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 Aactivated 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	Source	Activity				
transfected		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 ³²P-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.

^aMouse IL-3 represents the positive control. 10 μ g 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 CTCAGAGAGA	20 AAGGCTAAGG	30 TCCTGAGGAG	G ATG TGG MET Trp	46 CTG CAG AAT TT Leu Gin Asn Le	61 A CTT TTC CTG GGG u Leu Phe Leu Gly	1 C ATT GTG GTC TA y lle Val Val Ty	76 AC AGC CTC TCA G yr Ser Leu Ser A	91 GCA CCC ACC CGC TCA Na Pro Thr Arg Ser	106 CCC ATC ACT GTC ACC CGG CCT Pro lle Thr Val Thr Arg Pro
121 TGG AAG CAT Trp Lys His	1: T GTA GAG GO s Val Glu A	36 CC ATC AAA Ia Ile Lys	151 GAA GCC CTG Glu Ala Leu	AAC CTC CTG G Asn Leu Leu A	166 AT GAC ATG CCT G sp Asp MET Pro Va	181 TC ACG TTG AAT G al Thr Leu Asn G	196 GAA GAG GTA GAA Glu Glu Val Glu	21 GTC GTC TCT AAC GA Val Val Ser Asn Gl	1 226 G TTC TCC TTC AAG AAG CTA ACA u Phe Ser Phe Lys Lys Leu Thr
241 TGT GTG CAC Cys Val Gir	I G ACC CGC C n Thr Arg Lo	256 TG AAG ATA eu Lys Ile	TTC GAG CAG Phe Glu Gin	271 GGT CTA CGG G Gly Leu Arg G	286 GC AAT TTC ACC A/ Iy Asn Phe Thr Ly	AA CTC AAG GGC G ys Leu Lys Gly A	301 GCC TTG AAC ATG Ala Leu Asn MET	316 ACA GCC AGC TAC TA Thr Ala Ser Tyr Ty	331 346 C CAG ACA TAC TGC CCC CCA ACT r GIn Thr Tyr Cys Pro Pro Thr
CCG GAA ACC Pro Glu Thr	361 G GAC TGT GA r Asp Cys G	AA ACA CAA lu Thr Gln	376 GTT ACC ACC Val Thr Thr	39 TAT GCG GAT T Tyr Ala Asp Pl	91 TC ATA GAC AGC CI he lle Asp Ser Le	406 TT AAA ACC TTT (eu Lys Thr Phe L	421 CTG ACT GAT ATC Leu Thr Asp Ile	436 CCC TTT GAA TGC AA Pro Phe Glu Cys Ly	451 A AAA CCA AGC CAA AAA TGA s Lys Pro Sor Gin Lys .
467 GGAAGCCCAG	477 GCCAGCTCTG	487 AATCCAGCTT	497 CTCAGACTGC	507 TGCTTTTGTG CC	517 SIT SAGCCAGO	527 537 GAA CTTGGAATTT C	547 CTGCCTTAAA GGGAC	557 567 CCAAGA GATGTGGCAC A	577 587 GCCACAGTT GGAAGGCAGT
597 ATAGCCCTCT	607 GAAAACGCTG	617 ACTCAGCTTG	627 GACAGCGGAA	637 GACAAACGAG AG	647 6 ATATTTTC TACTGAT	657 667 AGG GACCATTATA 1	677 ITTATTTATA TATTT	687 697 ATATT TTITAAATAT T	707 717 TATTTATTT ATTTATTTAT
727 TTTTGCAACT	737 CTATTTATTG	747 AGAATGTCTT	757 ACCAGAATAA	767 TAAA TTATTA AA	777	787 AAA 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 μ g), digested with either *Eco*RI, *Hind*III or *BgI*II, was electrophoresed on 0.8% agarose gel and transferred to nitrocellulose paper. The ³²P-labeled *Pstl/AhaIII* fragment of human GM-CSF cDNA (5 × 10⁷ c.p.m./ μ g) 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 NH₂-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, λ GM-CSF12, was subcloned in pUC13 (Vieira and Messing, 1982) to yield p λ GM-CSF12 (Figure 3A). A human placental genomic λ phage library was screened using a *PstI/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 λ HGM11-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 λ GM-CSF12. The region (shown in an expanded scale) that contains the GM-CSF gene of $p\lambda$ GM-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, *XbaI*; Xm, *Xnn1*. (B) Restriction endonuclease cleavage map and sequence analysis of the human GM-CSF gene. λ HGM11-a and λ HGM15-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, *Bg/*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-

10 20 30 40 50 60 70 80 90 100 110 120 130 140 AGTICCTGAT TCCACAGAGA GCTTCTGGAG AGGGAGGTGG GCTGGAGCAA GGGGCTGTGC AACAGACTAA GCCTGGGAGA ACTTGCCAGG GAAGCGGGGG TTGGGGCCGCA C<u>GTGTGTGTG TGTGTGTGT GTGTGTGTGT</u>G 150 160 170 180 190 200 210 220 230 240 250 260 270 280 CGCGCGCGGG CGTGGTGGCG TGCCTGGTTA TTGTGTTACA GCTAGTTACT ATGTTACAGC CAGCCTCAGA GACCCAGGTA TCCCATAATG GTACAGATAG CAGTAATGTG TCCTTTTTGA ATGGCAGGCT GCTCTGAAGA 290 300 310 320 330 340 350 360 370 380 390 400 410 420 GGGTAAAGAG GCTCACATAA CTCAAAGGAA GAGTCGCTTG GCAAAGGGGG AGGGCAGAAT CAGCAGTCAA ATGGGAGTAG CCTTACCTGC CCCAGCCAGA AGCTCTAGCC ATGCTGTGAA TTTTTAGGAT TGAACATGCC 430 440 450 460 470 480 490 500 510 520 530 540 550 560 TITGTAACTG GGATIGTAGC AGTCAGTIGG GATGGTATCA TAAGAGAACA TGGTATAGTC TGCCTGGCTT CTATACCCAT TGCTGTCATT CTCACTGCTC CCAAGGTGGC AGGGGTGGGG CTTGGCTGGC CAGTITGCCC 570 580 590 600 610 620 630 640 650 660 670 680 690 700 AGAATGGTCT CCTCAGTGGG AGTCTGTGGA GACCATTATA GAATGGTGTC ATTGGGCTGG ACCTTATTTA ATAGATAATG AGGTGGACTT GTGAGAAGTG TATATCTCAG AAGGTGGCTG GAAAGAGAAC GGGCCAGGAC 710 720 730 740 750 760 770 780 790 800 810 820 830 840 TGGGGCTGGA ATGAGCCACC AGAGTAGGTA GAGCTTGCCC AAAGGCCTCC AGGAACAGCA GGTGCTATGG AAGCAAGAGC CCCACTCAGT ATCTCCCAAA CCCCGCCCCA GCCACTCCAG GCCAGGAAAT CCAAATATGC 980 920 970 870 880 890 90.0 910 930 940 950 960 860 CTGGAGGCCC CTCAAAAAGG AGAGGCTAGC CAGAGGCTGG GTCAGACTGC CCAGGCAGGG TGGGAAAGCC CTTTAATCAG CCCGCAGGTG GGCTGCCAGT TCTTGGAAGG GCTTATTAAT GAAAACCCCC CAAGCCTGAC 990 1000 1010 1020 1030 1040 1050 1060 1070 1080 1090 1100 <u>1110</u> 1120 AACCTGGGGG AAGGCTCACT GGCCCCATGT ATAGCTGATA AGGGCCAGGA GATTCCACAA CTCAGGTAGT TCCCCCGCCC CCCTGGAGTT CTGTGGTCAC CATTAATCAT TTCCTCTAAC TGTGT<mark>ATA A</mark>GAGCTCTTT 1130 *1140 1150 1160 1182 1197 1212 1227 1242 TGCAGTGAGC CCAGTACTCA GAGAGAAAGG CTAAGGTCCT GAGGAGG ATG TGG CTG CAG AAT TTA CTT ITC CTG GGC ATT GTG GTC TAC AGC CTC TCA GCA CCC CGC TCA CCC ACC ACC + MET Trp Leu Gin Asn Leu Leu Phe Leu Giy IIe Val Val Tyr Ser Leu Ser Ala Pro Thr Arg Ser Pro IIe Thr 1257 1272 1287 1302 1317 1327 1337 1347 1357 1367 GTC ACC CGG CCT TGG AAG CAT GTA GAG GCC ATC AAA GAA GCC CTG AAC CTC CTG GAT GAC ATG CCT GTC ACG TTG GTGAGTGAGG GAGAGAGTTG GGTTGTTCCC AGGCCACCCA TTGGCCCTTG Val Thr Arg Pro Trp Lys His Val Giu Ala Lie Lys Giu Ala Leu Asn Leu Leu Asp Asp Met Pro Val Thr Leu 1377 1387 1397 1407 1429 1444 1466 1476 1486 1496 TCTTGCCCTG CCTTTCAGCT TGGTAACACG GCACTTCCTT TTTACAG AAT GAA GAG GTA GAA GTC GTC TCT AAC GAG TTC TCC TTC AAG GTAAGCTGCT TCTCTTTTGC TCTTCTCTT TGGTGTAGCT Asn Giu Giu Val Giu Val Val Ser Asn Giu Phe Ser Phe Lys 1506 1516 1526 1536 1546 1556 1566 1576 1586 1596 1606 1616 1626 1636 TCCCGGGGT<u>IG CCCTGCAGGG CTGCT</u>ITAAA TAGCCTCTIG CAAGGCTITC TACACTCT<u>IC CCTGCCAGGG GCTGCCAGGG GGAGGGGGGCA GAATCTGGGT AGTCAAGATC TCTATATCCA TICCTCTGTA GCAGTCACG</u> 1686 1696 1706 1716 1726 1736 1746 1756 1756 1766 1776 ATGTGCACTC TITITATCAA GTGAGCAATA AGTTIGGGCA GAGCTAGCTT TCAAAGGTCA GAGGTCTTCA GGGATTGATG GGAGCATGAG AGTCACTTTA 1666 1676 1646 1656 TGTGGCAGTC GCACTCTTT CGTCCAGA GCCATGCCCC 1786 1796 1806 1816 1826 1836 1846 1856 1866 1876 1886 1896 1906 1916 TGAGTIGGCC TIGTIGCCTT CAAGTICTCT CTGTGCCCCG TGACCATGGA GAGAAGGTAT TGATTATACA GGGCACAGA GGGTATIGAT TATACAGGGC ACAGAGGGGC TITIGAAATA GTGCTTCCCC ACACAGAAGT 1926 1936 1946 1956 1966 1976 1986 1996 2006 2016 2026 2036 2046 2056 TIGGCTCTGG CCATCTCCTT CTICTAGAAT GGCAGCAAGT AGCCTITGCT TACAAATTIT TITITTGGGG GGGGGACAGC ACTTGGAACT GGGAACCTCT ATGGTACCAG ACTATAAACA TGGCTGTGGT AGCTGAGGTG 2066 