The EMBO Journal vol.4 no.10 pp.2561–2568, 1985

Structure of the chromosomal gene for granulocyte-macrophage colony stimulating factor: comparison of the mouse and human genes

S.Miyatake, T.Otsuka, T.Yokota, F.Lee and K.Arai

DNAX Research Institute of Molecular and Cellular Biology, 901 California Avenue, Palo Alto, CA 94304, USA

Communicated by P.Berg

A cDNA clone that expresses granulocyte-macrophage colony stimulating factor (GM-CSF) activity in COS-7 cells has been isolated from a pcD library prepared from mRNA derived from concanavalin A-activated mouse helper T cell clones. Based on homology with the mouse GM-CSF cDNA sequence, the mouse GM-CSF gene was isolated. The human GM-CSF gene was also isolated based on homology with the human GM-CSF cDNA sequence. The nucleotide sequences determined for the genes and their flanking regions revealed that both the mouse and human GM-CSF genes are composed of three introns and four exons. The organization of the mouse and human GM-CSF genes are highly homologous and strong sequence homology between the two genes is found both in the coding and non-coding regions. A 'TATA'-like sequence was found 20–25 bp upstream from the transcription initiation site. In the 5'-flanking region, there is a highly homologous region extending 330 bp upstream of the putative TATA box. This sequence may play a role in regulation of expression of the GM-CSF gene. These structures are compared with those of different lymphokine genes and their regulatory regions.

Key words: lymphokine/helper T cell/hematopoiesis/gene expression/DNA sequence analysis

Introduction

Colony stimulating factors (CSFs) are humoral factors required for the proliferation and differentiation of committed hematopoietic progenitor cells *in vitro*. In the mouse system, CSFs that stimulate the growth of granulocyte and macrophage lineages have been characterized and are classified as G-CSF, M-CSF, GM-CSF and multi-CSF, also known as IL-3 (Burgess *et al.*, 1977; Stanley and Heard, 1977; Ihle *et al.*, 1982; Nicola *et al.*, 1983) based on the characteristic colonies formed in a semisolid culture system. Two types of human CSFs have been described with activity for granulocytes and macrophages (Nicola *et al.*, 1979). One, designated CSF-α, stimulates neutrophil, macrophage, and eosinophil colonies. The other, termed CSF-β, stimulates exclusively neutrophil and macrophage colonies. However, human factors are less well characterized than those from the mouse and, therefore, the exact relationship between mouse and human factors has not as yet been established.

Helper T cells are known to produce a variety of lymphokines including GM-CSF after activation by either antigen or lectin (Nabel *et al.*, 1981; Prystowski *et al.*, 1982). We have previously isolated one class of human GM-CSF cDNA clone from concanavalin A (Con A)-activated human helper T cell clones (Lee *et al.*, 1985). This cDNA clone, designated human GM-CSF, has strong homology with a mouse GM-CSF cDNA clone (Gough *et al.*, 1984, 1985) and expresses in COS-7 cells a CSF activity specific for neutrophil, macrophage and eosinophil lineages. These properties correspond well with those of CSF-α. Despite their high degree of sequence homology, activities encoded by human and mouse GM-CSF cDNA clones are species-specific. Human GM-CSF expressed in COS-7 cells does not stimulate colony formation *in vitro* using mouse bone marrow cells (Lee *et al.*, 1985).

To analyse further the relationship between human and mouse GM-CSFs, we have examined the organization of their genes. In this paper, we first describe the isolation from a Con A-activated helper T cell cDNA library of a mouse GM-CSF cDNA clone, which expresses GM-CSF activity in COS-7 monkey cells. We then describe the isolation of the human and mouse GM-CSF genes and their entire nucleotide sequences, and compare the structural features of mouse and human GM-CSF genes. Our results indicate that there exists only one copy of the human GM-CSF gene in the haploid genome and that the organization of mouse and human genes is highly conserved. These results establish that our GM-CSF cDNA clone isolated from a Con A-activated human T cell clone does, in fact, encode the human homologue of mouse GM-CSF. The structure of human and mouse GM-CSF genes have been compared with the structure of other inducible T cell lymphokine genes such as IL-2 (Fujita

Table I. Assay of biological activity of COS-7 supernatants transfected with GM-CSF cDNA clones

cDNA clone transfected	Source	Activity	
		Number of colonies	Proliferation of NSF-60 (units/ml)
E1-11	E1	250	1585
E1-6	E1	200	188
C5-1-1	C5	320	N.D.
C5-1-10	C5	275	544
LB2-1-1	LB2-1	75	0
LB2-1-9	LB2-1	23	0
LB2-1-14	LB2-1	0	N.D.
IL-3[a]	$Ly1^+2^-/9$	400	N.D.
Mock	–	0	0

The 5'-proximal 74-bp fragment of the published mouse GM-CSF cDNA sequence was synthesized. Using this ^{32}P-labeled fragment as a probe, 3×10^4 clones from three different Con A-activated helper T cell clone cDNA libraries, E1 and C5 (Clayberger *et al.*, 1983) and LB2-1 (Giedlin *et al.*, unpublished data; Yokota *et al.*, 1985) were screened by colony hybridization. Supernatants from COS-7 cells transfected with cDNA clones were assayed by *in vitro* colony formation. The same supernatants were tested in a proliferation assay using the NFS-60 cell line. 1 unit of GM-CSF is defined as the amount of GM-CSF required to give 50% of the maximal signal using NFS-60 cells in a volume of 0.1 ml.

[a]Mouse IL-3 represents the positive control. 10 μg of pcD-IL-3 clone B9 (Yokota *et al.*, 1984) was transfected into COS-7 cells. The supernatant collected 72 h after transfection was added to 5% in the colony stimulation assay. Only assays containing pcD-IL-3 supernatant showed burst-promoting activity.

