The EMBO Journal vol.4 no.3 pp.645–653, 1985

Structure and expression of the mRNA for murine granulocyte-macrophage colony stimulating factor

Nicholas M.Gough, Donald Metcalf[1], Jill Gough, Dianne Grail and Ashley R.Dunn

The Ludwig Institute for Cancer Research, Melbourne Tumour Biology Branch, and [1]The Walter and Eliza Hall Institute for Medical Research, Post Office Royal Melbourne Hospital, Parkville, Victoria 3050, Australia

Communicated by W.Keller

A cDNA clone containing a virtually complete copy of the mRNA for the haemopoietic growth regulator, granulocyte-macrophage colony stimulating factor (GM-CSF), has been isolated from a murine T lymphocyte cDNA library. When a eukaryotic expression vector with this cDNA coupled to the SV40 late promoter was introduced into simian COS cells, significant quantities of GM-CSF were secreted. Since all of the biological activities previously ascribed to highly purified GM-CSF were exhibited in the COS cell-derived GM-CSF, all of these activities are intrinsic to the product of a single gene. There are two potential translational initiation codons in the GM-CSF mRNA; the first is buried in the stem and the second located in the loop of a very stable hairpin structure. Expression studies using deletion derivatives of the cDNA indicated that the second AUG is able to initiate the translation and secretion of GM-CSF. The amino acid sequence of the leader peptide is rather atypical for a secreted protein and we speculate that molecules which initiate at the first AUG might exist as integral membrane proteins whereas those initiating at the second are secreted.

Key words: cloned growth factor/eukaryotic expression vector/ nucleotide sequence/RNA secondary structure/transmembrane protein

Introduction

Haemopoietic cell proliferation and differentiation is controlled by a group of glycoprotein regulators, the colony stimulating factors (CSFs) (Metcalf, 1984). The CSFs are active at low concentrations ($10^{-11}-10^{-13}$ M) and are usually assayed by their ability to stimulate the proliferation of committed progenitor cells to form colonies of differentiating progeny in semisolid cultures. For murine granulocyte-macrophage populations, four major CSFs have been identified (GM-CSF, Multi-CSF, M-CSF and G-CSF) that interact to control the formation of granulocytes and macrophages (Metcalf, 1984). Each has been purified to homogeneity and its spectrum of biological activities determined. GM-CSF (mol. wt. 23 000) (Burgess *et al.*, 1977) stimulates the formation of granulocyte and macrophage colonies (Burgess and Metcalf, 1977) and, in cultures of foetal liver progenitor cells, is also an effective stimulus for eosinophil colony formation (Metcalf and Nicola, 1983). Whilst GM-CSF has the ability to support the survival and initial proliferation of some erythroid, multipotential and megakaryocyte precursors, GM-CSF cannot support the continued proliferation and differentiation of these cells to form colonies of mature progeny (Burgess and Metcalf, 1980; Metcalf *et al.*, 1980).

Although it has been possible to explore many aspects of the *in vitro* biology and biochemistry of purified CSFs, studies of the structure and *in vivo* effects have been hampered by the limited amounts of material available. This problem may be overcome by molecular cloning of the corresponding gene sequences to provide large amounts of clonally-derived material, an approach that also permits a wide variety of structure-function studies on the CSFs. We have recently cloned partial DNA copies of the GM-CSF mRNA from a mouse lung cDNA library (Gough *et al.*, 1984). We now describe the isolation and nucleotide sequence of a full length GM-CSF cDNA clone, its expression in eukaryotic cells and the biological activities of the cloned gene product. We also speculate on the significance of the structure of the leader peptide(s) and the mRNA that encodes it.