2076 2086 2096 2106 2116 2126 2136 2146 2156 2166 2176 2186 2196 GCCTGGACCA GGCAAGAAGA AGGGCAGTGC TCAGGACAGT GTTATGACTA GCCCCAACCT GTACCACTTG GTGGGAGCAG AAGGGAAGTA GCTTGTATTC TATGGACACT TGGCTTTTGC TTATCAACAA 2206 2228 2243 2258 2273 2288 2303 ATGGATGCTT CCCACAG AAG CTA ACA 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 Lys Leu Thr Cys Val Gin Thr Arg Leu Lys IIe Phe Giu Gin Giy Leu Arg Giy Asn Phe Thr Lys Leu Lys Giy Ala Leu Asn Met Thr Ala Ser Tyr 2318 2333 2349 2359 2369 2379 2389 2399 2409 2419 2429 2439 TAC CAG ACA TAC TGC CCC CCA ACT CCG GTGAGTACAC AGCATCAGGG CCCCAGGTCT TAGTCTGGGA GGGTGCGGCC AGTAAGGAGA GACTTCTGGC TGTAGCTGTT CAGGGCCCCA GTAAGTGGTT Tyr GIn Thr Tyr Cys Pro Pro Thr Pro 2449 2459 2469 2479 2489 2499 2509 2519 2529 2539 2549 2559 2569 2579 TCTGTATTGT TAGATGTTTC CTAGGAGCCC CCCAAGAACA GGCCATAGCT CCTATTCATC CCTAGACCCT AAGCTGAGAA ACAGCAGAGA GCCCACTGCT CTCCTAGGA GGAACCATCT GCCCATGTCA CACATGCCAT 2589 2599 2609 2619 2629 2639 2649 2659 2669 2679 2689 2699 2709 2719 TGTCAGTCAT GAGGTCAGAG GTGAGGCAAA CCCAAGGAAG CTCAGGGCCT GCCCAAGCCC AGGAGAGCCA AGAGCCAGCA CCCAGATGCC ACTTCCTGGC AGGACTTTCC TGTAAGGATG CCTGTATTCC ATGGATACAA 2729 2739 2749 2759 2769 2779 2789 2799 2809 2819 2829 2839 2849 2859 GGAGGAGGCC CCTAGAATGG GCAGACACCC CAAACATTCC TATTICCTTT CCTCATCAAA GACAGTGAGC CTGTGTCTT TTTAGAGGTG AGGAGATGCC TCAGGCTGGT TGCTTGCTTC TCGGTGGGAA CCAAAGAGGGC 2869 2879 2889 2899 2909 2919 2929 2939 2949 2959 2969 2979 2989 2999 AGACCTGAGA CAGTCCTTCT GTAAGGGCAC AGTCCCTGGG TGCCATTGGC AGAGCTGTGC ATGCCTTAGG CAGCACACGG GCAGCTATAG TGTCACCTGA TCCTACAAA TGGGCTTCTG GCCTCTGAGT TCTCAGCAGC 3009 3019 3029 3039 3049 3059 3069 3079 3096 3111 3126 GAGAGAGAAA TACACAGTTC ITTCCCCACG TGGCCCACTT ACCTGGACTC AAGTGCTGTT TATTTITITC CATTCCCTCC AG GAA ACG GAC TGT GAA ACA CAA GTT ACC ACC TAT GCG GAT TTC ATA Glu Thr Asp Cys Glu Thr Gin Val Thr Tyr Ala Asp Phe lie 3141 3156 3171 3186 3199 3209 3219 3229 3239 3249 GAC AGC CTT AAA ACC TTT CTG ACT GAT ATC CCC TTT GAA TGC AAA AAA CCA GGC CAA AAA TGA GGAAGCCCAG GCCAGCTCTG AATCCAGCTT CTCAGACTGC TGCTTTTGTG CCTGCGTAAT Asp Ser Leu Lys Thr Phe Leu Thr Asp IIe Pro Phe Giu Cys Lys Lys Pro Giy Gin Lys . 3259 3269 3279 3289 3299 3309 3319 3329 3339 3349 3359 3369 3379 3389 GAGCCAGGAA CTTGGAATTT CTGCCTTAAA GGGACCAAGA GATGTGGCAC AGCCACAGTT GGAAGGCAGT ATAGCCCTCT GAAAACGCTG ACTAGCTTG GACAGCGGAA GACAAACGAG AGATATTTTC TACTGATAGG

3539 3549 3559 3569 3579 3589 3599 3609 TITIGAAAGG GGAAAAATIT GGGCATAGGT GGAGTGGGGG AGCTATIGGG ATATGGTATT GATGAGAGTC AATGCTGTCA TG

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-

TTOTO ACACT CUCTOCACT	0 50	40	50	60	70	80	90	100	110
TICICAGAGI GGCIGCAGI	C TCGCTGCTGG	ATGTGCACAT	GGTGGTCATT	CCCTCTGCTC	ACAGGGGCAG	GGGTCCCCCC	TTACTGGACT	GAGGTTGCCC	CCTGCTCCAG
		0150 CTOCCCCACC	TOTOTOACAC	170	180	190	200	210	220
230 24	250	260	270	200	ACICAAGICI	TUTUTUAUAGI	GGUUAGAGAA	GAGGAAGGCT	GGAGICAGAA
TGAGGCACCA GGGCGGGCA	AGCCTGCCCA	AAGGCCCCTG	GGATTACAGG	CAGGATGGGG	AGCCCTATCT	AAGTGTCTCC		CCCAGCCATT	000
340 350	360	370	380	390	400	410	420	430	440
AAGTCCAAAC TGTGCCCCT	AGAGGGAGGG	GGCAGCCTCA	GGCCCAT TC A	GACTGCCCAG	GGAGGGCTGG	AGAGCCCTCA	GGAAGGCGGG	TGGGTGGGCT	GTCGGTTCTT
450 460	470	480	490	500	510	520	530	540	550
GGAAAGGIIC ATTAATGAA	ACCCCCAAGC	CTGACCACCT	AGGGAAAAGG	CTCACCGTTC	CCATGTGTGG	CTGATAAGGG	CCAGGAGATT	CCACAGTICA	GGTAGTTCCC
		590	500	610	620	630	640	650	660
670	GOICACCAIT	691		706	THECCAGIG	•721	CACAGAGAGA	736	751
G ATG TGG CTG CAG AG	CTG CTG CT	C TTG GGC A	CT GTG GCC 1	IGC AGC ATC	TCT GCA CC	с бос сво то	CG CCC AGC	CCC AGC ACG	CAG CCC TGG
MET Trp Leu Gin Sei	· Leu Leu Le	u Leu Gly T	hr Val Ala (Cys Ser Ile	Ser Ala Pr	o Ala Arg Se	er Pro Ser F	Pro Ser Thr	Gin Pro Trp
766		781		796		811 [×]		830	840
GAG CAT GTG AAT GCC	TC CAG GAG	GCC CGG CGT	CTC CTG AAG	C CTG AGT AG	GA GAC ACT	GCT GCT GAG	ATG GTAAGTO	GAGA GAATGTO	GGC
850 860		A La Arg Arg	Leu Leu Asr	Leu Ser Ai	"g Asp inr	Ala Ala Glu	MEI	20	047
CTGTGCTAGG CACCAGTGG	CCTGACTGGC	CACECCTETC	AGCTIGATAA	CATGACATIT	TCCTTTTCTA	CAG AAT GA	ACA GTA G	AA GTC ATC 1	
				0		Asn GI	u Thr Val G	lu Val Ile S	Ser Glu MET
	965	975	985	995	1005	1015	1025	1035	1045
TTT GAC CTC CAG GTAAC	SATGCT TCTCT	CTGAC ATAGCI	TTCC AGAAG	CCCT GCCCT	GGGGT GGAGG	TGGGG ACTCC	ATTTT AGATGO	GCACC ACACAG	GGTT
Phe Asp Leu Gin	1076	1005	1005						
GICCACITIC TOTOCAGIO		000	GTAGCAACTO	GGTGCTC		CCGTGCCCCCT	ATGGCAGTGA	CATGACCTCC	TTTATCACCT
1165 1175	1185	1195	1205	1215	1225	1235	1245	1255	1265
GAGCGGCCAT GGGCAGACCT	AGCATTCAAT	GGCCAGGAGT	CACCAGGGGA	CAGGTGGTAA	AGTGGGGGTC	ACTTCATGAG	ACAGGAGCTG	TGGGTTTGGG	GCGCTCACTG
1275 1285	1295	1305	1315	1325	1335	1345	1355	1365	1375
IGCCCCGAGA CCAAGTCCT	TTGAGACAGT	GCTGACTACA	GAGAGGCACA	GAGGGGTTTC	AGGAACAACC	CT TGCCCACC	CAGCAGGTCC	AGGTGAGGCC	CCACCCCCCT
	14U5	1415	1425	1435	1445	1455	1465	1475	1485
1495 150	1515	1525	1535	1545	1555	1565	1575	GAAGGGGCAGG	1505
CCATGGACAG GGCAGGGTCI	ATGACTGGAC	CCAGCCTGTG	CCCCTCCCAA	GCCCTACTCC	TGGGGGCTGG	GGGCAGCAGC	AAAAAGGAGT	GGTGGAGAGT	TOTIGTACCA
1605 1615	1625	1635			1657	1	672	168	10110110011
CTGIGGGCAC TICCCCACTO	0101000100								
CIGIOGOCAC TIGOCCACIO	CICACUGAUG	AACGACATTT	TCCACAG GAG	G CCG ACC TO	GC CTA CAG /	ACC CGC CTG	GAG CTG TAG	C AAG CAG GO	SC CTG CGG
1702	LICALLGALG	AACGACATTT	TCCACAG GAG	GCCGACCTO ProThrCy	SC CTA CAG /	ACC CGC CTG Thr Arg Leu	GAG CTG TAG Giu Leu Tyr	C AAG CAG GO Lys Gin Gi	C CTG CGG y Leu Arg
		AACGACATTT	TCCACAG GAG GI 1732	G CCG ACC TO J Pro Thr Cy	SC CTA CAG / /s Leu Gln 1747	ACC CGC CTG Thr Arg Leu	GAG CTG TAG Giu Leu Tyr 762	C AAG CAG GO Lys Gin Gi 1778	C CTG CGG y Leu Arg 1788
1702 GGC AGC CTC ACC AAG (Gly Ser Leu Thr Lys L	TC AAG GGC (AACGACATTT	TCCACAG GAO GIU 1732 ATG ATG GCO MET MET ALA	G CCG ACC TO Pro Thr Cy C AGC CAC TA Ser His Ty	GC CTA CAG / /s Leu Gln 1747 AC AAG CAG (// Lys Gln /	ACC CGC CTG Thr Arg Leu CAC TGC CCT HIS CVS Pro	GAG CTG TAG Giu Leu Tyr 1762 CCA ACC CCC Pro The Pro	C AAG CAG GG r Lys Gin Gi 1778 G GTGAGTGCCT	GC CTG CGG y Leu Arg 1788 ACGGCAGGGC
1702 GGC AGC CTC ACC AAG C Gly Ser Leu Thr Lys L 1798 1808	1717 TC AAG GGC (eu Lys Giy f 1818	AACGACATTT CCC TTG ACC Pro Leu Thr 1828	TCCACAG GAO GIU 1732 ATG ATG GCO MET MET AIa 1838	G CCG ACC TO Pro Thr Cy C AGC CAC TA Ser His Ty 1848	SC CTA CAG / /s Leu Gln 1747 NC AAG CAG (/r Lys Gln) 1858	ACC CGC CTG Thr Arg Leu CAC TGC CCT His Cys Pro 1868	GAG CTG TAC Giu Leu Tyr 762 CCA ACC CCC Pro Thr Pro 1878	C AAG CAG GG Lys Gin Gi 1778 G GTGAGTGCCT D 1888	C CTG CGG y Leu Arg 1788 ACGGCAGGGC 1898
1702 GGC AGC CTC ACC AAG (Gly Ser Leu Thr Lys L 1798 1800 CTCCAGCAGG AATGTCTTAA	1717 TC AAG GGC (eu Lys Gly F 1818 TCTAGGGGGT	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA	TCCACAG GAO GIU 1732 ATG ATG GCO MET MET AIa 1838 TGGGGAGAGA	G CCG ACC TG J Pro Thr Cy C AGC CAC TA Ser H1s Ty 1848 TCTATGGCTG	GC CTA CAG / /s Leu Gln 1747 AC AAG CAG (/r Lys Gln) 1858 TGGCTGTTCA	ACC CGC CTG Thr Arg Leu CAC TGC CCT HIS CyS Pro 1868 GGACCCCAGG	GAG CTG TAC Glu Leu Tyr 762 CCA ACC CCC Pro Thr Pro 1878 GGGTTTCTGT	C AAG CAG GC Lys Gln Gl 1778 G GTGAGTGCC1 D 1888 GCCAACAGTT	C CTG CGG y Leu Arg 1788 ACGGCAGGGC 1898 ATGTAATGAT