N.D., not done.

```
        10          20          30                  46                  61                  76                  91                 106
CTCAGAGAGA AAGGCTAAGG TCCTGAGGAG G ATG TGG CTG CAG AAT TTA CTT TTC CTG GGC ATT GTG GTC TAC AGC CTC TCA GCA CCC ACC CGC TCA CCC ATC ACT GTC ACC CGG CCT
                                   MET Trp Leu Gln Asn Leu Leu Phe Leu Gly Ile Val Val Tyr Ser Leu Ser Ala Pro Thr Arg Ser Pro Ile Thr Val Thr Arg Pro

121                 136                 151                 166                 181                 196                 211                 226
TGG AAG CAT GTA GAG GCC ATC AAA GAA GCC CTG AAC CTC CTG GAT GAC ATG CCT GTC ACG TTG AAT GAA GAG GTA GAA GTC GTC TCT AAC GAG TTC TCC TTC AAG AAG CTA ACA
Trp Lys His Val Glu Ala Ile Lys Glu Ala Leu Asn Leu Leu Asp Asp MET Pro Val Thr Leu Asn Glu Glu Val Glu Val Val Ser Asn Glu Phe Ser Phe Lys Lys Leu Thr

        241                 256                 271                 286                 301                 316                 331                 346
TGT GTG CAG ACC CGC CTG AAG ATA TTC GAG CAG GGT CTA CGG GGC AAT TTC ACC AAA CTC AAG GGC GCC TTG AAC ATG ACA GCC AGC TAC TAC CAG ACA TAC TGC CCC CCA ACT
Cys Val Gln Thr Arg Leu Lys Ile Phe Glu Gln Gly Leu Arg Gly Asn Phe Thr Lys Leu Lys Gly Ala Leu Asn MET Thr Ala Ser Tyr Tyr Gln Thr Tyr Cys Pro Pro Thr

                361                 376                 391                 406                 421                 436                 451
CCG GAA ACG GAC TGT GAA ACA CAA GTT ACC ACC TAT GCG GAT TTC ATA GAC AGC CTT AAA ACC TTT CTG ACT GAT ATC CCC TTT GAA TGC AAA AAA CCA AGC CAA AAA TGA
Pro Glu Thr Asp Cys Glu Thr Gln Val Thr Thr Tyr Ala Asp Phe Ile Asp Ser Leu Lys Thr Phe Leu Thr Asp Ile Pro Phe Glu Cys Lys Lys Pro Ser Gln Lys  .

        467         477         487         497         507         517         527         537         547         557         567         577         587
GGAAGCCCAG GCCAGCTCTG AATCCAGCTT CTCAGACTGC TGCTTTTGTG CCTGCGTAAT GAGCCAGGAA CTTGGAATTT CTGCCTTAAA GGGACCAAGA GATGTGGCAC AGCCACAGTT GGAAGGCAGT

        597         607         617         627         637         647         657         667         677         687         697         707         717
ATAGCCCTCT GAAAACGCTG ACTCAGCTTG GACAGCGGAA GACAAACGAG AGATATTTTC TACTGATAGG GACCATTATA TTTATTTATA TATTTATATT TTTTAAATAT TTATTTATTT ATTTATTTAT

        727         737         747         757         767         777         787
TTTTGCAACT CTATTTATTG AGAATGTCTT ACCAGAATAA TAAATTATTA AAACTTTAAA AAAAAAAAAA AAAAAAA
```

Fig. 1. Nucleotide sequence and predicted amino acid sequence for mouse GM-CSF. The nucleotide sequence of cDNA begins with position 1 at the first nucleotide following the oligo(dG) segment.

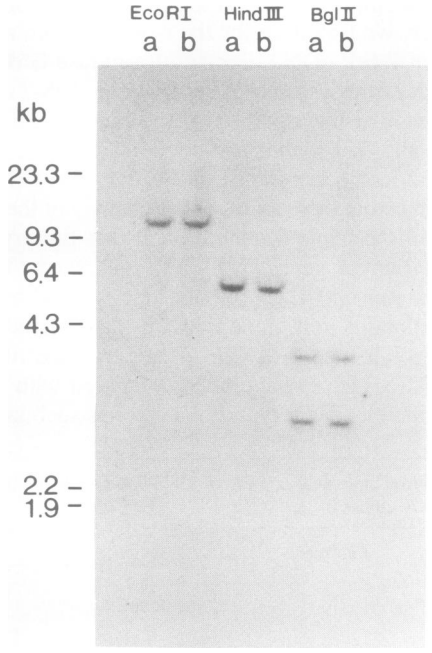

Fig. 2. Southern blotting of human chromosomal DNA with human GM-CSF cDNA probe. High mol. wt. DNA was prepared from HeLa cells (**lane a**) and T lymphoblast (Mo) cells (**lane b**). Each DNA sample (10 μg), digested with either *Eco*RI, *Hind*III or *Bgl*II, was electrophoresed on 0.8% agarose gel and transferred to nitrocellulose paper. The ^{32}P-labeled *Pst*I/*Aha*III fragment of human GM-CSF cDNA (5×10^7 c.p.m./μg) was hybridized with the filter in 6 × SSC at 65°C. The filter was washed with 2 × SSC at 65°C and then with 0.1 × SSC at room temperature. Numbers on the left indicate the position of size markers shown in kb.

et al., 1983; Fuse *et al.*, 1984), IL-3 (Miyatake *et al.*, 1985) and IFN-γ (Gray and Goeddel, 1982, 1983).

Results

Isolation of a functional mouse GM-CSF cDNA clone from a Con A-activated helper T cell library

The GM-CSF cDNA clone isolated from a mouse lung cell cDNA library (Gough *et al.*, 1984) was not full length. To isolate cDNA clones containing a complete copy of the GM-CSF mRNA, we have isolated seven GM-CSF clones from Con A-activated helper T cell clone cDNA libraries by colony hybridization. Approximately 0.5% of the libraries hybridized with the GM-CSF cDNA probe. The longest clones contained inserts of ~1 kb in length. Since the cDNA libraries were established in a pcD vector (Okayama and Berg, 1983) which promotes the expression of cDNA inserts in mammalian cells, the DNA of seven clones (E1-6, E1-11, C5-1-1, C5-1-10, LB2-1-1, LB2-1-9 and LB2-1-14) was transfected into COS-7 cells. Supernatants were assayed by *in vitro* colony forming assays (Rennick *et al.*, 1985) and proliferation assays with the GM-CSF-dependent NFS-60 cell line (provided by J.Ihle). As shown in Table I, four clones, E1-6, E1-11, C5-1-1 and C5-1-10 were active in the colony-forming assay. Three clones (E1-6, E1-11, C5-1-10) also exhibited biological activity in the NFS-60 proliferation assay.

Nucleotide sequence of mouse GM-CSF cDNA

The nucleotide sequence of clone E1-11, one of the functional cDNA clones, was determined (Figure 1). The first ATG is found 32 – 34 nucleotides from the 5′ end and is followed by 141 codons before the termination codon TGA. The NH_2-terminal segment of the predicted GM-CSF amino acid sequence is hydrophobic as would be expected for a signal peptide.

Southern blotting analysis of human chromosomal DNA

There is a single copy gene encoding mouse GM-CSF (Gough *et al.*, 1985). Southern blotting analysis was performed with chromosomal DNA isolated from HeLa cells which are non-producers and Mo cells which are constitutive producers of human GM-CSF (Wong *et al.*, 1985) to elucidate the number of genes in the human genome.

Chromosomal DNA from both HeLa and Mo cells showed the same hybridization patterns (Figure 2), indicating that a single *Bgl*II site exists but neither *Eco*RI nor *Hind*III sites occur within the human GM-CSF gene. The results also indicated that each haploid genome contains a single copy of the GM-CSF gene. Furthermore, no detectable rearrangement was found around the GM-CSF gene in the Mo cell line.