Results

Isolation of a full-length GM-CSF cDNA clone

We recently isolated two cDNA clones (pGM37 and pGM38 in Figure 1A) complementary to the GM-CSF mRNA from the lungs of endotoxin-treated mice (Gough *et al.*, 1984). Their nucleotide sequences indicated that neither contains the translational initiation codon: pGM38 starts at amino acid residue 6 of the mature protein, and pGM37 at a position 20 nucleotides 5′

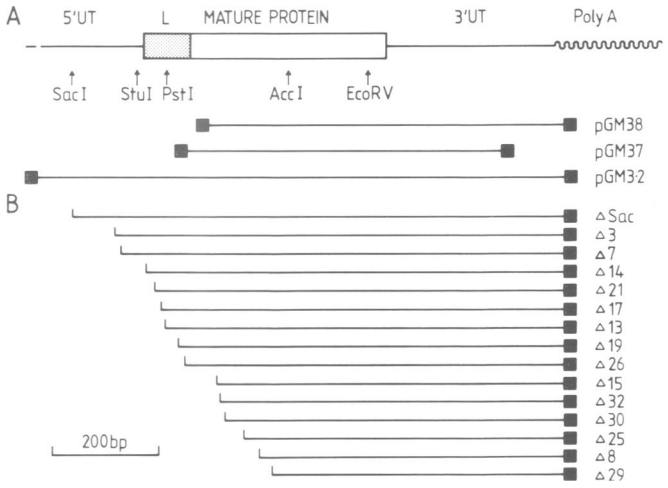

Fig. 1. (**A**) Map of the murine GM-CSF mRNA and of cDNA clones pGM38, pGM37 and pGM3.2. The mRNA is 1200 nucleotides in length (Gough *et al.*, 1984). The region of the mRNA encoding the mature protein is shown as an open block and the NH_2-terminal hydrophobic leader peptide (L) as a stippled block. The untranslated regions are designated by UT. The locations of certain restriction endonuclease cleavage sites are indicated below the mRNA. The regions contained within cDNA clones pGM37, pGM38 (Gough *et al.*, 1984) and pGM3.2 (this work) are indicated by lines; the blocks at the ends of these lines indicate the oligo(dG-dC) tails. (**B**) Regions of the GM-CSF mRNA contained in various 5′ deletion mutants derived from pGM3.2. The locations of the 5′ end-points of clones pGM5′△Sac, and pGM5′△3, 7, 14, 21, 17, 13 and 19 were determined by nucleotide sequence analysis and of the other clones by restriction mapping.