1702 GGC AGC CTC ACC AAG (GIy Ser Leu Thr Lys L 1798 1800 CTCCAGCAGG AATGTCTTAA 1908 1918	1717 TC AAG GGC (eu Lys Gly f 1818 TCTAGGGGGT 1928	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA 1938	TCCACAG GAO GIU 1732 ATG ATG GCC MET MET AIa 1838 TGGGGAGAGA 1948	G CCG ACC TO J Pro Thr Cy C AGC CAC TA Sor H1s Ty 1848 TCTATGGCTG 1958	SC CTA CAG is Leu Gln 1747 AC AAG CAG (ir Lys Gln 1858 TGGCTGTTCA 1968	ACC CGC CTG Thr Arg Leu CAC TGC CCT HIS Cys Pro 1868 GGACCCCAGG 1978	GAG CTG TAC Glu Leu Tyr 762 CCA ACC CCC Pro Thr Pro 1878 GGGTTTCTGT 1988	C AAG CAG GC Lys GIn GI 1776 G GTGAGTGCCT 0 1888 GCCAACAGTT 1998	C CTG CGG y Leu Arg 1788 ACGGCAGGGC 1898 ATGTAATGAT 2008
1702 GGC AGC CTC ACC AAG (GIY Ser Leu Thr Lys L 1798 1800 CTCCAGCAGG AATGTCTTA/ 1908 1918 TAGCCCTCCA GAGAGGAGG	1717 TC AAG GGC (eu Lys Gly f 1818 TCTAGGGGGT 1928 AGACAGCCCA	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA 1938 TTTCATCCCA	TCCACAG GAG GI 1732 ATG ATG GCC MET MET AIa 1838 TGGGGAGAGA 1948 AGGAGTCAGA	GCCG ACC TO Pro Thr Cy AGC CAC TA Ser H1s Ty 1848 TCTATGGCTG 1958 GCCACAGAGC	GC CTA CAG / /s Leu GIn 1747 IC AAG CAG (/r Lys GIn) 1858 TGGCTGTTCA 1968 GCTGAAGCCC	ACC CGC CTG Thr Arg Leu CAC TGC CCT 11s Cys Pro 1868 GGACCCCAGG 1978 ACAGTGCTCC	GAG CTG TAC Glu Leu Tyr 762 CCA ACC CCC Pro Thr Pro 1878 GGGTTTCTGT 1988 CCAGCAGGAG	C AAG CAG GG Lys Gin Gi G GTGAGTGCCT 1888 GCCAACAGTT 1998 CTGCTCCTAT	C CTG CGG y Leu Arg 1788 ACGGCAGGGC 1898 ATGTAATGAT 2008 CCTGGTCATT
GC AGC CTC ACC AAG (Gly Ser Leu Thr Lys L 1798 1800 CTCCAGCAGG AATGTCTTA/ 1908 1918 CTCCAGCAGG AGAGGGG 2018 2022 ATTGCCTTCA GGGTAATGA	1717 TC AAG GGC (eu Lys Gly f 1818 TCTAGGGGGT 1928 AGACAGCCA 2038 GGTCAGCGGT	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA 1938 TTTCATCCA 2048 GAGGGCAAAC	TCCACAG GAG GIU 1732 ATG ATG GCC MET MET AIa 1838 TGGGGAGAGA 1948 AGGAGTCAGA 2058	GCCG ACC TO Pro Thr Cy AGC CAC TA Ser H1s Ty 1848 TCTATGGCTG 1958 GCCACAGAGC 2068	GC CTA CAG / /s Leu GIn 1747 IC AAG CAG (/r Lys GIn) 1858 TGGCTGTTCA 1968 GCTGAAGCCC 2078	ACC CGC CTG Thr Arg Leu 1 CAC TGC CCT 41s Cys Pro 1868 GGACCCCAGG 1978 ACAGTGCTCC 2088 ACAGGAGAGTG	GAG CTG TAC Glu Leu Tyr 762 CCA ACC CCC Pro Thr Pro 1878 GGGTTTCTGT 1988 CCAGCAGGAG 2098	C AAG CAG GG Lys GIn Gi 1778 G GTGAGTGCCT 1888 GCCAACAGTT 1998 CTGCTCCTAT 2108	C CTG CGG y Leu Arg 1788 ACGGCAGGGC 1898 ATGTAATGAT 2008 CCTGGTCATT 2118 TCGCCAGGA
1702 GGC AGC CTC ACC AAG C GIY Ser Leu Thr Lys L 1798 1800 CTCCAGCAGG AATGTCTTA/ 1908 1918 TAGCCTCCA GAGAGGAGG 2018 2028 ATTGTCATTA CGGTTAATG 2128 2158	1717 TC AAG GGC (eu Lys Gly f 1818 TCTAGGGGG T 1928 AGACAGCCCA 2038 GGTCAGAGGT 2148	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA 1938 TTTCATCCCA 2048 GAGGGCAAAC 2158	TCCACAG GAG GIU 1732 ATG ATG GCC MET MET AIa 1838 TGGGGAGAGA 1948 AGGAGTCAGA 2058 CCAAGGAAAC 2168	G CCG ACC TC Pro Thr Cy C AGC CAC TA Ser H1s Ty 1848 TCTATGGCTG 1958 GCCACAGAGC 2068 TTGGGGCCTG 2178	C CTA CAG /s Leu GIn 1747 IC AAG CAG (r Lys GIn) 1858 TGGCTGTTCA 1968 GCTGAAGCCC 2078 CCCAAGGCCC 2188	ACC CGC CTG Thr Arg Leu 1 CAC TGC CCT 41s Cys Pro 1868 GGACCCCAGG 1978 ACAGTGCTCC 2088 AGAGGAAGTG 2198	GAG CTG TA(Glu Leu Tyr 762 CCA ACC CCC Pro Thr Pro 1878 GGGTTTCTGT 1988 CCAGCAGGAG 2098 CCCAGGCCCA 2208	C AAG CAG GG Lys GIn Gi GTGAGTGCCT 1888 GCCAACAGTT 1998 CTGCTCCTAT 2108 AGTGCCACCT 2218	C CTG CGG y Leu Arg 1788 ACGGCAGGGC 1898 ATGTAATGAT 2008 CCTGGTCATT 2118 TCTGGCAGGA 2228
1702 GC AGC CTC ACC AAG (Gly Ser Leu Thr Lys L 1798 1800 CTCCAGCAGG AATGTCTTA/ 1908 1918 TAGCCCTCCA GAGAGGAGG 2018 2022 ATTGTCATTA CGGTTAATG/ 2128 2138 CTTTCCTCTG GCCCCACATG	1717 TC AAG GGC (eu Lys GI) f 1818 TCTAGGGGGT 1928 AGACAGCCA 2038 GGTCAGAGGT 2148 GGGTGCTTGA	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA 1938 TTTCATCCCA 2048 GAGGGCAAAC 2158 ATTGCAGAGGG	TCCACAG GAG GIU 1732 ATG ATG GCC MET MET AIZ 1838 TGGGGAGAGA AGGAGTCAGA 2058 CCAAGGAAAC 2168 ATCAAGGAAG	GCCG ACC TC Pro Thr Cy AGC CAC TA Ser H1s Ty 1848 TCTATGGCTG 1958 GCCACAGAGC 2068 TTGGGGCCTG 2178 GGAGGCTACT	GC CTA CAG , rs Leu GIn ⁻ 1747 IC AAG CAG (r Lys GIn) 1858 TGGCTGTTCA 1968 GCTGAAGCCC 2078 CCCAAGGCCC 2188 TGGAATGGAC	ACC CGC CTG Imr Arg Leu CAC TGC CCT 11s Cys Pro 1868 GGACCCCAGG 1978 ACAGTGCTCC 2088 AGAGGAAGTG 2198 AAGGACCTCA	GAG CTG TAG Glu Leu Tyn 762 CCA ACC CCC Pro Thr Pro 1878 GGGTTTCTGT 1988 CCAGCAGGAG 2098 CCCAGGCCCA 2008 GGCACTCCTT	C AAG CAG GG Lys GIn GI 1778 G GTGAGTGCCT 1888 GCCAACAGTT 1998 CTGCTCCTAT 2108 AGTGCCACCT 2218 CCTGCGGGAA	CCTG CGG y Leu Arg 3 1788 ACGGCAGGGC 1898 ATGTAATGAT 2008 CCTGGTCATT 2118 TCTGGCAGGA 2228 GGGAGCAAAG
GC AGC CTC ACC AAG (Gly Ser Leu Thr Lys L 1798 1800 CTCCAGCAGG AATGTCTTA/ 1908 1918 CTCCAGCAGG AAGAGGG 2018 2022 ATTGTCATTA CGGTTAATG/ 2128 2138 CTTTCCTG GCCCCACATG CTTTCCCTG GCCCCACTG	1717 TC AAG GGC (eu Lys Gly f 1818 TCTAGGGGGG 1928 AGACAGCCA 2038 GGTCAGAGGT 2148 GGGTGCTTGA 2258	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA 1938 TTTCATCCCA 2048 GAGGGCAAAC 2158 ATTGCAGAGG ATTGCAGAGG 2268	TCCACAG GAG GI U 1732 ATG ATG GCC MET MET A12 1838 TGGGGAGAGA 1948 AGGAGTCAGA 2058 CCAAGGAAAC 2168 ATCAAGGAAG	GGAGCCTCC Pro Thr Cy AGC CAC TA Ser HIS Ty 1848 TCTATGGCTG 1958 GCCACAGAGC 2068 TTGGGGCCTG 2178 GGAGGCTACT 2288	GC CTA CAG / rs Leu GIn 1 1747 AC AAG CAG (rr Lys GIn) 1858 GCGGAAGCCC 2078 CCCAAGGCCC 2188 TGGAATGGAC 2298	ACC CGC CTG Thr Arg Leu CAC TGC CCT 115 Cys Pro- 1868 GGACCCCAGG 1978 ACAGTGCTCC 2088 AGAGGAAGTG 2198 AAGGACCTCA 2308	GAG CTG TAG G1u Leu Tyr 762 CCA ACC CCC Pro Thr Pro 1878 GGGTTTCTGT 1988 CCAGCAGGAG 2098 CCCAGGCCCA 2208 GGCACTCCTT 2318	C AAG CAG GG Lys GIn GI 1778 G GTGAGTGCC1 D 1888 GCCAACAGTT 1998 CTGCTCCTAT 2108 AGTGCCACCT 2218 AGTGCCACCT 2218 CCTGCGGGAA 2328	C CTG CGG y Leu Arg 1788 ACGGCAGGGC 1898 ATGTAATGAT 2008 CCTGGTCATT 2118 TCTGCCAGGA 2228 GGGAGCAAAG 2338
1702 GGC AGC CTC ACC AAG C GI y Ser Leu Thr Lys L 1798 B806 CTCCACCAGG AATGCTTAA 1908 1918 2018 2022 ATGCTCATCA GAGGAGG 2018 2022 ATGTCATTA CGGTTAATG 2128 2138 CTTTCCTCT6 GCCCCCACTC 2238 2248 TTTGTGGCCT TGACTCCACT	1717 TC AAG GGC u eu Lys Gly i 1818 TCTAGGGGT 1928 AGACAGCCCA 2038 GGTCAGAGGT 2148 GGGTGCTTGA 2258 CCTTCTGGGT	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA 1938 TTTCATCCCA 2048 GAGGCCAAAC 2158 ATTGCACAGG 2268 GCCCAGACAC	TCCACAG GAG GIU 1732 ATG ATG GCC MET MET AIL 1838 TGGGAGAGA 1948 AGGAGTCAGA 2058 ACCAAGGAAC 2168 ATCAAGGAAG 2278 GACCTCACCC	 GCG ACC TG Pro Thr Cy AGC CAC TA Ser HIS Ty 1848 TCTATGGCTG 1958 GCCACAGAGC 2068 TTGGGGCCCTG 2178 GGAGGCTACT 2288 CAGCTGCCCTA 2288 	C CTA CAG / /s Leu GIn 1 1747 C AAG CAG (/r Lys GIn 1 1858 TGGCTGTTCA 1968 GCTGAAGCCC 2078 CCCAAGGCC 2188 TGGAATGGAC 2298 GCTCTGCCT	ACC CGC CTG Thr Arg Leu CAC TGC CCT 115 Cys Pro 1868 GGACCCAGG GGACCCAGG 2198 AAGGGAACTG 2308 GGGACCCAAA 2308	GAG CTG TAG G1u Leu Tyr 762 CCA ACC CCC Pro Thr Pro 1878 GGTTTCTGT 1988 CCAGCAGGAG 2098 GCCAGGCCCA 2208 GGCACTCCTT 2318 AGGCAGGCGT	C AAG CAG GG Lys GIn GI 1776 G GTGAGTGCC1 1888 GCCAACAGTT 1998 CTGCTCCTAT 2108 AGTGCCACCT 2218 CCTGCGGGAA 2328 TTGACTGCCC	C CTG CGG y Leu Arg 1788 ACGCAGGCC 1898 ATGTAATGAT 2008 CCTGGTCATT 2118 TCTGGCAGGA 2228 GGGACCAAG 2338 AGAAGGCCAA