Nucleotide sequences of mouse and human GM-CSF genes

The mouse GM-CSF gene was isolated from a sperm DNA library in the λ phage vector Charon 4A and the 12-kb insert of one clone, λGM-CSF12, was subcloned in pUC13 (Vieira and Messing, 1982) to yield pλGM-CSF12 (Figure 3A). A human placental genomic λ phage library was screened using a *Pst*I/*Aha*III fragment derived from the human GM-CSF cDNA as a probe. Seven positive phage plaques were identified. The *Hind*III/*Eco*RI fragment of λHGM11-a (Figure 3B) which contains the entire human GM-CSF gene was subcloned into pUC13 to yield pHG23. The nucleotide sequences of the mouse and human GM-CSF genes and their flanking regions were deter-

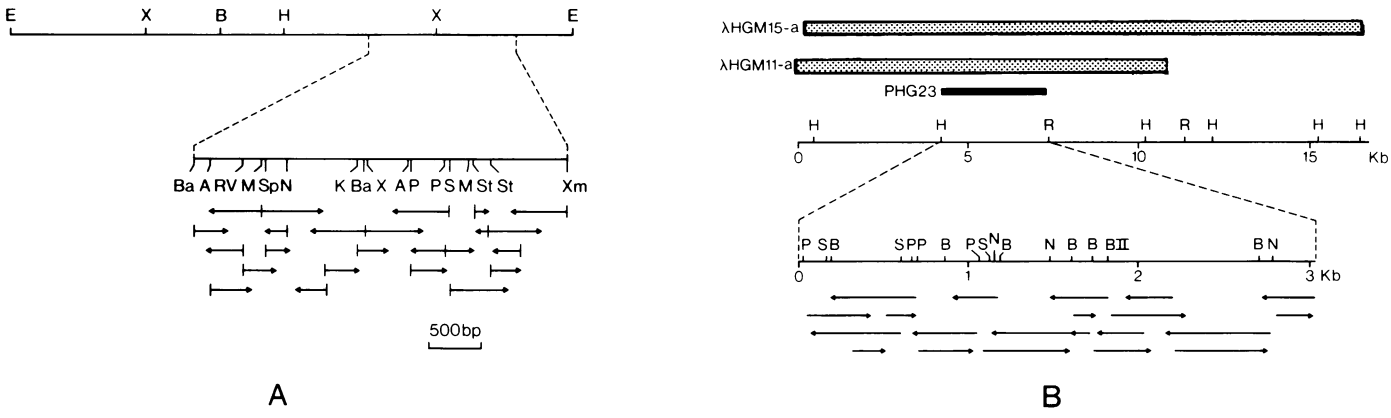

Fig. 3. (**A**) Restriction cleavage map of the mouse GM-CSF gene region. Seven positive clones were isolated from 6×10^5 plaques of mouse sperm DNA library. Six of them contained the entire region of the chromosomal gene as judged by hybridization with either 5′- or 3′-specific probes. The upper line shows the restriction map of the entire 12-kb insert of λGM-CSF12. The region (shown in an expanded scale) that contains the GM-CSF gene of pλGM-CSF12 was sequenced from various restriction sites by the dideoxy chain termination method (solid arrows). A, *Aha*III; B, *Bam*HI; Ba, *Bal*I; E, *Eco*RI; H, *Hind*III; K, *Kpn*I; M, *Mbo*I; N, *Nco*I; P, *Pst*I; RV, *Eco*RV; S, *Sst*I; Sp, *Sph*I; St, *Stu*I; X, *Xba*I; Xm, *Xmn*I. (**B**) Restriction endonuclease cleavage map and sequence analysis of the human GM-CSF gene. λHGM11-a and λHGM15-a, whose regions are indicated as the upper two lines, were isolated from 1×10^6 plaques of the human genomic library. The *Hind*III/*Eco*RI fragment (thick line) that contains the GM-CSF gene, subcloned from HGM11-a into pUC13 (pHG23), is shown in an expanded scale. Arrows indicate the direction and the extent of sequence analysis. B, *Bal*I; BII, *Bgl*II; R, *Eco*RI; H, *Hind*III; N, *Nco*I; P, *Pst*I; S, *Sst*I.

mined by the strategy shown in Figure 3. When these sequences are compared with the sequence of each cDNA, four exons and three introns can be identified (Figures 4 and 5).

As is the case for several other lymphokine genes (Gray and Goeddel, 1982, 1983; Fujita *et al.*, 1983; Fuse *et al.*, 1984; Miyatake *et al.*, 1985), each intron interrupts the reading frame precisely between codons. The consensus sequences for splicing junctions, G/GTA(G)AG for the donor splice site and AG for the acceptor splice site, were found in all splice sites (Breathnach *et al.*, 1978; Lerner *et al.*, 1980).

'TATA'-like sequences (positions 1107 – 1111 in mouse, positions 597 – 603 in human) are found 20 – 25 bp upstream from the transcription initiation sites of both genes determined by S1 mapping as described below (Figures 4 and 5). Another sequence which is relatively conserved in the promoter region of eukaryotic genes is the sequence GGC(T)CAATCT located ~80 bp upstream of the cap site (Benoist *et al.*, 1980). A homologous sequence is not found around 80 bp upstream of the cap site of the GM-CSF genes.

Determination of the transcription initiation site by S1 nuclease mapping

The site of transcription initiation was mapped by measuring the RNA-dependent protection of a radiolabelled DNA probe from an S1 nuclease digestion (Berk and Sharp, 1977).

The protected fragments were detected predominantly with termini corresponding to position 1136 (A) and position 1135 (T) of mouse GM-CSF gene using mRNA from Con A-induced helper T cell clone $Ly1^+2^-/9$ (Figure 6). Essentially the same result was obtained using poly(A)$^+$ RNA from Con A-induced helper T cell clone LB2-1 (data not shown). It should be noted that the 5′ end of the nearly full-length GM-CSF cDNA (clone E1-11) is at position 1137 (C). Since most transcripts synthesized by RNA polymerase II begin with a purine, the A at 1136 is the most likely site of transcription initiation. This residue is located 32 bp upstream of the first ATG codon (position 1168 – 1170) and 25 bp downstream of a 'TATA'-like sequence (position 1107 – 1111).

The sequence of the 5′ end of Mo cell-derived human GM-CSF mRNA was determined by S1 nuclease mapping. The results indicate that positions 620 – 622 (Figure 5) may serve equally well as a cap site and suggest that microheterogeneity of the GM-CSF mRNA might be the case (data not shown).