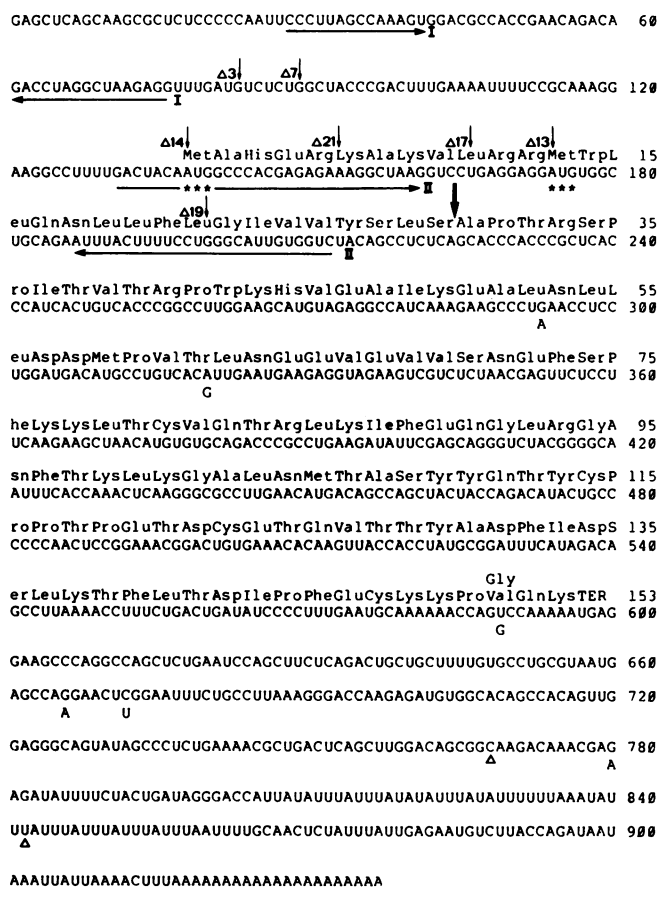

Fig. 2. Nucleotide sequence of the murine GM-CSF mRNA. The sequence of the mRNA is listed 5′ to 3′ with the predicted amino acid sequence of GM-CSF given above; numbers at the ends of the lines indicate the position of the final residue (amino acid or nucleotide) on that line. The amino acid sequence numbering starts with the first methionine of the major open reading frame. Two potential initiation codons are indicated with asterisks. The location of the 5′ end-points of various 5′ deletion mutants (Figure 1B) are indicated with arrows. Regions involved in potential secondary structure (Figure 3) are underlined. The bold arrow indicates the site at which the leader peptide is cleaved from the mature protein (A.W.Burgess, E.C.Nice, N.A.Nicola, R.Cutler and R.J.Simpson, personal communications). The sequence presented was derived from clone pGM3.2 (Figure 1A) which is of BALB/c origin and differs at 12 positions from the sequence of clones pGM37 and pGM38 (Gough *et al.*, 1984) which are of C57BL/6 origin. Four of these differences reflect errors in our previous sequence: position 245 was previously assigned as an A, positions 247 and 251 as T's and position 897 as AA rather than A; a T residue at 247 previously predicted an Ile residue; this residue is now predicted to be Thr, in accord with the protein sequence of Sparrow *et al.*, (1984). The other eight differences probably reflect allelic variation between the two mouse strains: the C57BL/6 alternatives are given below the line; two nucleotides (positions 768 and 842) not present in the C57BL/6 sequence are marked by △.

to the first amino acid residue detected in asialo GM-CSF. As a prelude to the expression of GM-CSF in cell culture under the direction of a cloned cDNA sequence, we wished to isolate a cDNA clone containing a complete copy of the GM-CSF mRNA. Since the abundance of GM-CSF sequences in mouse lung mRNA is extremely low (we originally isolated two clones from a library of 100 000), we used a cloned T lymphocyte line, LB3 (Kelso *et al.*, 1984; Kelso and Metcalf, 1985), in which the synthesis of high levels of GM-CSF mRNA is inducible by concanavalin A (Gough *et al.*, 1984). We constructed a library of cDNA clones complementary to concanavalin A-stimulated LB3 mRNA and searched this library for GM-CSF clones by colony hybridization using as a probe a fragment of DNA derived from pGM38. Of 24 GM-CSF clones isolated and examined, one (pGM3.2) appeared to contain a substantial, if not complete, copy of the GM-CSF mRNA. The cDNA insert in pGM3.2 is ~1.05 kbp in length which includes ~50 bp for the oligo(dG.dC) tails at either end of the cDNA molecule. This compares with an estimated length for the GM-CSF mRNA of 1.2 kb (Gough *et al.*, 1984), which includes a poly(A) tail of probably 50−200 residues (Jeffrey and Brawerman, 1974; Adams, 1977). Thus, the cDNA insert in pGM3.2 is at most 150 bp short of being a complete copy of the GM-CSF mRNA, depending upon the length of the poly(A) tail. As determined by mapping the location of various restriction endonuclease cleavage sites and by nucleotide sequence analysis (below), the cDNA insert in pGM3.2 extends from the poly(A) tail of the mRNA to a position some 200 bp 5′ to the start of the coding region (see Figure 1A).