1702 GGC AGC CTC ACC AAG C GIY Ser Leu Thr Lys L 1798 1800 CTCCAGCAGG AATGTCTTA/ 1908 1918 TAGCCCTCCA GAGAGGAGG 2018 2022 ATTGTCATTA CGGTTAATG 2128 2138 CTTTCCTCTG GCCCCACATC 2348 2358 CTTGGCCT TGACTCCACT 2348 2358	1717 TC AAG GGC 1 eu Lys Gly 1 1818 AGACAGCCCA 2038 GGTCAGAGGC 2148 GGGTGCTTGA 2258 CCTTCTGGGT 2368 CCTTCTGGGT 2368 CGTCTTG	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA 1938 TTTCATCCAA 2048 GAGGGCAAAC 2158 ATTGCAGAGG 2268 ATTGCAGAGG 2378 ACTCTGCGTG	TCCACAG GAG GII 1732 ATG ATG GCC MET MET AI2 1838 AGGGAGAGAA 2058 CCAAGGAAAG 2278 ATCAAGGAAAG 2278 ATCAAGGAAG 2278 CCACGGCAGC	 CCG ACC TG Pro Thr Cy AGC CAC TH Ser H1s Ty 1848 TCTATGGCTG 2068 GCCACAGCC 2078 GGAGGCTACT 2178 GGAGGCTACT 2398 CACCTGCCCT 2398 CACCTGCCCC 	CCTA CAG IT47 IT47 IC AAG CAG (/r Lys GIn) 1858 IGGCTGTTCA 1968 GCTGAAGCCC 2188 IGGAATGGAC 2298 GCTCTGCCCT 2408 GCTCTGCCCT	ACC CGC CTG Thr Arg Leu CAC TGC CCT 115 Cys Pro 1868 GGACCCCAGG 1978 ACAGTGCTCC 2088 AGAGGAACTG 2198 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308	GAG CTG TAG G1u Leu Tyr 762 CCA ACC CCC Pro Thr Pro 1878 GGGTTTCTGT 1988 CCAGCAGGAG 2098 CCAGCAGGAG 2098 CCAGCAGCAG 2098 GGCACTCCTT 2318 AGGCAGCCGT 2428	C AAG CAG GG Lys GIn GI 177E G GTGAGTGCCT 1888 GCCAACAGTT 1998 CTGCTCCTAT 2108 AGTGCCACCT 2218 CCTGCGGGAA 2328 CTGCCGGGAA 2328 CCGCGGGAA 2328 CCGCGGGAA 2328 CCGCCGCGAA 2328 CCGCCGCGAA 2438 CCACCCCACC	CCTG CGG y Leu Arg 1788 ACGCAGGGC 1898 ATGTAATGAT 208 CCTGGCAGGA CCTGGCAGGA 2228 GGGAGCAAAG 2338 GGAGCCAA 2488 2485 2455 2485 2
1702 GGC AGC CTC ACC AAG (Gly Ser Leu Thr Lys L) 1798 1800 CTCCAGCAGG AATGTCTTA/ 1908 1918 CTCCAGCAGG AATGTCTTA/ 1908 2018 2018 2018 2128 217CTCCTG GCCCACATA 2238 2248 TTTGTCCTTG GCCCACTT 2340 2358 CCTCAGGCTG GCACTTAGTG 2458 2458	1717 TC AAG GGC 1 1818 TCTAGGGGT 1928 AGACAGCCA 2038 GGTCAGAGGT 2148 CGGTGCTTGA 2258 CCTCTGGGT 2368 CAGCCCCTTG 2368 CAGCCCCTTG	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA 1938 TTTCATCCCA 2048 GAGGGCAACC 2158 GCCCAGAGC 2268 GCCCAGAGAC 2378 ACTCTGGCTG 2488	TCCACAG GAG GI (1732 ATG ATG GCC MET MET AI2 1838 TGGGAGAGAA 2058 CCAAGGAAG 2278 GACCTCAGCC 2388 CCACTGGCAG 2498	CCG ACC TC J Pro Thr C) C AGC CAC TA 1848 TCTATGGCTG 1958 GCCACAGAGC 2068 TTGGGGCCTG 2178 CAGCTGCCCT 2398 AGCTATGCAC 2508	C CTA CAG / /s Leu GIn 1 1747 C AAG CAG (/r Lys GIn 1 1858 TGGCTGTTCA 1968 GCTGAAGCCC 2188 GCTCAAGGCC 2298 GCTCTGCCA 2298 GCTCTGCCA 2408 TCCTTGGGAA 2518	ACC CGC CTG Thr Arg Leu 12AC TGC CCT 115 Cys Pro 1868 GGACCCAGG 1978 ACAGTGCTCC 2088 AGAGGAAGTG 2198 GGGACCTCA 2308 GGGACCTCA 2418 ACGACGTGGGT 2528	GAG CTG TAX Glu Leu Tyr 762 CCA ACC CCC Pro Thr Prr 1988 CCAGCAGGAG GGATTCTGT 1988 CCAGCAGGAG GGACTCCTT 2318 AGCAGGCGT 2428 GGCAGCAGCG GGCAGCAGCG 2538	C AAG CAG GG Lys GIn GI 1776 G GTGAGTGCCT 1888 GCCAACAGTT 1998 CTGCTCCTAT 2108 AGTGCCACCT 2218 AGTGCCACCT 2218 AGTGCCACCT 2238 TTGACTGCCC 2438 TCACCTCACC	C CTG CGG y Leu Arg 1788 ACGGCAGGGC 1898 ATGTAATGAT 2008 CCTGGTCATT 2118 TCTGGCAGGA 238 AGAAGGCCAAG 238 AGAAGGCCAA 248 CAGGTCAGTG
1702 GGC AGC CTC ACC AAG C GI y Ser Leu Thr Lys L 1798 B80 CTCCAGCAGG AATGTCTTA/ 1908 1918 CTTCCAGCAGGAGGGC 2018 2022 ATGTCATTA CGGTTAATG/ 2128 2138 CTTTCCTCT6 GCCCCACATC 2238 2248 TTGTGGCCT TGAATGGCC 2348 2358 CCTCAGGCTG GCACTTAAGT 2458 2456 GGTGGTGCT GGAATGGGCC	1717 TC AAG GGC (eu Lys GIy I 1818 TCTAGGGGT 1928 AGACAGCCCA 2038 GGTCAGAGGT 2148 GGTGCTTGA 2258 CCTTCTGGGT 2368 CAGGCCCTTG 2478 CAGGCCCTTG 2478	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA 1938 GTTCATCCCA 2048 GAGGCCAACA 2158 ATTCCAGAGG 2268 GCCCAGACACC 2378 ACTCTGGCTG 2488 CGACTTCTAA	TCCACAG GAC GII 1732 ATG ATG GCC MET MET AII 1838 AGGAGTCAGA 2058 CCAAGGAAGC 2168 ATCAAGGAAG 2278 GACCTCAGCC 2388 CCACTGCCAG 2498 GAGCAGTAG	2 CCG ACC TC J Pro Thr C; 2 AGC CAC TJ 3 Ser HIS T; 1848 GCCACAGCC 2068 TTGGGGCTACT 2288 CAGCTGCCCT 2398 AGCTATGCAC 2508 AGAAACATGC	C CTA CAG) 's Leu Gin 1 1747 C AAG CAG (Gr (C C C C C C C C C C C C C C C C C C C	ACC CGC CTG Thr Arg Leu ICAC TGC CCT 115 Cys Pro 1868 GGACCCCAGG 2088 AGAGGACTCA 2198 AAGAGCACTCA 2308 GGGACCAAAA 2418 ACACGTGGGT 2528 ACACGTGGGT 2528	GAG CTG TAX Glu Leu Tyr 762 CCA ACC CCC Pro Thr Prc 1878 GGGTTICTGT 1988 CCAGCAGGAG 2098 CCCAGGCCCA 2208 GGCATCCTT 2428 GGCAGCAGCG 2538 GGCAGCAGCG 2538 CCCAGCTCA 2428 CCCAGCCCA	C AAG CAG GC Lys GIn GI G GGAGTGCCD 1888 CGCAACAGTT 1998 CTGCTCCTAT 2108 AGTGCCACCT 2218 CCTGCGGAA CZ328 TTGACTGCCC 2438 TCACCTGACC 2548 TCACCTGACC	CC CTG CGG y Leu Arg 1788 ACGGCAGGGC 1898 ATGTAATGAT 2008 CCTGGTCATT 2018 CCTGGTCATT 2018 CCTGGTCATT 2018 CCTGGTCATA 2228 GGGACCAAG 2338 AGAAGGCCAA 2428 CAGGTCACTG 2558 AGATGTTTTT
1702 GGC AGC CTC ACC AAG C GI y Ser Leu Thr Lys L 1798 1806 CTCCAGCAGG AATGTCTTA/ 1908 1916 TAGCCTCCAC AGAGGGAGG 2018 2018 2018 2018 2018 2018 2138 CTTGCCTCG CCCCACATG 238 2248 TTGTGGCCT TGACTCACT 2348 2358 2458 2458 2458 2458 2458 2458 2558	1717 TIC AAG GGC (eu Lys Gfy f 1818 AGACAGCCA 2038 GGTAGAGAGT 2148 GGGTGCTTGA 2258 CCTTCTGGGT 2348 CCTTCTGGGT 2348 CCTTCTGGGT 2478 CCGGCCCTG	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA 1938 TTTCATCCAA 2048 GAGGGCAAAC 2158 ATTGCAGAGG 2268 GCCCAGAGAC 2378 ACTCTGCTG 2488 TGAGTTCTAA 2592	TCCACAG GAG GIU 1732 ATG ATG GCC MET MET AI2 1838 AGGGAGCAGA 1948 AGGGGAGCAGA 2058 CCACGGGAAC 2168 ATCAAGGAAG 2278 GACCTCAGCC 2388 CCACTGGCAG CACTGGCAG GAGGCAGTAG	 CCG ACC TG	C CTA CAG : 's Leu GIn : 1747 C AAG CAG (Cr Lys GIn) 1858 GGCTGTTCA 1968 GCTGAAGCCCC 2188 TGGAATGGAC 2298 GCCCAGGCCC 2408 TCCTTGGGGA 2518 TGGTGCTCCC 2408	ACC CGC CTG Thr Arg Leu CAC TGC CCT 115 Cys Pro 116 Cys Pro 1868 GGACCCCAGG 2198 AGAGGAAGTG 2198 AAGGACCTCA 2308 AAGGACCTCA 2308 GGGACCAAAA 2418 ACAGGTGGGT 2528 TTCCCCCACG 2522	GAG CTG TAX Glu Leu Tyr 762 CCA ACC CCC Pro Thr Prc 1878 GGGTTTCTGT 1988 CCACAGGAG 2098 CCCAGGCCG 2208 GGCACTCCTT 2318 GGCAGCGGT 2428 GGCAGCGC 2538 TTACCACT 263	C AAG CAG GC Lys GIn GI GGGAGTGCCT 177E GGCAACACTT 1998 CTGCTCCTAT 2108 AGTGCCACCT 2218 CCTGCGGGAA 2528 TCACCTGACC 2438 TCACCTGACC 2548 GCCTGGGACTC 37	CC CCG CGG y Leu Arg 1788 ACGGCAGGGC 1898 ATGTAATGAT 2008 CCTGGTCATT 2118 TCTGGCAGGA 2228 GGGAGCAAGG 2388 CAGGTCAATG 2558 AAGTGTTTTT 2552