Discussion

Nucleotide sequence of mouse GM-CSF cDNA and predicted amino acid sequence of mouse GM-CSF protein

The mouse GM-CSF cDNA contains a single open reading frame consisting of 141 codons. Analysis of the hydrophobicity of the polypeptide and comparison with a proposed consensus sequence for the processing of signal peptides (Perlman and Halvorson, 1983) suggest that processing of the precursor polypeptide would probably occur following either of the serine residues at positions 15 and 17, or the alanine residue at position 18.

The mature protein would then consist of ~120 amino acid residues corresponding to a protein with a calculated mol. wt. of 12 000, while the reported mol. wt. of mouse GM-CSF is 29 000 (Burgess *et al.*, 1977) or 33 000 (Sparrow *et al.*, 1985). This difference could be explained by post-translational glycosylation of the molecule.

The isolation of a full-length mouse GM-CSF cDNA clone (pGM3.2) from a cloned T cell line was recently reported (Gough *et al.*, 1985). In comparing the sequences of pGM3.2 and clone E1-11 described in this paper, we find that pGM3.2 contains an additional 146 bp at its 5′ end which are not present in E1-11. With the exception of 5 bp at the 5′ end, the entire E1-11 sequence is contained in pGM3.2. The origin of the extra 146 bp is at present unknown. This sequence could not be found in the 1.1 kb of 5′-flanking sequence that has been sequenced.

The predicted major open reading frame of pGM3.2 has two ATG initiator codons. The initiator ATG codon in E1-11 corresponds to the second ATG codon in pGM3.2. Therefore, the predicted amino acid sequence of E1-11 lacks 12 NH_2-terminal residues which are charged and therefore hydrophilic. It was suggested that the hydrophilic region which preceeds the putative hydrophobic signal peptide may yield a membrane-bound form of GM-CSF. However, S1 mapping showed that the predomi-

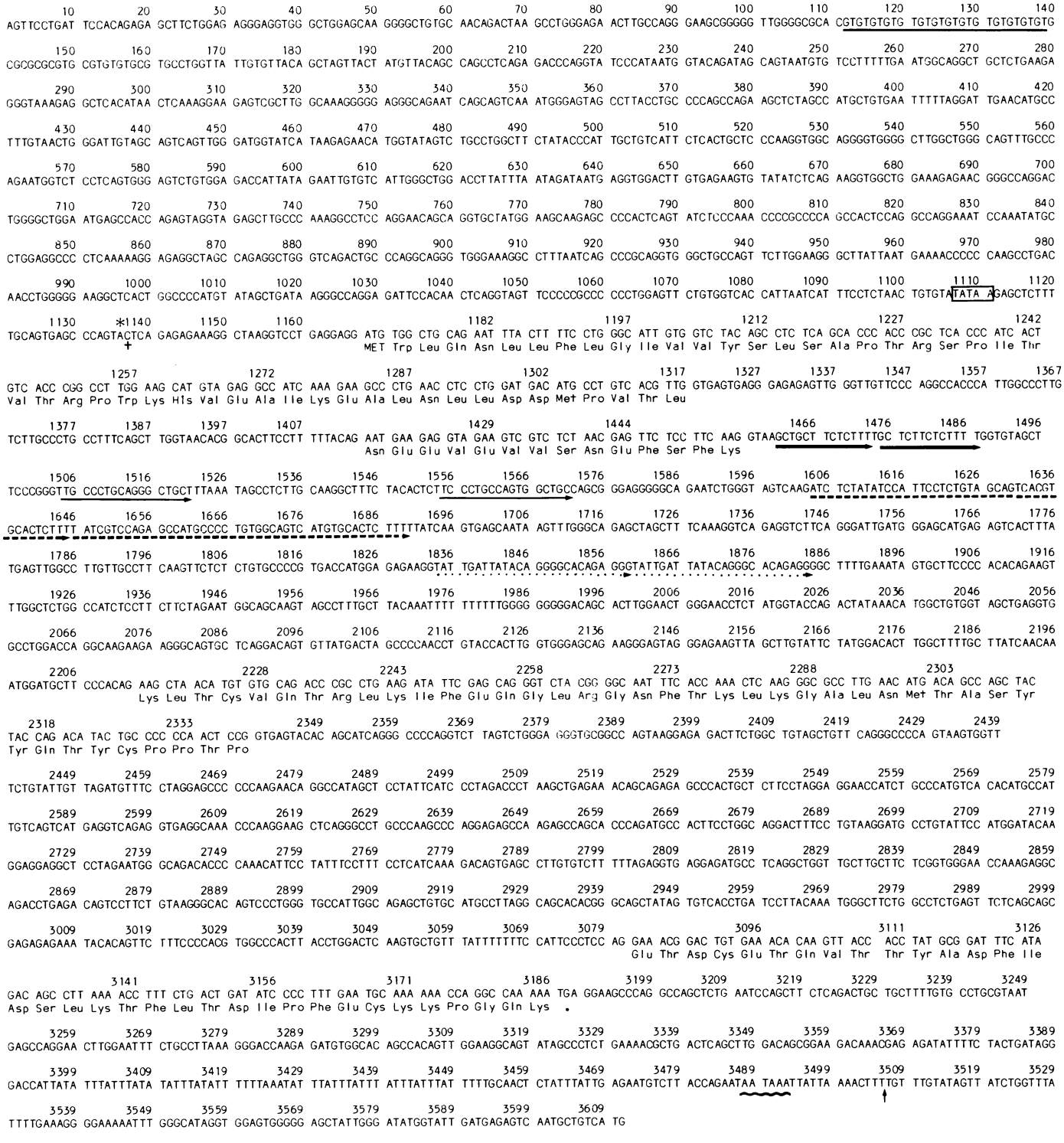

Fig. 4. Nucleotide sequence of the mouse GM-CSF gene and its flanking regions. The coding sequences of the exons have been translated. From left to right the following sequences are identified: 14 contiguous GT dinucleotides (underline, 112 – 139), TATA box (box, 1107 – 1111), transcription initiation site (asterisk, 1136), 5′ end of the cDNA clone E1-11 (cross, 1137), direct repeat 1 (thick arrows, 1461 – 1474, 1475 – 1487), direct repeat 2 (thin arrows, 1504 – 1520, 1554 – 1571), direct repeat 3 (dashed arrows, 1604 – 1646, 1647 – 1690), direct repeat 4 (dotted arrows, 1834 – 1859, 1860 – 1884), polyadenylation site (wavy line, 3488 – 3493), and position of junction with poly(A) tail (vertical arrow, 3506).

nant transcription initiation sites of the GM-CSF mRNAs in two helper T cell clones Ly1^{+}2^{-}/9 and LB2-1, are 32 bp upstream of the first ATG codon in the E1-11 cDNA sequence. These results suggest that the mRNA for a putative membrane-bound form of GM-CSF may be very rare or absent in the T cell clones used in our studies.