Nucleotide sequence of the GM-CSF mRNA

The nucleotide sequence of the cDNA of pGM3.2 was determined 3′ to the *Sac*I restriction endonuclease site that is located close to the 5′ end of the cDNA (see Figure 1A). The deduced GM-CSF mRNA sequence and the amino acid sequence predicted to be encoded by this mRNA are shown in Figure 2. The sequence of pGM3.2 differs from the previous sequence at 12 positions. Four of these were in error in our previous sequence (for details, see legend to Figure 2). However, the other eight residues appear to represent genuine differences between the two sequences. Since pGM3.2 is derived from BALB/c mice, whereas the two previous clones were from C57BL/6, it is likely that these differences represent allelic variation. Of these eight allelic differences, three (all base substitutions) occur in the coding region (positions 293, 320 and 589); two of these are silent substitutions and one results in an amino acid interchange (Val:Gly at position 151 in the amino acid sequence). The other five differences (three base substitutions and two insertions in the BALB/c sequence) occur within the 3′ untranslated region of the mRNA.

The primary nucleotide sequence of the GM-CSF cDNA indicates that two 'hairpin loop' structures (illustrated in Figure 3) might exist within the secondary structure of the mRNA in the 5′ untranslated region and at the start of the coding region. Whilst it is difficult to predict secondary structure models from primary nucleotide sequence data, these hairpin loops appear to be highly stable structures: using the rules of Tinoco *et al.* (1971, 1973) and Salser (1977) the free energy values for structures I and II were estimated as −17.35 and −25 kcal respectively. Although structures of comparable stability have been proposed for some eukaryotic mRNAs [e.g. the SV40 viral VP1 mRNA (Ghosh *et al.*, 1978), the rat β casein mRNA (Blackburn *et al.*, 1982) and the HMG CoA reductase mRNA (Reynolds *et al.*, 1984)] the structures in the GM-CSF mRNA appear to be much more stable than most other proposed mRNA secondary structures, e.g. the rabbit β-globin, chicken ovalbumin and conalbumin and the hepatitis B viral core antigen mRNAs, which have △G values of −4 to −12.51 kcal (Efstratiadis *et al.*, 1977; McReynolds *et al.*, 1978; Cochet *et al.*, 1979; Pasek *et al.*, 1979).

Predicted amino acid sequence of GM-CSF

There are two potential initiation codons for the major open reading frame of the GM-CSF mRNA. The first (5′-proximal) of these is at position 138 and the second at 174 in the nucleotide sequence (indicated with asterisks in Figure 2). A sequence of predominantly hydrophobic amino acids typical of a secreted pro-