GC AGC CTC ACC AAG C Gly Ser Leu Thr Lys L 1798 1800 CTCCAGCAGG AATGCTTA 1908 1800 CTCCAGCAGG AATGCTTA 1908 1800 CTCCAGCAGG AATGCTTA 1908 2022 CTTCCCTCC GGAGAGAGG 2018 2022 CTTCCCTCG GCCCACTA 2128 2150 CTTCGCCT GGCCCACTA 2348 2358 CCTCAGGCTG GCACTAAGG 2458 2466 GGTGTGTCCT GGAGTGGGCC 2568 TATTTTCCT TTTTTAAG	1717 TC AAG GGC (eu Lys Giy) CTCTAGGGGGT 1928 GGTCAGAGGT 2038 GGTCAGAGGT 2148 GGGTCCTTCGGGT 2258 CCTCTGGGCTT 2368 CAGGCCCTTG 2478 TCCTGGGCTC GAA ACT TCC	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA 1938 GAGGGCAAAC 2048 ATTGCACGACA 2048 ATTGCACGACA 2268 GCCCAGAGAC 2378 ACTCTGCTG 2488 TGAGTTCTAA 2592 TGT GCA ACC	TCCACAG GAC GII 1732 ATG ATG GCC MET MET AIZ 1838 TGGGGAGAGA 1948 ATGAGGATCAGA 2058 ATCAAGGAAAC 2168 ATCAAGGAAAC 278 GACCTCAGCC 2388 CCACTGGCAG 2498 GAGGCAGTAG	 CCG ACC TC Pro Thr C; CAGC CAC TA Ser H1s T; 1848 TCTATGECTG 1958 TCTATGECTG 2068 CCACAGACC 2178 GGAGECCTG CAGC CACAGC 2288 CAGCTGCCT 2398 AGCTATGEAC 2508 AGAAACATGC 2607 C ACC TTT CC 	C CTA CAG / /s Leu GIn / 1747 C AAG CAG (/r Lys GIn) 1858 TGGCTGTTCA 1968 GCTGAAGCCC 2188 TGGAATGAC 2298 GCTCAGCCCT 2408 TCCTTGGGAA 2518 TGGTGCTCCC CAA AGT TTC	ACC CGC CTG Thr Arg Leu CAC TGC CCT 11s Cys Pro 1868 GGACCCCAGG 2088 AGAGGAAGTG 2088 AAGAGAAGTG 2198 AAGGACCTCA 2308 GGGACCAAAA 2418 ACAGTGGGT 2528 TTCCCCCACG 2622 AAA GAG AAC	GAG CTG TAX GIU Leu Tyr 762 CCA ACC CCC Pro Thr Prc 1878 GGGTTTCTGT 1988 CCACGCAGGAG 2098 CCCAGGCCCA 2208 GGCACTCTT 2318 AGCCAGGCGT 2428 GGCAGCACCG 2538 TTACCCACTT 261 CCG AAG G	C AAG CAG GG Lys GIn GI 1776 G GTGAGTGCCT 1888 GCCAACAGTT 2108 AGTGCCACCT 2188 AGTGCCACCT 2188 AGTGCCACCT 2388 CTGCGGGAA C 2548 GCCTGGACTC 57 AG TIT CTG CC	CCTG CGG y Leu Arg 1788 ACGCCAGGGC 1898 ATGTAATGAT 2008 CCTGGCAGGA 2118 TCTGGCAGGA 2338 AGAAGGCCAA 2448 CAGGTCAGTG 2558 AAGTGTTTTT 2652 TT GTC ATC
1702 GGC AGC CTC ACC AAG (GIY Ser Leu Thr Lys L) 1798 1800 CTCCAGCAGG AATGTCTTA/ 1908 1918 CTCCAGCAGG AATGTCTTA/ 1908 1918 COTCCAGCAGG AATGTCTTA/ 1908 2018 2018 2018 2018 2128 CTTCCCTG GCCCCACATG 2238 2248 CTTGCCTT GACTCACT 2348 2350 CCTCAGGCTG GCACTTAAG 2458 GGTGTGTCCT GGAGTGGGCC 2568 TATTITCTT TTTTTAAG 2424	1717 TC AAG GGC (eu Lys Giy f 1818 TCTAGGGGT 1928 GGTCAGAGGCT 2038 GGTCAGAGCTTG 2478 CCTTCTGGGT 2478 CCTGCGCCT GAA ACT TCC GI u Thr Ser	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGCCGACA 1938 GAGGGCCACA 2048 GAGGGCAAAC 2158 ATTCCACAGGG 2268 GCCCAGAGACC 2488 TGACTCTAA 2592 TGT GCA ACC Cys Ala Thr 2600	TCCACAG GAC GII 1732 ATG ATG GCC MET MET AI2 1838 CCAGGAGAGAA 2058 CCAAGGAAAC 2168 ATCAAGGAAG 2278 GACCTCAGCC 2398 CCACTGGCAG 2498 GAGGCAGTAG CAG TT AT	2 AGC ACC TC J Pro Thr C; AGC CAC TA Ser HIS T; 1958 CTATGCCTG 1958 TTGGGCCCC 2088 TTGGGGCCCC 2178 GGAGGCTACT 2288 CACCTGCCCT 2398 AGCTATGCAC 2508 AGAAACATGC 2607 C ACC TTT C 9 Thr Phe C 707	CCTA CAG) s Leu Gin : 1747 CAAG CAG (CG (CC) 1858 GCTGAAGCCC 2078 CCCAAGCCCC 2188 GCTGAAGCCC 2188 GCTCAGCCC 2288 GCTCTGCCT 2408 TCCTTGGGGA 2518 TCGTGCTTCC 2408 TCCTTGGGGA 2518 TCGTGCTTCC 2408 TCCTTGGGGA 2518 TCCTTGCGGA 2518 TCCTTGCGCA 2518 TCCTTGCGCA 2518 TCCTTGCGCA 2518 TCCTTGCGCA 2518 TCCTTGCCA 2518 TCCTTCC 258 TCCTTCC 258 TCCTTCC 258 TCCTTCC 258 TCCTCC 258 TCCC 258 TCCTCC 258 TCCC	ACC CGC CTG Thr Arg Leu CAC TGC CCT 115 Cys Pro 115 Cys Pro 115 Cys Pro 115 Cys Pro 1978 ACAGTGCTCC 2088 AGAGGAACTCA 2198 AAGGAACTCA 2308 GGGACCAAAA 2418 ACACGTGGGT 2528 TCCCCCACG 2622 AAG GAG AAC Lys GI u Ass	GAG CTG TAX Glu Leu Tyr 762 CCA ACC CCC Pro Thr Prc 1878 CCGGCTTCTGT 1988 CCGCAGGAG CCCAGGAG GGCACCAGGAG 2208 GGCAGCAGCG 2218 AGGCAGCAGCG 2538 GGCAGCAGCG 2538 CTG AAG GC CTG AAG GC Leu Lys A	C AAG CAG GC Lys GIn GI G GTGAGTGCCT 1776 G GTGAGTGCCT 1998 CTGCTCCTAT 2108 AGTGCCACCT 2218 CCTGCGGGAA 2328 TTGACTGCCC 2438 TCACCTGACC 2548 CCCTGGACTC 37 AC TTT CTG C 59 Phe Lag	CCTG CGG y Leu Arg 1788 ACGCACGGCC 1898 ACGCACGGC ATGTAATGAT 2008 CCTGGTCATT 2118 TCTGGCAGGA 2228 GGGAGCAAAG 2338 AGAAGGCCAA 248 CAGGTCACTG 258 AGAAGGCCAA 258 AGAAGGCCAA 258 AGAAGCCAA 258 AGAAGCCAA 258 AGAAGCCAA 258 AGAAGCCAA 258 AGAAGCCAA 258 AGAAGCCAA 258 AGAAGCCAA 258 AGAAGCCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 AGAACGCAA 258 258 258 258 258 258 258 258
1702 GGC AGC CTC ACC AAG C GI y Ser Leu Thr Lys L 1798 1806 CTCCACCAGG AATGCTTAA 1908 1918 2018 2018 2018 2128 2128 2128 2128 2128 2128 2238 2246 TITGTGCCT TGACTCACT 2348 2458 CCTCAGGCTG GCACTTAAG 2458 CATCT GGAGTGGCCC 2568 TATTITCTT HTHTAAG 2667 CCT TI GAC TGC TGC TG	1717 TC AAG GGC (eu Lys GIy f 1818 1618 16276 1628 16276 1628 16276 1628 16276 1628 16276 1628 16276 1628 16276 1628 16276 1628 16276 1628 16276 1628 16276 1628 16276 1628 16276 16286 16276 162866 16286 16286 16286 16286 16	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA 1938 TTCATCCCA 2048 GAGGCCAACA 2158 ATTCCACAGGG 2268 GCCCAGAGACA 2488 GCCCAGAGAC 2488 GAGTCTCTAA 2592 TGT GCA ACC Cys Ala Thr 2682 Cys GaG TC2	TCCACAG GAC GII 1732 ATG ATG GCC MET MET AII 1838 AGGAGTCAGA 1948 AGGAGTCAGA 2058 CCAAGGAAAC 2168 ATCAAGGAAG 2278 GACCTCAGCCAG 2498 CACGGCAGTAG CAG ATT AT GIN LIE LII 26055 CACCGCCAG	 CCG ACC TC Pro Thr C; AGC CAC T/ Ser His T; 1848 GCACAGAGC 2068 TTGGGGCTGC 2178 GGAGGCTACT 2288 CAGCTGCCCT 2508 AGCTATGCAC 2607 C AGCC TTT G C Thr CC 2709 C ATGCAGCTCC 	CCTA CAG) 's Leu GIn ; 1747 1747 1747 1748 1748 1748 1748 1749 1749 1858 1660TGTTCA 1858 1660TGTTCA 1858 1660TGTTCA 2078 1747 1858 1660TGTTCA 2078 1747 1858 1660TGTCA 2078 1747 1858 1660TGTCA 2078 1747 1858 1660TGTCA 2078 1747 1858 1660TGTCA 2078 1747 1858 1660TGTCA 2078 1747 1858 1747 18588 1858 1858 1858 1858 1858 1858 18	ACC CGC CTG Thr Arg Leu CAC TGC CCT 115 Cys Pro 1868 GGACCCCAGG 2198 AGAGGACCTCA 2308 GGACCCAAAA 2418 GGGACCAAAA 2418 GGGACCAAAA 2528 GGGACCAAAA 2528 GGGACCAAAA 2528 GGGACCAAAA 2528 GGGACCAAAA 2528 GGGACCAAAA 2538 GGGACCAAAA 2538 GGGACCAAAA 2538 GGGACCAAAA 2538 GGGACCAAAA 2538 GGGACCCAAAA 2538 GGGACCCAAAA 2538 GGGACCCAAAA 2538 GGGACCCAAAA 2538 GGGACCCCACG 25388 2538 2538 2538 2538 25388 2538 2538 25388 25388 2538	GAG CTG TAX Glu Leu Tyr 762 CCA ACC CCC Pro Thr Prc 1878 GGGTTICTGT 1988 CCAGCAGAGAG 2098 CCCAGCCAGAC 2208 GGCACTCCTT 2428 CCCAGCCCA 2538 TACCCACTT 263 CTG AAG GG 2753 CTG CAAGAGC 2753 CTG CAAGAGC CTG CAAGAGC 2753 CTG CACCACT 2753 CTG CAAGAGC 2753 CTG CACAGAGC 2753 CTG CACAGC 2753 CTG CACAGC 2753 CTG CACAGC 2753 CTG CACAGC 2753 CTG CACAGAGC 2753 CTG CACAGC 2753 CTG CACAGC 2753 CTG CACAGAGC 2753 CTG CACAGAGC 2753 CTG CACAGAGC 2753 CTG CACAGAGC 2753 CTG CACAGC 2753 CTG CACAGAGC 2753 CTG CACAGAGC 2753 CTG CACAGAGC 2753 CTG CACAGAGC 2753 CTG CACAGAGC 2753 CTG CACAGAGC 2753 CTG CACAGAGC 2753 CTG CACAGAGC 2753 CTG CTG CACAGAGC 2753 CTG CACAGAGC 2753 CTG CACAGAGC 2753 CTG CTG CACAGAGC 2753 CTG CTG CACAGAGC 2753 CTG CTG CACAGAGC 2753 CTG CTG CACAGAGC 2753 CTG CTG CACAGAGC 2753 CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG	C AAG CAG GC Lys GIn GI G GGAGGCCT 1776 G GGAGGCCT 1888 GGCAACAGTT 1998 CTGCTCCTAT 2108 AGGCCACGTA 2218 CCGCGGGAAA 2328 TTGACTGCC 2438 TTGACTGCC 2548 GCCTGGACTC 2548 GCCTGGCGCCAC 2745 CCTCTCACC 2548 CCACCTGCCCC 2438 TCACCTGACC 2548 CCACCTGCCCC 2548 CCACCTCACC 2548 CCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CC CTG CGG y Leu Arg 1788 ACGCCAGGGC 1898 ACTGATGAT 2008 CCTGGTCATT 2118 TCTGCCAGGA 2238 AGAAGGCCAAAG 2558 AAGTCATC ATC 2552 TT GTC ATC 2755 CASTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACGAT 2755 ACTCACCACGAT 2756 ACTCACCACGAT 2756 ACTCACCACGAT 2758 ACTCACCCACCAC