Comparison of mouse and human GM-CSF genes

We have shown that a human haploid genome contains only one copy of the GM-CSF gene (Figure 2) as is the case for mouse GM-CSF (Gough *et al.*, 1984). The results described in this paper established that both human and mouse GM-CSF genes are com-

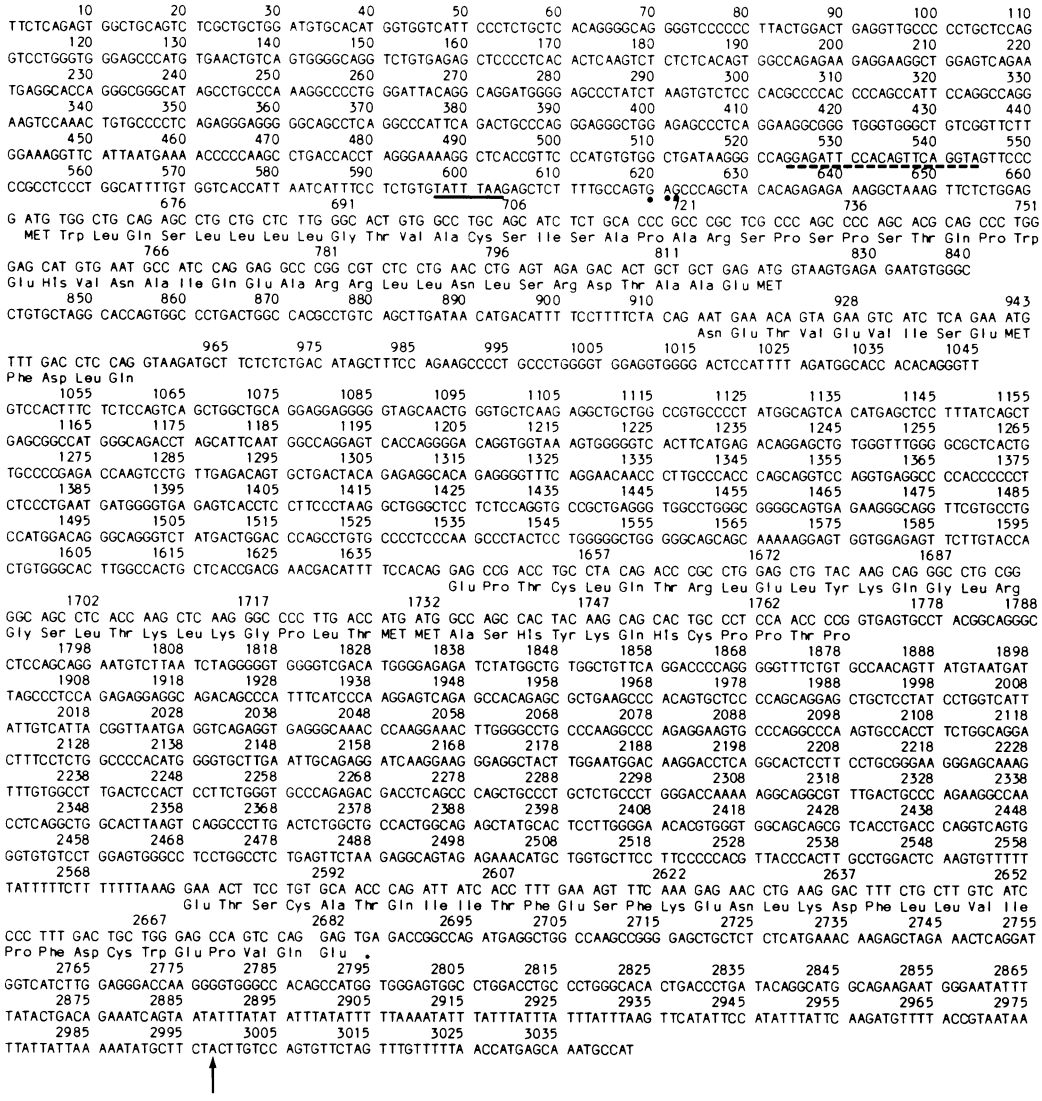

Fig. 5. Nucleotide sequence of human GM-CSF gene and the flanking regions. The coding sequences of the four exons have been translated. From left to right, the following sequences are identified: common sequence in the 5′-flanking region found in both mouse and human GM-CSF and in mouse IL-3 genes (dashed line, 524 − 544), TATA box (underline, 597 − 603), transcription initiation site (dot, 620 − 622), and junction between cDNA and poly(A) tail (vertical arrow, 2998).

posed of three introns and four exons and are organized in a similar manner (Figure 7). The sizes of exons 2, 3 and 4 (defined from the beginning of exon 4 to the stop codon TGA) are identical in both species and therefore, each exon encodes exactly the same number of amino acid residues in both species. However, exon 1 of human GM-CSF is 9 bp longer than exon 1 of mouse GM-CSF. In addition, the length of each intron is nearly the same in both copies. The mouse GM-CSF gene contains four direct repeats in the second intron. No such repeated sequences are found in the introns of the human GM-CSF gene. Nucleotide sequences of the mouse and human GM-CSF cDNA clones share ~70% homology and the amino acid sequences share ~50% homology (Lee *et al.*, 1985). In general, intron sequences show more diversity than exon sequences. However, stretches which show >70% homology are clustered in each intron. The most highly conserved sequences were found in the 5′-flanking region extending ~330 bp upstream of putative TATA boxes (Figure 7). Upstream sequences beyond this point show very little or no homology. The remarkable conservation of the overall structures and the nucleotide sequences of human and mouse GM-CSF genes may indicate that they evolved from a common ancestral gene. From these results, we conclude that the GM-CSF cDNA clone isolated from a Con A-stimulated human helper T cell clone encodes a human homologue of mouse GM-CSF.

Strong conservation in the 5′-flanking regions of mouse and human GM-CSF genes suggests that these sequences may play a role in regulating expression of the genes during T cell activation. Similarly, mouse and human IL-2 genes show strong homology in the 5′-flanking regions (Fuse *et al.*, 1984).