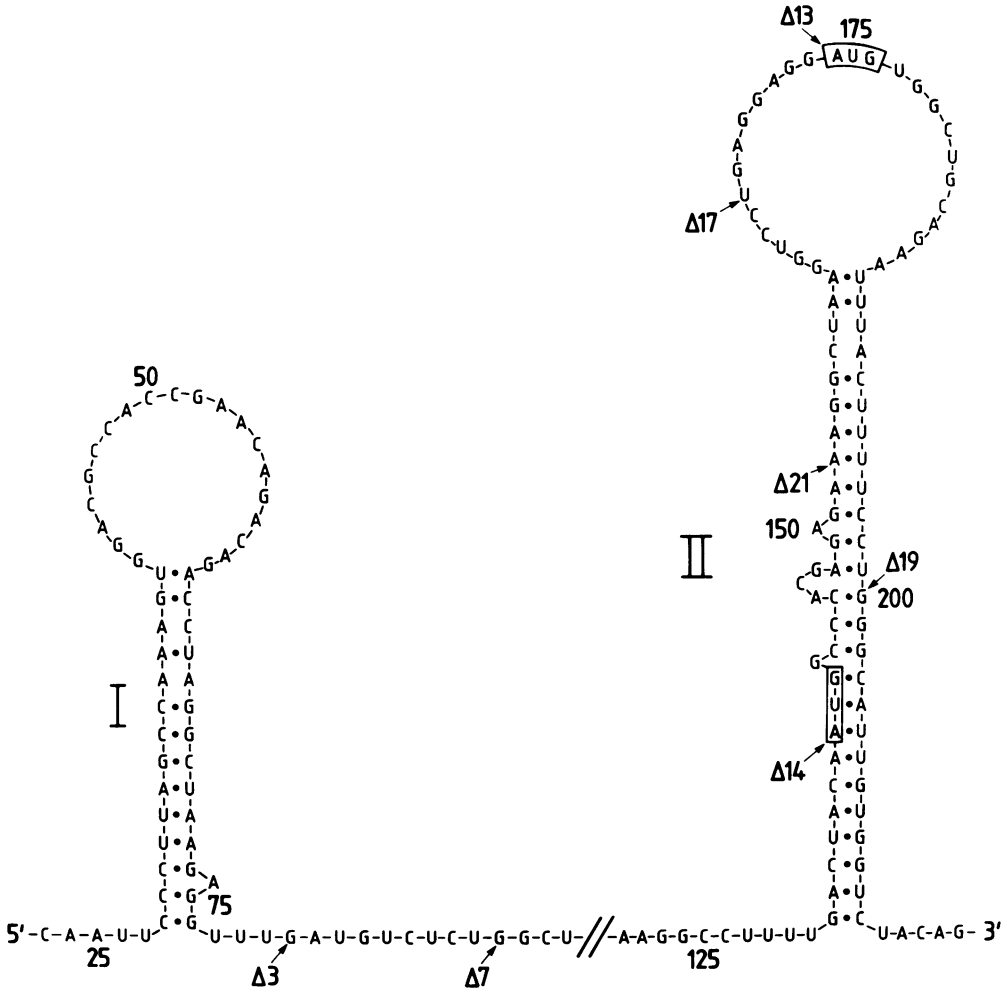

Fig. 3. Possible secondary structures towards the 5′ end of the GM-CSF mRNA. The numbering used is the same as in Figure 2. The two potential initiation codons are boxed and the 5′ end points of various deletion mutants are indicated. In calculating the free energy of structure I, the A.U bp involving residues 38 and 65 was taken to have a free energy of −1 kcal (Tinoco *et al.*, 1971). Otherwise the values of Salser (1977) were used.

tein occurs between residues 13 and 25. Cleavage of this leader peptide from the secreted protein appears to occur between Ser 29 and Ala 30 (arrowed in Figure 2), since the NH_2-terminal sequence of murine lung-derived and T cell-derived GM-CSF is Ala-Pro-Thr (A.W.Burgess, E.C.Nice, N.A.Nicola, R.Cutler and R.J.Simpson, personal communications). Moreover, the amino acid sequence of a human analogue of murine GM-CSF begins at the analogous Ala residue (Wong *et al.*, 1985). This is different to the NH_2-terminal amino acid sequence for asialo GM-CSF determined by Sparrow *et al.* (1984), which starts at Ile 36. Conceivably the NH_2-terminal sequence determined by Sparrow *et al.* was generated either by protease activity in the neuraminidase with which the GM-CSF was treated or during one of the other purification steps. The mature protein consists of 124 amino acids and has a calculated mol. wt. of 14 138 daltons. The first 12 residues of the predicted amino acid sequence of pre-GM-CSF are quite atypical for a secreted protein, since seven of them are charged (one acidic and six basic). Of the 90 sequences compiled by von Heijne (1983) 34 have no charged residues in the region NH_2-terminal to the hydrophobic core, 41 have one, 12 have two, one has three and only two have four.

Expression of GM-CSF by simian cells

To express eukaryotic cDNA sequences in simian cells [CVI and COS (Gluzman, 1980)] in culture, we have constructed a pair of SV40-based eukaryotic expression vectors, pJL3 and pJL4 (Figure 4). These vectors contain the β-lactamase gene and an origin of DNA replication from the bacterial plasmid pAT153 (Twigg and Sherratt, 1980), which lacks those plasmid sequences known to inhibit SV40 DNA replication in animal cells (Lusky and Botchan, 1981); the SV40 origin of DNA replication and T antigen coding sequences (Tooze, 1980); and a 'multi-cloning site' adjacent to the SV40 late promoter. The two vectors differ in the array of sites available for cloning in the multi-cloning site (see Figure 4). When introduced into cultured simian cells, such as CVI or COS, these vectors are able to replicate and a DNA sequence inserted at the multi-cloning site is transcribed and, if an initiation codon is present, translated.