1702 GGC AGC CTC ACC AAG (GIY Ser Leu Thr Lys L 1798 1806 CTCCAGCAGG AATGCTTAA 1918 CTCCAGCAGG AATGCTTAA 1918 TAGCCCTCCA GAGAGGAGGG 2018 2028 2128 21716TCATTGF GCCCCACATG 2348 CTTGCGTCTG GCCCACATG 2458 2446 GGTGTGTCC GGAGTGGCC 2556 CACCTTAAGG 2458 CACCTTAGGCTG GCACTTAAGG 2556 CACTTAGGCTG GCACTTAAGG 2568 CACTTAGCTG GCACGACTG 2458 2568 TATTITTCTT TITTTAAG 2667 CCC TTT GAC TGC TGG CGC	1717 TC AAG GGC (eu Lys Giy i CTTAGGGGGT 1928 GGTCAAGGT 2038 GGTCAAGGT 2148 GGTCAAGGT 2258 CCTTCTGGGT 2368 CAGCCCTTG 2368 CAGCCCTTG 2478 CCTCTGGCTC GAA ACT TCC GI U Thr Ser AG CCA GTC (U Pro Val (AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA 1938 TTTCATCCAA 2048 GAGGGCAAAC 2158 ATTGCAGAGG 2268 GCCCAGAGAC 2378 ACTCTGCTG 2488 TGAGTTCTAA 2592 TGT GCA ACC Cys Ala Thr 2682 CAG GAG TG/ 216 GA U L	TCCACAG GAC GIU 1732 ATG ATG GCC MET MET AI2 1838 AGGGAGCAGA 2058 CCAAGGAAGA 2058 CCAAGGAAGA 2168 ATCAAGGAAG 2278 GACCTCAGCC 2388 CCACTGGCAG 2498 GAGCCAGTAG CAG ATT AT GIN II = II 2695 AGACCGGCAG	 CCG ACC TG	C CTA CAG) 's Leu GIn) 1747 1747 1747 1748 1747 1748 1767 1858 1766 1968 1767 1968 1767 1968 1767 1968 1767 1968 196	ACC CGC CTG Thr Arg Leu CAC TGC CCT 115 Cys Pro- 1868 GGACCCCAGG 1978 ACAGTGCTCC 2088 AGAGGAACTG 2198 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 AAGGACTCCA 2308 AAGGACTCCA 2308 AAGGACTCA 2308 AAGGACTCA 2308 AAGGACCTCA 2308 AAGGACTCA 2308 AAGGACCTCA 2308 AAGGACCTCA 2308 ACCGTGGCTCT AAGGACTCCCA 2308 AAGGACTCA 2308 2	GAG CTG TAA Glu Leu Tyr 762 CCA ACC CCC Pro Thr Pro 1878 GGGTTTCTGT 1988 CCAGCAGGAG 2098 CCCAGGCCCA 2098 CCCAGGCCCA 2208 GGCACTCCTT 2318 GGCAGCGCC 2538 TTACCACTT 263 CCTG AAG G 2532 CTG AAG G 2735 CTCATGAAAC	C AAG CAG GC Lys GIn GI 1776 G GGAAGTGCCT 1888 GGCAACACTT 1998 CTGCTCCTAT 2108 AGTGCCACCT 2108 AGTGCCACCT 2108 AGTGCCACCT 2218 CCTGCGGGAA 2328 TCACTGGCC 2438 TCACTGACC 2548 GCCTGGACTC 37 C TTT CTG C 50 AGGGCTAGA	CCTG CGG y Leu Arg 1788 ACGGCAGGGC 1898 ATGTAATGAT 2118 TCTGGCAGGA CCTGGTCATT 2118 TCTGGCAGGA CAGGCCAAAG 2338 AGAGGCCAA 2488 CAGGCCAGTG 2558 AAGTGTTTTT 2652 TT GTC ATC 2755 AACTCAGGAT
1702 GGC AGC CTC ACC AAG (Gly Ser Leu Thr Lys L) 1798 1800 CTCCAGCAGG AATGTCTTA/ 1908 1918 CTCCAGCAGG AATGTCTTA/ 1908 2018 2018 2018 2018 2018 2018 2128 2117GTCATTA CGGTTAATG/ 2128 2239 2248 CTTCCATG GCCCCACATG 2458 2458 GGTGTGTCCT GGAGTGGCCC 2568 TATITITCTT THTTAAG 2667 CCT TI GAC TGC TGG CGC GGC 2567 CCT TI GAC TGC TGG CG 2765 2775	1717 TC AAG 6GC (eu Lys Giy f 1818 TCTAGGGGT 1928 GGTCAGGGGT 1928 GGTCAGAGA 2038 GGTCAGAGA 2148 GGGTGCTTGA 248 CCTTCTGGGT 2478 CCTGCCCTC GAA ACT TCC GI u Thr Ser AG CCA GTC (1 u Pro Val (2785	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGCTCGACA 1938 GAGGGCAAAC 2048 GAGGGCAAAC 2158 ATTGCACAGG 2268 GCCCAGAGACC 2378 ACTCTGGCTG 2488 TGAGTTCTAA 2592 TGT GCA ACC Cys Ala Thr 2682 CAG GAG TGA Sin Glu 2795	TCCACAG GAC GII 1732 ATG ATG GCC MET MET AI2 1838 TGGGGAGAGA 1948 CCAAGGAATCA 2058 CCAAGGAAAC 2058 CCAAGGAAAC 2168 ATCAAGGAAG 2278 GACCTCAGCC 2388 CCACTGGCAG 2498 GAGGCAGTAT AT GIN IIe II 2695 GACCGCCAGCAB 2805	 CCG ACC TC Pro Thr C; AGC CAC T; Ser H15 T; 1958 CTATGCCTG 1958 TTGGGGCCTG 2068 TGGGGCCCCG 2178 GGAGCTACT 2288 CAGCTGCCCT 2398 AGCTATGCAC 2508 AGAAACATGC 2607 C ACC TTT C e Thr Phe C 2705 ATGAGCTGC 2815 	C CTA CAG) s Leu Gin 1 1747 10 AAG CAG (C r Lys Gin) 1858 GCTGAAGCCC 2078 CCCAAGCCCC 2188 TGGATGTTCA 2298 GCTCTGCCCT 2408 TCCTTGGGGA 2518 TGGTGCTTCC 10 Ser Phe 2715 CCCAAGCCCG 2825	ACC CGC CTG Thr Arg Leu CAC TGC CCT 115 Cys Pro 115 Cys Pro 115 Cys Pro 115 Cys Pro 115 Cys Pro 115 Cys Pro 115 Cys Pro 1978 ACGGCCCAGG 2198 AAGGAACTCC 2088 AAGGAACCTCA 2508 GGGACCAAAA 2508 GGGACCAAAA 2508 GGGACCAACA 2508 GGGACCCACA 2508 GGGACCCACA 2508 GGGACCCACA 2508 CACCGTGGCT 2528 CACCGTGGCTCT 2835	GAG CTG TAA Glu Leu Tyr 762 CCA ACC CCC Pro Thr Prc 1878 CCGGCTTCTGT 1988 CCAGGAGGAG 2098 CCCAGGACGAG 2098 CCCAGGACCC 208 GGCACCCCT 2318 AGGCAGCAGCG 2538 GGCAGCAGCG 2538 GGCAGCAGCG 2538 TTACCCACT 726 CTG AAG GA Leu Lys As CCCATGAAAC 2845	C AAG CAG GC Lys GIn GI G GTGAGTGCT 1776 G GTGAGTGCT 1776 G GGCAACAGTT 1998 CTGCTCCTAT 2108 AGTGCCACCT 218 CCTGCGGGAA 2548 TCACCTGACC 2548 TCACCTGACTC 2548 CCTGGACTC 50 Phe Leu L 50 C AAGAGCTAGA 2855	CCTG CGG y Leu Arg 1788 ACGCCAGGGC 1898 ACGCAGGGC ATGTAATGAT 2008 CCTGGTCATT 2118 TCTGGCAGGA 2338 AGAAGCCAAA 2448 CAGGTCAGTG 2538 AGAAGCCAGTG 2538 AGAAGCCAGTG 2538 AGAAGCCAGTG 2538 AGATGTTTTT 2652 AACTCAGGAT 2865
1702 GGC AGC CTC ACC AAG C GI y Ser Leu Thr Lys L 1798 1806 CTCCAGCAGG AATGTCTTA/ 1908 1916 1908 1917 1908 2018 2018 2018 2128 2138 CTTGCCATTA CGGTTAATGA 238 238 CCTCAGGCTG GCACTTAGT 2498 2568 CATTITCT GACGCTG GCACTAGGCC 2568 TATTITCT TITTAAG 2667 CCC TIT GAC TGC TGG CG 2765 2775 GGTCATCTT GAGGGCCACATA	1717 TC AAG 6GC (eu Lys GIy I 1818 TCTAGGGGT 1928 AGACAGCCA 238 GGTCAGAGAGT 2148 GGTGCTTGA 2258 CCTTCTGGGT 2478 TCCTGGCCTC GAA ACT TCC GIU Thr Ser AG CCA GTC (1 u Pro Val C 1 u Pro Val C GGGGGGGGGCCC	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGTCGACA 1938 GTTCATCCCA 2048 GAGGCCAACA 2158 ATTCCAGAGG 2268 GCCCAGACAC 2488 ACTCTGGCTG 2488 ACTCTGGCTG 2488 ACTCTGGCTG 2483 ACTCTGGCTG 2483 GAGTCTCTAA 2592 TGT GCA ACC Cys Ala Thr 2682 CAG GCA Sin Glu 2795 ACAGCCATGG	TCCACAG GAC GII 1732 ATG ATG GCC MET MET AIZ 1838 AGGAGTCAGA 2058 CCAAGGAAGA 2168 ATCAAGGAAG 2278 GACCTCACCC 2388 CCACTGGCAG 2498 CACCTGCCAG CACGCAGTAG CAGCATT AT GIN LIE JI 2655 GACCGCCCCC 2805 TGGGAATGGC	 CCG ACC TC Pro Thr C; AGC CAC TI Ser His T; 1848 GCACAGAGC 1958 GCACAGAGC 2068 TTGGGGCTGC 2178 GGAGGCTACT 2288 CAGCTGCCTT 2398 AGCTATGCAC 2508 AGCTATGCAC 2509 AGAAACATGC 2607 C ACC TTT 6 2705 ATGAGCTGCTGCC 2815 CTGGACTGCC 	CCTA CAG) 's Leu GIn ; 1747 1747 1747 1748 1747 1748 1748 1747 1747 1747 1747 1858 1858 1858 1858 1858 1968	ACC CGC CTG Thr Arg Leu CAC TGC CCT 115 Cys Pro 115 Cys Pro 115 Cys Pro 1868 GGACCCCAGG 2088 AGAGGAGTG 2198 AAGGACCTCA 2308 GGGACCAAAA 2418 ACACGTGGGT 2528 GGGACCAAA 2528 GGGACCAAA 2535 CTCCCCCGGA 2725 GAGCTGCTG	GAG CTG TAX GIU LEU Tyr 762 CCA ACC CCC Pro Thr Prc 1878 CGGTTICTGT 1988 CCAGCAGAGA 2098 CCCAGGCCA 2098 CCCAGGCCA 2238 AGGCAGCGC 2538 AGGCAGCGC 2538 CCCAGCACC 2538 TACCCACTT 2428 CCCAGCACCA 2538 CCCAGACCC 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCCA 2538 CCCAGCCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCA 2538 CCCAGCCCCCCCAC 2538 CCCAGCCCCCCCCCCAC 2538 CCCAGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	C AAG CAG GC Lys GIn GI G GGAGGCCC 1776 G GGAGGCCC 1888 GCCAACAGTT 1998 CTGCTCCTAT 2108 AGGCCACCT 2218 CCGCGGGAA CCTGCGGACTC 2438 TTGACTGACC 2548 GCCTGGACTC 37 CACTGACC 2745 C AAGAGCTAGA 2855 GCAGAAGAAT	CC CTG CGG y Leu Arg 1788 ACGCCAGGGC 1898 ACGACAGGGC 1898 CCTGGTCATT 2118 TCTGCCAGGA 2238 AGAAGGCCAAA 2338 AGAAGGCCAA 2538 AGGTCACTG 2558 AAGTGTTTTT TCTG CATC 2558 AGGTCATCATG 2558 AGGTCATCATG 2558 AGGTCATCATG 2558 AGGTCATCATCATG 2558 AGGTCATTTT 2655 GGGAATATTT