A sequence composed of 14 contiguous GT dinucleotides is found ~1 kb upstream from the transcription initiation site of mouse GM-CSF gene (position 112 − 139 in Figure 4). Such a sequence may adopt a left-handed conformation (Z-DNA) (Wang *et al.*, 1979; Haniford and Pulleyblank, 1983; Nordheim and Rich, 1983a, 1983b) and is reported to have an enhancer-like

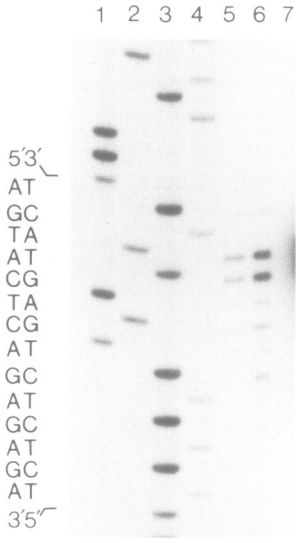

Fig. 6. S1 nuclease mapping of mouse GM-CSF mRNA in activated helper T cells. The 1.8-kb *Acc*I fragment from the GM-CSF chromosomal gene which encompasses the 5′ end of exon 1 and ~1.7 kb of the 5′-flanking region was isolated from an agarose gel and the 5′ ends were labeled with [γ-^{32}P]ATP (~1 × 10^7 c.p.m. ^{32}P/μg). 4 μg of poly(A)$^+$ RNA or 30 μg of total RNA isolated from a cloned helper T cell line 4 h after Con A activation was mixed with 0.3 μg of the labeled DNA probe and incubated at 75°C for 10 min in a volume of 15 μl (Berk and Sharp, 1977). The mixture was immediately transferred to 45°C and incubated for at least 3 h. The hybrids were treated with 75 units of S1 nuclease for 45 min at 37°C in a volume of 165 μl. Then the products were analysed on an 8% sequencing gel. An 18-bp synthetic fragment of the anti-sense strand, with one end at the *Acc*I site in exon 1 exending 18 bp upstream, was used as a primer to sequence plasmid DNA carrying the GM-CSF gene by the supercoiled dideoxy chain termination method. **Lanes 1–4** represent the dideoxy sequence (G,A,T,C). **Lanes 5** and **6** represent poly(A)$^+$ RNA and total RNA of helper T cell clone Ly1$^+$2$^-$/9 isolated 4 h after Con A stimulation, respectively. **Lane 7** represents the total RNA from resting helper T cells of the same cell line. The nucleotide pairs at the left hand side represent the corresponding sequence of the mouse GM-CSF gene.

activity (Hamada *et al.*, 1984a, 1984b). It is tempting to speculate that this sequence may have an enhancing effect on expression of the mouse GM-CSF genes.

Comparison of the structure of lymphokine genes

When helper T cells are activated by lectin or antigen, various lymphokines are induced at high levels (Nabel *et al.*, 1981; Prystowsky *et al.*, 1982). Therefore, many lymphokine genes may contain specific sequence(s) required for this inducible expression. If this is true, it might be possible to find potential consensus sequences which would be required for regulated expression of lymphokine genes by comparing their 5′-flanking sequences. We have performed this comparison using the 5′-flanking region sequences of IL-2, IL-3, IFN-γ and GM-CSF. Although no convincing homologies were found which are shared by all, some sequence homologies were detected in some of the lymphokine genes.

Efficient transcription of some of the eukaryotic genes requires the GC-rich sequences 5′-CCGCCC-3′ or 5′-CACCC-3′ (Fromm and Berg, 1982; Orkin *et al.*, 1982; Dynan and Tijan, 1983; Dierks *et al.*, 1983; Treisman *et al.*, 1983; McKnight *et al.*, 1984). Similar sequences are found in a GC-rich region preceding the TATA box of the mouse IL-3 gene (Miyatake *et al.*, 1985) and upstream of the transcription initiation sites of both the mouse and human IL-2 (Fujita *et al.*, 1983; Fuse *et al.*, 1984) and GM-CSF genes (Figure 8). These GC-rich sequences are not found in the 5′-flanking region of the human IFN-γ gene.

There are two regions of shared homology between the IL-3 and GM-CSF genes. One is 90–130 bp, the other 240–260 bp, upstream of the transcription start sites of the IL-3 and GM-CSF genes (Figure 8). An additional region of homology between mouse GM-CSF and human IFN-γ is located 50–70 bp and 80–100 bp upstream of transcription initiation sites of the GM-CSF and IFN-γ genes, respectively (Figure 8) (Gray and Goeddel, 1982, 1983).

The absence of a consensus sequence shared by all of the known lymphokine genes, and the presence of several homologies

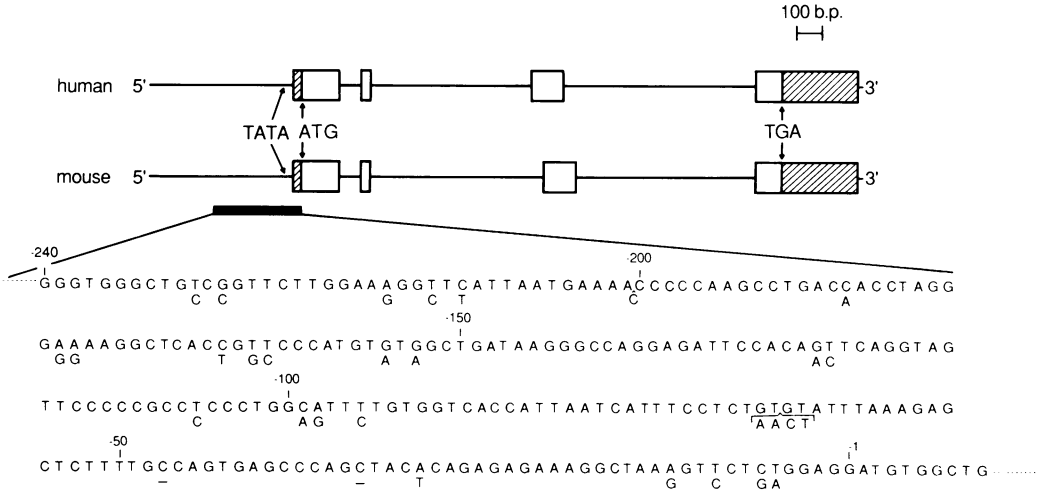

Fig. 7. Schematic representation of the organization of the human and the mouse GM-CSF genes. Coding region of exons, open box; untranslated region, hatched box; introns and the flanking region, thin line; sequence of the highest homology in the 5′-flanking region, thick bar. The nucleotide sequence of this region of the human GM-CSF gene is shown at the bottom. Mouse nucleotides differing from the human sequence are shown underneath. Nucleotides missing in the mouse gene are indicated by bars. Nucleotides are numbered negatively upstream from the first base of the initiation codon.

