *Blood*, **58**, 630–641.
Das,S.K. and Stanley,E.R. (1982) *J. Biol. Chem.*, **257**, 13679–13684.
Devlin,J.J., Devlin,P.E., Myambo,K., Lilly,M., Rado,T. and Warren,M.K. (1987) *J. Leukocyte Biol.*, in press.
Dynan,W.S., Sazer,S., Tjian,R. and Schimke,R. (1986) *Nature*, **319**, 246–248.
Horiguchi,J., Warren,M.K., Ralph,P. and Kufe,D. (1986) *Biochem. Biophys. Res. Comm.*, **141**, 924–930.
Huynh,T.V., Young,R.A. and Davis,R.W. (1985) In Glover, D.M. (ed.), *DNA Cloning–A Practical Approach*. IRL Press, Oxford, Vol. I., pp. 49–78.
Ishii,S., Merlino,G. and Pastan,I. (1985a) *Science*, **230**, 1378–1381.
Ishii,S., Xu,Y.-H., Stratton,R., Roe,B., Merlino,G. and Pastan,I. (1985b) *Proc. Natl. Acad. Sci. USA*, **82**, 4920–4924.
Kawasaki,E., Ladner,M.B., Wang,A.M., Van Arsdell,J.N., Warren,M.K., Coyne,M.Y., Schweickart,V.L., Lee,M.-T., Wilson,K.J., Boosman,A., Stanley,E.R., Ralph,P. and Mark,D. (1985) *Science*, **230**, 291–296.
Kim,S.H., Moores,J.C., David,D., Respess,J.G., Jolly,D.J. and Friedmann,T. (1986) *Nucleic Acids Res.*, **14**, 3103–3118.
Laimins,L.A., Kessel,M., Rosenthal,N. and Khoury,G. (1983) In Gluzman,Y. and Shenk,T. (eds), *Communications in Molecular Biology. Enhancers and Eukaryotic Gene Expression.* Cold Spring Harbor Laboratory, NY, pp. 28–37.
Lehrach,H., Diamond,D., Wozney,J.M. and Boedtker,H. (1977) *Biochemistry*, **16**, 4743–4751.
Luthman,H. and Magnusson,G. (1983) *Nucleic Acids Res.*, **11**, 1295–1308.
Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Connell,C., Quon,D., Sim,G.K. and Efstratiadis,A. (1978) *Cell*, **15**, 687–701.
Mason,A.J., Evans,B., Cox,D., Shine,J. and Richards,R. (1983) *Nature*, **303**, 300–307.
Metcalf,D. (1986) *Blood*, **67**, 257–267.
Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.
Nagata,S., Tsuchiya,M., Asano,S., Yamamoto,O., Hirata,Y., Kubtota,N., Oheda,M., Nomura,H. and Yamazaki,T. (1986) *EMBO J.*, **5**, 575–581.
Padgett,R., Grabowski,P.J., Konarska,M.M., Seiler,S. and Sharp,P.A. (1986) *Ann. Rev. Biochem.*, **55**, 1119–1150.
Ralph,P., Warren,M.K., Lee,M.-T., Csejtey,J., Weaver,J., Broxmeyer,H., Williams,D., Stanely,E.R. and Kawasaki,E. (1986) *Blood*, **68**, 633–639.
Reed,R. and Maniatis,T. (1986) *Cell*, **46**, 681–690.
Reynolds,G.A., Basu,S.K., Osborne,T.F., Chin,D.J., Gil,G., Brown,M.S., Goldstein,J.L. and Luskey,K.L. (1984) *Cell*, **38**, 275–285.
Rich,A., Nordheim,A. and Wang,A. (1984) *Ann. Rev. Biochem.*, **53**, 792–846.
Sanger,T., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Shaw,G. and Kamen,R. (1986) *Cell*, **46**, 659–667.
Sompayrac,L.M. and Dana,K.J. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 7575–7578.
Valerio,D., Duyvesteyn,M., Dekker,B., Weeda,G., Berkrens,T., van der Voorn,L., van Ormondt,H. and van der Eb,A. (1985) *EMBO J.*, **4**, 437–443.
Wong,G., Witek,J., Temple,P., Wilkens,K., Leary,A., Luxenberg,D., Jones,S., Brown,E., Kay,R., Orr,E., Shoemaker,C., Golde,D., Kaufman,R., Hewick,R., Wang,E. and Clark,S. (1985) *Science*, **228**, 810–815.
Wong,G.G., Temple,P.A., Leary,A.C., Witek-Giannotti,J.S., Yang,Y.-C., Ciarletta,A.B., Chung,M., Murtha,P., Kriz,R., Kaufman,R.J., Ferenz,C.R., Sibley,B.S., Turner,K.J., Hewick,R.M., Clark,S.C., Yanai,N., Yokata,H., Yamada,M., Saito,M., Motoyoshi,K. and Takaku,F. (1987) *Science*, **235**, 1504–1508.
Yang,Y.-C., Ciarletta,A.B., Temple,P.A., Chung,M.P., Kovacic,S., Witek-Giannotti,J.S., Leary,A.C., Kriz,R., Donahue,R.E., Wong,G.G. and Clark,S.C. (1986) *Cell*, **47**, 3–10.
Yanisch-Perron,C., Vierra,J. and Messing,J. (1985) *Gene*, **33**, 103–119.

Received on May 22, 1987