GGC AGC CTC ACC AAG C GIY Ser Leu Thr Lys L 1798 1800 CTCCAGCAGG AATGCTTAA 1908 1918 TAGCCCTCCA GAGAGGAGGG 2018 2022 ATGCTATTA CGGTTAATG 2128 2128 CTTCCTCTG GCCCCACATG 2238 238 CCTCAGGCT GACCCACTTAAG 248 2466 GGTGTGTCC GGAGTGGGCC 2568 TATTTTCTT TTTTTAAG CCC TTT GAC TGC TGG C 2765 277 GGTCATCTTG GAGGGACCAA 2875 2875	1717 TIC AAG GGC 1 1818 1646 1646 1646 1646 1717 1818 192	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGCTCGACA 1938 TTTCATCCCA 2048 GAGGGCAAAC 2158 ATTGCAGAGG 2268 GCCCAGAGAC 2378 GCCCAGAGAC 2378 TGT GCA ACC Cys Ala Thr 2682 CAG GAG TGA 210 U 2795 CAG GAG TGA 210 U 2795 211 U 2795 211 U 2795 211 U 2795 201 U 2795 211 U 2795 201 U 2795 201 U 2795 201 U 2795 201 U 2795 201 U 2795 201 U 2795 201 U 2795 201 U 2795 270	TCCACAG GAC GIU 1732 ATG ATG GCC MET MET AI2 1838 AGGGGAGAGA 2058 CCAAGGAAC 2168 ATCAAGGAAG 2278 GACCTCAGCC 2388 GAGCCAGTAG CCACTGCAG CACTGCAG CACTGCAG CACGCAGCAG CAG ATT AT GIN LIE LI 2695 GAGCCGCAG 2805 TGGGAGTGGC 2915	 CCG ACC TG Pro Thr C; AGC CAC T/ Ser His T; 1848 GCACAGAGC 2068 TTGGGGCTG 2178 GGAGGCTACT 2398 CAGCTGCCCT 2398 AGCATATGAC 2607 C ACC TTT G; ATGAGGCTGC 2815 CTGGACCTGC 2925 	C CTA CAG : 's Leu GIn : 1747 1747 1747 1748 1747 1748 1748 1749 1749 1749 1749 1858 1766 1968 1767 1968 196	CACC CGC CTG Thr Arg Leu CAC TGC CCT 115 Cys Pro- 116 Cys Pro- 1868 GGACCCCAGG 1978 ACAGTGCTCC 2088 AGAGGAAGTG 2198 AAGGACCTCA 2308 AAGGACCTCA 2528 TTCCCCAGG CZ22 AAA GGA AAG 418 ACAGTGGGT 2528 TTCCCCCAGG 2418 CCCCCCCCAG 2525 CACGCCCTGA 2725 CTGACCCCTGA 2945 TTCCCCCGAG 2945 TTCCCCCCGA 2945 TTCCCCCCGA 2945 TTCCCCCCGA 2945 TTCCCCCCGA 2945 TTCCCCCCGA 2945 TTCCCCCCGA 2945 TTCCCCCCGA 2945 TTCCCCCCGA 2945 TTCCCCCCGA 2945 TTCCCCCCGA 2945 TTCCCCCCGA 2945 TTCCCCCCGA 2945 TTCCCCCCGA 2945 TTCCCCCCGA 2945 TTCCCCCCCGA 2945 TTCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCCGA 295 TTCCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCGA 295 TTCCCCCCCCGA 295 TTCCCCCCCCGA 295 TTCCCCCCCCGA 295 TTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	GAG CTG TAN GI L Leu Tyr 762 CCA ACC CCC Pro Thr Prc 1878 GGGTTTCTGT 1988 CCACCAGGAG 2098 CCCAGGCCG 2208 GGCACTCCTT 2428 GGCAGCCGC 2538 TTACCAGCT 2428 CCC GAAG GC 2538 TACCACCAC 2538 CCCAGACCG 2538 CCCAGACCG 2538 CCCAGACCG 2538 CCCAGACCG 2538 CCCAGACCG 2538 CCCAGACCG 2538 CCCAGACCG 2538 CCCAGACCG 2538 CCCAGACCG 2538 CCCAGACCG 2538 CCCAGACCG 2538 CCCAGACCG 2538 CCCAGCCCA 2538 CCCAGCCCC 2538 CCCACCCC 2538 CCCACCCC 2538 CCCACCCC 2538 CCCACCCC 2538 CCCCCCCCC 2538 CCCCCCCCC 2538 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	C AAG CAG GC Lys GIn GI GGGAAGTGCCT 1776 GGGAAGTGCCT 1888 GGCAACACTT 1998 CTGCTCCTAT 2108 AGTGCCACCT 2218 CCTGCGGGAA 2228 GCCTGCGGACTC 2438 GCCTGGACTC 2438 GCCTGGACTC 2438 GCCTGGACTC 2438 GCCTGGACTC 2438 GCCTGGACTC 2438 GCCTGGACTC 2438 GCCTGGACTC 2438 GCCTGGACTC 2438 GCCTGCGACTC 2438 GCCTGCGACTC 2438 GCCTGCGACTC 2438 GCCTGCGACTC 2438 GCCTGCGACTC 2438 GCCTGCGACTC 2438 GCCTGCGACTC 2438 GCCTGCGACTC 2438 GCCTGCGACTC 2438 GCCTGCCCCC 2438 GCCTGCCCCC 2438 GCCTGCCCCC 2438 GCCTGCCCCC 2438 GCCTGCGACTC 2438 GCCTGCCCCCC 2438 GCCTGCCCCCC 2438 GCCTGCCCCC 2438 GCCTGCCCCC 2438 GCCTGCCCCCC 2438 GCCTGCCCCCC 2438 GCCTGCCCCCCCCCCCCCCCCCCC 2438 GCCTGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CC CTG CGG y Leu Arg 1788 ACGGCAGGGC 1898 ATGTAATGAT 2008 CCTGGTCATT 2118 TCTGGCAGGA 2328 GGGAGCAAG 2528 AGGTCATG 2552 TT GTC ATC aCG52 AGCTCAGGAT 2652 TT GTC ATC 2755 AACTCAGGAT 2865 GGGAATATTT 2865 GGGAATATTT 2975
1702 GGC AGC CTC ACC AAG (GIY Ser Leu Thr Lys L) 1798 1800 CTCCAGCAGG AATGCTTAA 1908 1918 CTCCAGCAGG AATGCTTAA 1908 2018 2018 2018 2018 2018 2018 2128 2117CCTCTG GCCCCACAT 2238 2248 CTTCCTCTG GCCCCACTT 2348 2358 CCTCAGGCTG GGACTTAAGG 2568 TATITTCTI THTTTAAAG 2667 CCT TI GAC TGC TGE GC 2568 CTTI T GAC TGC TGE G Pro Phe ASp Cys Trp 0 2765 275 GGTATCTTG GAGGGACCAA 2855 275 GTATCTTGCAAGCAAATCACTAAT 285	1717 1717 1717 1717 1717 1818 1717 1928 1717 1928 1717 1928 1717 1928 1717 1928 1717 1928 1939 1938 19 8 19 8 19 19 19 19 19 19 19 19 19 19 19 19 19	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGCCGACA 1938 GAGGGCCAACA 2048 GAGGCCAACA 2048 GAGGCCAACA 2588 ATTGCAGAGG 2488 GCCCAGAGACC 2378 ACTCTGCTG 2488 TGAGTTCTAA 2592 TGT GCA ACC Cys Ala Thr 2682 CAG GAG TGA SIN Glu 2795 ACACCATGG 2905 ATTTATATTT	TCCACAG GAC GII 1732 ATG ATG GCC MET MET AI2 1838 TGGGGAGAGA 1948 CCAGGGCCAGA 2058 CCAGGGAAAC 2058 CCAGGGAAAC 2168 ATCAAGGAAG 2278 GACCTGCCAG 2388 CCACTGGCAG 2498 GAGGCAGTAGC 260 CAG ATT AT 160 CAG ATT AT 160 CAG ATGGCCAG 2915 TTAAAATATT	 CCG ACC TC Pro Thr C; AGC CAC TA Ser H1s T; 1958 CTATGCCTG 2068 CTGGGGCCTG 2178 GGAGGCTACT 2288 CAGCTAGCCC 2508 AGAAACATGC 2607 CACCTTC CACCTTC 2815 CTGGACCTGC 2925 TATLTATTA 2038 	C CTA CAG) 1747 1747 16 AAG CAG (C r Lys GIn) 1858 GCTGAAGCCC 2078 CCCAAGCCCC 2188 TGGATGTTCA 2298 GCTCTGCCCT 2408 TCCTTGGGGA 2518 TGGTGCTTCC 2408 TCCTTGGGGACA 2715 CCCAGCCGG 2825 CCTGGGCACA 2935 TTTATTTAAG	ACC CGC CTG Thr Arg Leu ICAC TGC CCT 115 Cys Pro 136 Cys Pro 136 Cys Pro 137 Cys Pro 137 Cys Pro 137 Cys Pro 137 Cys Pro 138 Cys Cys Pro 138 Cys	GAG CTG TAA Glu Leu Tyr 762 CCA ACC CCC Pro Thr Pro 1878 CCAGCAGCA 2098 CCAGGAGGAG 2098 CCCAGGACCC 208 GGCACCAGCAG 2318 AGCCAGCAGCA 2318 AGCCAGCAGCA 238 TACCACCTT 2428 CTG AAG GA Leu Lys A CTCATGAAAC 2845 TACCAGCAT 2955 ATATTTATTC	C AAG CAG GC Lys GIn GI G GTGAGTGCT 1776 G GTGAGTGCT 1998 AGTGCCACATT 1998 AGTGCCACTT 2108 AGTGCCACCT 218 CCTGCGGCAA 2328 TTGACTGCCC 2438 TCACCTGACC 2548 GCCTGGACTC 37 C ATCTGCC 2745 C AAGAGCTAGA 2855 GCAAAGATTTTT	CCTG CGG y Leu Arg) 1788 ACGGCAGGGC 1898 ATGTAATGAT 2008 CCTGGTCATT 2008 CCTGGTCATT 2118 TCTGGCAGGA 2238 AGAAGCCAAG 2558 AGAAGCCAAG 2558 AGAAGCCAGTG 2558 AGAGTCAGTG 2558 AGATGTTTTT 2652 GGGAATATTT 2975 ACCCGTAATAA