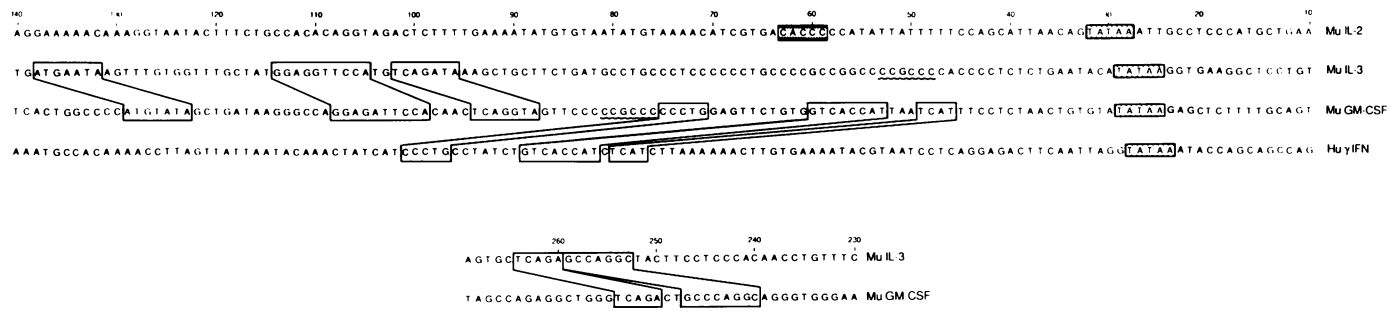

Fig. 8. The 5'-flanking sequences of four different lymphokine genes (mouse IL-2, mouse IL-3, mouse GM-CSF and human IFN-γ). Nucleotides are numbered negatively upstream from the transcription initiation site. Boxes with diagonal lines define a 'TATA'-like sequence. The wavy line shows the GC-rich sequence that is found both in the SV40 GC-rich region upstream of the early promoter and in the herpes simplex virus tk gene promoter. A similar sequence found in mouse IL-2 is defined by a heavily shaded box. The sequences enclosed by solid lines show homologous regions between mouse IL-3 and mouse GM-CSF, and mouse GM-CSF and human IFN-γ.

for subsets of these genes suggest that there might be several different mechanisms for the activation of lymphokine genes in helper T cells.

Materials and methods

Cell lines and isolation of mRNA

Helper T cell clones E1 and C5 are reactive to either trinitrophenol (TNP) on a BALB/c MHC background or to dinitrophenol (DNP-ovalbumin conjugate) on a BL/6 background, respectively (Clayberger *et al.*, 1983). Helper T cell clone LB2-1 recognizes chicken red blood cells on a BL/6 background (Giedlin *et al.*, unpublished data). These helper T cell clones were stimulated with 2 μg/ml Con A. 4 h after the addition of Con A, all cells were collected and the total cellular RNA was extracted by the guanidium-thiocyanate method (Chirgwin *et al.*, 1979). The Mo cell line is an HTLV-II transformed human T lymphoblast cell (Kalyanaraman *et al.*, 1982). Poly(A)$^+$ RNA was selected by chromatography on oligo(dT)-cellulose.

Construction of cDNA libraries

cDNA libraries are established with mRNA from E1, C5 and LB2-1 cells using the pcDV1 vector-primer and the pL1 linker fragment according to the procedure of Okayama and Berg (1983).

Screening of cDNA library by hybridization

The 5'-proximal 74-bp fragment of the reported mouse GM-CSF cDNA (Gough *et al.*, 1984) was synthesized and a labeled probe was made by nick translation (1×10^8 c.p.m. ^{32}P/μg). This probe was used to screen 3×10^4 colonies of a cDNA library transferred onto nitrocellulose filters. Relatively low stringency hybridization conditions were used: 6 × SSPE (1.08 M NaCl, 60 mM sodium phosphate pH 7.4 and 6 mM EDTA) (Maniatis *et al.*, 1982), 20% formamide, 0.1% SDS, and 100 μg/ml carrier tRNA overnight at 42°C. The filters were washed with 2 × SSPE, 0.1% SDS at 65°C for ~1 h.

Identification of GM-CSF cDNA clones by transfection

Plasmid DNA of cDNA clones which hybridized with the probe were isolated by cesium chloride-ethidium bromide gradients (Maniatis *et al.*, 1982). 10^6 COS-7 monkey cells were transfected with 25 μg of plasmid DNA using DEAE-dextran as described previously (Yokota *et al.*, 1984). After 4 h incubation at 37°C, the cells were washed in Dulbecco-modified Eagle's medium (DME) containing 150 μM chloroquine as described previously (Yokota *et al.*, 1985). After 3 h this medium was replaced with DME containing 4% fetal calf serum. After 72 h the medium was collected and assayed in a colony formation assay.

In vitro colony forming assays

1.3×10^5 non-adherent bone marrow cells isolated from mouse femur were added to 35 mm Petri dishes in a total volume of 1 ml of Iscove's medium containing 25% fetal calf serum, 10 μM 2-mercaptoethanol, 0.9% methylcellulose, and 0.3 ml test sample. Cultures were incubated for 3 days, then 1 unit of mouse erythropoietin was added (Rennick *et al.*, 1985). Colonies were scored after 7 days.

Proliferation of the NFS-60 cell line

The NFS-60 cell line (kindly provided by J.Ihle) was originally isolated from a myeloid leukemia in NFS mice (Holmes *et al.*, 1985). Proliferation was assayed using a colorimetric method (Mosmann, 1983). A subline of NFS-60 which responds to GM-CSF (provided by D.Rennick) was used in the present study. This assay is sensitive and requires only 2 days.

DNA sequence analysis

Nucleotide sequences were determined using the phage M13 dideoxy chain termination method (Sanger *et al.*, 1977) and the supercoiled dideoxy chain termination method with denatured plasmid DNA as a template (S.Hattori and Y.Sakaki, unpublished data).

Southern blotting of human chromosomal DNA

Chromosomal DNAs, extracted as described (Maniatis *et al.*, 1982), from HeLa cells and Mo cells were digested with restriction endonuclease and electrophoresed on a 0.8% agarose gel. Blotting analysis was done according to the method of Southern (1975). Hybridization was performed using nick-translated ^{32}P-labeled *Pst*I/*Aha*III fragment of human GM-CSF cDNA (Lee *et al.*, 1985). The filter was washed with 2 × SSC (0.3 M sodium chloride/30 mM sodium citrate) at 65°C and then with 0.1 × SSC at room temperature.

Cloning of the mouse and human GM-CSF genes

A mouse (BALB/c) sperm DNA library established in the λ phage vector Charon 4A was provided by Mark Davis. 6×10^5 plaques of this library were screened by plaque hybridization using a *Pst*I/*Aha*III fragment containing the mouse GM-CSF cDNA clone as the probe.

A human genomic library (kindly provided by T.Maniatis) was screened with a human GM-CSF cDNA probe under stringent conditions as described above. The library was established in the *Eco*RI site of λ Charon 4A using human placental DNA partially digested with restriction endonucleases *Hae*III and *Alu*I. 1×10^6 plaques were screened by ^{32}P-labeled human GM-CSF probe. DNA from positive λ phages were purified and subjected to restriction mapping. A 3.1-kb *Hin*dIII/*Eco*RI restriction fragment of one of the positive clones, λHGM11-a, was isolated and subcloned into the plasmid vector pUC13.

Acknowledgements

We thank Sandra M.Zurawski and John Mason for sequencing clone E1-11, Joan W.Conaway for advice in S1 mapping, Donna Rennick and Lorraine Gemmell for *in vitro* colony formation assays, Nancy Larson and Chris Alafi for screening cDNA and genomic libraries, and Tim Mosmann for providing LB2-1 mRNA. We thank Naoko Arai and Gerard Zurawski for helpful discussions. S.Miyatake is the recipient of a DNAX Research Institute Graduate Fellowship.

References

Benoist,C., O'Hare,K., Breathnach,R. and Chambon,P. (1980) *Nucleic Acids Res.*, **8**, 127-142.

Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721-732.

Breathnach,R., Benoist,C., O'Hare,K., Gannon,F. and Chambon,P. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 4853-4857.

Burgess,A.W., Camakaris,J. and Metcalf,D. (1977) *J. Biol. Chem.*, **252**, 1998-2003.

Clayberger,C., Dekruyff,R.H., Aisenberg,J. and Cantor,H. (1983) *J. Exp. Med.*, **157**, 1906-1916.

Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry (Wash.)*, **18**, 5294-5299.

Dierks,P., van Ooyen,A., Cochran,M.D., Dobkin,C., Reiser,J. and Weissmann,C. (1983) *Cell*, **32**, 695-706.

Dynan,W.S. and Tijan,P. (1983) *Cell*, **32**, 669-680.

Fromm,M. and Berg,P. (1982) *J. Mol. Appl. Genet.*, **1**, 457-481.

Fujita,T., Takaoka,C., Matsui,H. and Taniguchi,T. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 7437-7441.

Fuse,A., Fujita,T., Yasumitsu,H., Kashima,N., Hasegawa,K. and Taniguchi,T. (1984) *Nucleic Acids Res.*, **12**, 9323-9331.
Gough,N.M., Gough,J., Metcalf,D., Kelso,A., Grail,D., Nicola,N.A., Burgess,A.W. and Dunn,A.R. (1984) *Nature*, **309**, 763-767.
Gough,N.M., Metcalf,D., Gough,J., Grail,D. and Dunn,A.R. (1985) *EMBO J.*, **4**, 645-653.
Gray,P.W. and Goeddel,D.V. (1982) *Nature*, **298**, 859-863.
Gray,P.W. and Goeddel,D.V. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5842-5846.
Hamada,H., Petrino,M.G., Kakunaga,T., Seidman,M. and Stollar,B.D. (1984a) *Mol. Cell. Biol.*, **4**, 2610-2621.
Hamada,H., Seidman,M., Howard,B.H. and Gorman,C.M. (1984b) *Mol. Cell. Biol.*, **4**, 2622-2630.
Haniford,D.B. and Pulleyblank,D.E. (1983) *Nature*, **302**, 632-634.
Holmes,K.L., Palaszynski,E., Morse,H.C.,III and Ihle,J.N. (1985) *Proc. Natl. Acad. Sci. USA*, in press.
Ihle,J.N., Keller,J., Henderson,L., Frederick,K. and Palaszynski,E. (1982) *J. Immunol.*, **129**, 2431-2436.
Kalyanaraman,V.S., Sarngadharan,M.G., Robert-Guroff,M., Miyoshi,I., Blayney,D., Golde,D. and Gallo,R.C. (1982) *Science (Wash.)*, **218**, 571-573.
Lee,F., Yokota,T., Otsuka,T., Gemmell,L., Larson,N., Luh,J., Arai,K. and Rennick,D. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4360-4364.
Lerner,M.R., Boyle,J.A., Mount,S.M., Wolen,S.L. and Steitz,J.A. (1980) *Nature*, **283**, 220-224.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY, USA.
McKnight,S.L., Kingsbury,R.C., Spence,A. and Smith,M. (1984) *Cell*, **37**, 253-262.
Miyatake,S., Yokota,T., Lee,F. and Arai,K. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 316-320.
Mosmann,T. (1983) *J. Immunol. Methods*, **65**, 55-63.
Nabel,G., Greenberger,J.S., Sakakeeny,M.A. and Cantor,H. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 1157-1161.
Nicola,N.A., Metcalf,D., Johnson,G.R. and Burgess,A.W. (1979) *Blood*, **54**, 614-627.
Nicola,N.A., Metcalf,D., Matsumoto,M. and Johnson,G.R. (1983) *J. Biol. Chem.*, **258**, 9017-9021.
Nordheim,A. and Rich,A. (1983a) *Nature*, **303**, 674-679.
Nordheim,A. and Rich,A. (1983b) *Proc. Natl. Acad. Sci. USA*, **80**, 1821-1825.
Okayama,H. and Berg,P. (1983) *Mol. Cell. Biol.*, **3**, 280-289.
Orkin,S.H., Kazazian,H.H., Antonarakis,S.E., Goff,S.C., Boehm,C.D., Sexton,J.P., Waber,P.G. and Giardina,P.J.V. (1982) *Nature*, **296**, 627-631.
Perlman,D. and Halvorson,H.O. (1983) *J. Mol. Biol.*, **167**, 391-409.
Prystowsky,M.B., Ely,J.M., Beller,D.I., Eisenbert,L., Goldman,J., Remold,H., Vogel,S.N. and Fitch,F.W. (1982) *J. Immunol.*, **129**, 2337-2344.
Rennick,D.M., Lee,F.D., Yokota,T., Arai,K., Cantor,H. and Nabel,G.J. (1985) *J. Immunol.*, **134**, 910-914.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.
Sparrow,L.G., Metcalf,D., Hunkapiller,M.W., Hood,L.E. and Burgess,A.W. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 292-296.
Stanley,E.R. and Heard,P.M. (1977) *J. Biol. Chem.*, **252**, 4305-4312.
Treisman,R., Orkin,S.H. and Maniatis,T. (1983) *Nature*, **302**, 591-596.
Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259-268.
Wang,A.H.-J., Quigley,G.J., Kolpack,F.J., Crawford,J.L., Van Boom,J.H., van der Marel,G. and Rich,A. (1979) *Nature*, **282**, 680-686.
Wong,G.G., Witek,J.S., Temple,P.A., Wilkens,K.M., Leary,A.C., Luxenberg, D.P., Jones,S.S., Brown,E.L., Kay,R.M., Orr,E.C., Shoemaker,C., Golde, D.W., Kaufmann,R.J., Hewick,R.M., Wang,E.A. and Clark,S.C. (1985) *Science (Wash.)*, **228**, 810-815.
Yokota,T., Lee,F., Rennick,D., Hall,C., Arai,N., Mosmann,T., Nabel,G., Cantor,H. and Arai,K. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1070-1074.
Yokota,T., Arai,N., Lee,F., Rennick,D., Mosmann,T. and Arai,K. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 68-72.

Received on 31 May 1985; revised on 24 July 1985