1702 GGC AGC CTC ACC AAG (GIY Ser Leu Thr Lys L) 1798 1800 CTCCAGCAGG AATGTCTTA/ 1908 1918 CTCCAGCAGG AATGTCTTA/ 1908 1918 CTCCAGCAGG AATGTCTTA/ 1908 2018 2018 2018 2018 2018 2018 2018 2018 2018 2018 2018 2018 2018 2018 2024 CTTGCCATTA CGGTTAATGA 2138 CTTCCCG GCCCCACATG 2238 2248 2238 2248 2458 2458 2458 2458 2667 CCC TTT GAC TGC TGG GC 275 GGTCATCTT GAGGGCC 275 GGTCATCTTG GAGGACCAR 2875 2955 2955 2955	1717 1717 1717 1717 1818 1717 1718 1717	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 GGGGCCGACA 1938 GAGGGCAAAC 2158 ATTCCACAGGG 2268 GCCCAGAGACC 2488 GCCCAGAGAC 2488 ACTCTGGCTG 2488 ACTCTGGCTG 2488 CCCAGCACC Cys Ala Thr 2692 Cys Ala Thr 2795 ACAGCCATGG 2905 ATTTATATTT 3015	TCCACAG GAC GII 1732 ATG ATG GCC MET MET AII 1838 CGGGAGAGAA 1948 CCAGGAAGCA 2058 CCAGGAAGA 2168 ATCAAGGAAG 2168 ATCAAGGAAG 2278 GACCTCACCC 2388 CCACTGCCAG 2498 CCACTGCCAG 2498 CACCCACAGCAG 2498 CACCCACAGCAG 2498 CACCCACAGCAG 2495 TGGGAAGGCAGCAG 2915 TTAAAATATT 3025 TTGTITTATA	 CCG ACC TC Pro Thr C; AGC CAC TA Ser HIS T; 1958 GCACACAGC 2068 TTGGGGCCTG 2068 TTGGGGCTACT 2288 CAGCTAGCC 2398 AGCTATGCAC 2508 AGCATATGCAC 2509 AGAAACATGC 2607 C ACC TTT C 2815 CTGGGACCTGC 2925 TATTATTATA 3035 	C CTA CAG) 's Leu Gin 1747 1747 1747 1747 1748 1748 1747 1748 1747 1747 1858 GCTGAGCC 2188 GCTGAGCCC 22188 GCTCTGCCCT 2408 TCCTTGGGA 2518 TGGTGTTCC 2408 TCCTGGGA 2515 CCCAGCCGG 2825 CCCGGGCACA 2935 TTATTTACAG AATGCCAT	CACC GCC CTG Thr Arg Leu CAC TGC CCT 115 Cys Pro 115 Cys Pro 115 Cys Pro 1978 ACAGTGCTCC 2088 AGAGGAGTG 2198 AAGAGAACTCA 2308 GGGACCAAA 2308 GGGACCAAA 2308 GGGACCAAA 2528 GGGACCAAA 2528 GGGACCAAA 2528 GGGACCAAA 2528 CACCGTGGT 2525 GAGCTGCTC 2835 CTGACCCTGA 2945 TTCATATTCC	GAG CTG TAX GLU LEU TYI 762 CCA ACC CCC Pro Thr Prc 1878 CGGTTTCTGT 1988 CCAGCAGGAG 2098 CCCAGGCCCA 2208 GGCACCAGGCCA 2218 AGGCAGGAGCG 2538 GGCAGCAGCG 2538 GGCAGCAGCG 2538 TACCCACTT 1 LEU LYS AS 275 CTCTATGAAAG 2845 TACCAGGCATG 2845 TACCAGGCATG	C AAG CAG GC Lys GIn GI G GGAGGCCD 177E G GGAGGCCD 1888 CGCACAGTT 2108 AGTGCCCCTAT 2108 AGTGCCCCTAT 2218 CCTCCGGAA 2328 TGACTGCCC 2438 TCACCTGACC 2548 TCACCTGACC 2548 CCCGGGACTAC 2545 CCAGAGCTAGA 2855 GCAGAAGAAT 2965 AAGATGTTT	CCTG CGG y Leu Arg 1788 ACGCCAGGGC 1898 ACGACAGGGC 1898 CCTGGTCATT 2008 CCTGGTCATT 2018 CCTGGTCATC 2018 CCTGGTCATC 2018 CCTGGTCATCATCATT 2017 CCTCATCATT 2017 CCTCATCATTT 2017 CCTCATCATTT 2017 CCTGTCATCATTTT 2017 CCTGTCATCATTTT 2017 CCTGTCATCATTTT 2017 CCTGTCATCATTTT 2017 CCTGTCATCATTTT 2017 CCTGTCATCATTTTT 2017 CCTGTCATCATTTTTTTTTTTTTTTTTTTTTTTTTTTTT
1702 GGC AGC CTC ACC AAG (GIY Ser Leu Thr Lys L) 1798 1806 CTCCAGCAGG AATGCTTAA 1918 TAGCCCTCCA GAGAGGAGG 2018 2028 2128 2171CCATCTG GCCCCACATG 2238 2238 CTTCCTCTG GCCCCACATG 2458 2458 CCTCAGGCTG GCACTTAAGG 2458 CCTCAGGCTG GCACTACACT 256 CTTCGCTCT GCAGCGACGAC 2568 CCC TTT GAC TGCC TGG 2755 CCC TTT GCAGGACCAA 2875 CCC TTAGC GCACAACAGTA 2875 CCC TTACACAGAACAGTA 285	1717 TIC AAG GGC 1 1818 1CTAGGGGT 1928 AGACAGCCG 2038 GGTCAGAGGT 2148 GGTCATGA 2258 CCTTCTGGGT 2478 CCTGCGCTC GAA ACT TCC GGA ACT TCC GGA ACT TCC 2785 GGCGTGGCC 2895 GGCGTGGGCC 2895 GGCGTGGGCC 2895 GGCGTGGGCC 2895 GCTATTTTATAT 3005 CTACTGTCC 4	AACGACATTT CCC TTG ACC Pro Leu Thr 1828 TTTCATCCGA 1938 TTTCATCCCA 2048 GAGGCCAAAC 2158 ATTGCACAGAG 2268 GCCCAGAGAC 2488 GACTCTGAC 2488 TGACTCTAA 2592 TGT GCA ACC Cys Ala Thr 2682 CAG GCA 219 ACACCCATGG 2905 ATTTATATTT 3015 AGTGTTCTAG	TCCACAG GAC GII 1732 ATG ATG GCC MET MET AI2 1838 AGGAGTCAGA 2058 CCAAGGAAGA 2168 ATCAAGGAAG 2278 GACCTCAGCC 2388 GAGCCAGTAG CACTGCCAG 2498 GAGCCAGTAG CACTGCCAG 2498 GAGCCAGTAG CACGCCAGC 2805 TGGCAGTGGC 2915 TGAAATATT 3025 TTTGTTTTA	 CCG ACC TG Pro Thr C; AGC CAC T/ Ser His T; 1848 GCACAGAGC 1958 GCCACAGAGC 2068 TTGGGGCTG 2178 GGAGGCTACT 2288 CAGCTGCCCT 2398 CAGCTGCCCT 2508 CAGCAGAGC 2607 C ACC TIT C 2815 CTGGACCTGCCCT 2925 TATGAGCTGCCCT 2815 CTGGACCTGCCC 2925 TATTATTA 3035 ACCATGAGCA 	CCTA CAG) 's Leu GIn) 1747 1747 1747 1748 1747 1748 1748 1748 1747 1748 1858 1858 1867 1858 1867 1878 1867 18788 1878 1878 1878 1878 1878 1878 187	ACC CGC CTG Thr Arg Leu CAC TGC CCT 18 Cys Pro 1868 GGACCCCAGG 2198 AGAGGACCCA 2198 AAGGACCTCA 2308 GGACCAAAA 2418 GGGACCAAAA 2418 GGGACCAAAA 2418 GGGACCAAAA 2528 GGGACCAAAA 2418 GGGACCAAAA 2528 GGGCCCCAGG 2528 ACAGGTGGGT 2528 GGGCCCCCAG 2528 CTCCCCCCCAG 2528 CTCCCCCCCAG 2528 CTCCCCCCCAG 2528 CTCCCCCCCAG 2528 CTCCCCCCCAG 2528 CTCCCCCCCCAG 2528 CTCCCCCCCCAG 2528 CTCCCCCCCCAG 2528 CTCCCCCCCCAG 2528 CTCCCCCCCCAG 2528 CTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	GAG CTG TAN GIU LEU Tyr 762 CCA ACC CCC Pro Thr Prc 1878 GGGTTICTGT 1988 CCAGCAGGAG 2098 CCCAGGCCA 2208 GGCACTCCTT 2428 CCCAGGCCA 2538 TACCACTT 2428 CCCAGCAGCG 2538 TACCACTT 2428 CCCAGCACG 2538 CCCAGCACG 2538 CCCAGCACG 2538 CCCAGCACG 2535 CTCATGAAACC 2845 TACCAGCATG 2955 ATATTTATTC	C AAG CAG GC Lys GIn GI GGCAGGGCCT 177E GGCAACAGTT 1998 CTGCTCCTAT 2108 CTGCCCCTAT 2218 CCGCGGGAA 2328 TTGACTGCCC 2438 TTGACTGCC 2438 CCACCGGACTAC 2548 GCCGGGACTAGAC C 2745 C AAGAGCTAGAC 2855 GCAGAAGAAT 2965 AAGATGTTTT	CCTG CGG y Leu Arg 1788 ACGCCAGGGC 1898 ATGTAATGAT 2008 CCTGGTCATT 2118 TCTGCCAGGA 2238 AGAGGCCAAAG 2558 AGGGTCATT 2552 AGGTCATG 2552 TT GTC ATC 2755 AACTCAGGAT 2865 GGGAATATTT 2975 ACCGTAATAA

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 $\sim 70\%$ 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 \sim 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 AccI 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 $[\gamma^{-32}P]ATP$ (~1 × 10⁷ c.p.m. ³²P/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 μ g of the labeled DNA probe and incubated at 75°C for 10 min in a volume of 15 μ l (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 μ l. 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 AccI 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.

TGATGAATAAGTTTGIGGIIIGCTATGGAGGTTCCATCTCAGATAAAGCIGCITCIGATGCCTGCCCCCCCCCC	
TCACTGGCCCALGIAIAAGGCCAGATAAGGGCCAGGAGATTCCAACAGTAGTAGTAGTAGTAGTAGCCCCCCCC	CSF
ΑΛΑΤ GCCA CA ΑΛΑCCΤΤΑGTTATTAATACAAACTATCAT <mark>CCCT GCCTATCTGTCACCATGTCATACTTAAAAAACTTGTGAAAATACGTAATCCTCAGGAGACTTCAATTAGG<mark>TATAA</mark>ATACCAGCAGCCAG HUYIFN</mark>	I

200 200 200 200 200 200 200 200 200 A G T C A G A G C C A G G C A G G C A G C C A C C T C C C C A A C C T G T T C Mull. 3 T A G C C A G A G G C T G G C A G C T G C C A G C A G G T G G G A A MUGM CSF

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 GCrich 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 μ g/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 $(1 \times 10^8 \text{ c.p.m. } {}^{32}\text{P}/\mu\text{g})$. 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: $6 \times \text{SSPE}$ (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 $\mu\text{g}/\text{ml}$ 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). 10⁶ COS-7 monkey cells were transfected with 25 μ g 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 μ M 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 μ M 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 ³²P-labeled *Pstl/AhaIII* fragment of human GM-CSF cDNA (Lee *et al.*, 1985). The filter was washed with $2 \times SSC$ (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 \times 10⁵ plaques of this library were screened by plaque hybridization using a *PstI/AhaIII* 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 *HaeIII* and *AluI*. 1 × 10⁶ plaques were screened by ³²P-labeled human GM-CSF probe. DNA from positive λ phages were purified and subjected to restriction mapping. A 3.1-kb *HindIII/Eco*RI restriction fragment of one of the positive clones, λ HGM11-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, Č., 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