Plasmid DNA from clone pGM3.2 or vector (pJL3) DNA or other negative control DNAs were introduced into COS cells (see Materials and methods), and the conditioned medium assayed 72 h post-transfection for CSF activity using the FDC P1 cell line, which is responsive to both GM-CSF and Multi-CSF (Hapel *et al.*, 1984; Metcalf, 1985). As shown in Figure 5A, medium from COS cells transfected with pGM3.2 DNA stimulated the proliferation of FDC P1 cells, whereas COS cells which received vector DNA (or DNA from recombinants containing only partial copies of GM-CSF mRNA) were quite inactive on FDC P1 cells. As the medium from COS cells transfected with pGM3.2 DNA did not stimulate 32D C13 cells (not shown), which respond only to Multi-CSF (Metcalf, 1985), we concluded provisionally that the factor in the COS cell-conditioned medium was

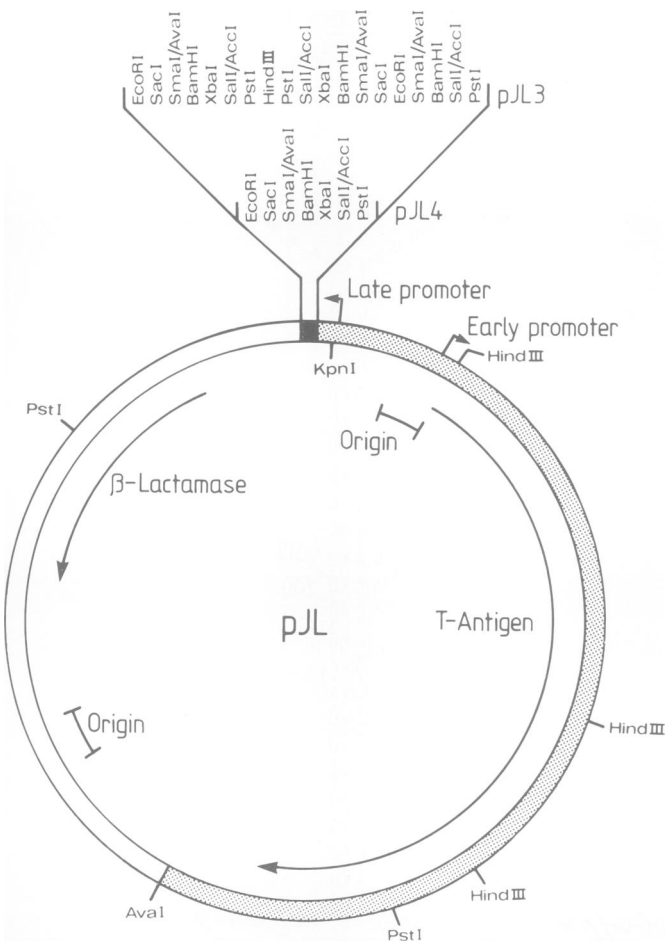

Fig. 4. Maps of the eukaryotic expression vectors pJL3 and pJL4. The portion derived from pAT153 (Twigg and Sherratt, 1980), indicated by an open segment, includes the β-lactamase gene and the origin of replication and extends from the *Ava*I site at position 1424 to the *Eco*RI site at 3740. The portion derived from SV40 (Tooze, 1980), indicated by a stippled segment, includes the T antigen coding sequences, the origin of DNA replication and the promoter of late transcription and extends from the *Nae*I (*Hpa*II) site at position 345 to the *Bam*HI site at 2533 (neither of these sites are retained in the pJL constructs). The multi-cloning site is filled-in. The arrangement of restriction endonuclease cleavage sites within the multi-cloning sites of pJL3 and pJL4 are indicated above the vector. The segments of DNA included in these multi-cloning sites were derived from pUC9 and 13 (Vieira and Messing, 1982). pJL3 is 5490 bp and pJL4 5417 bp in length.

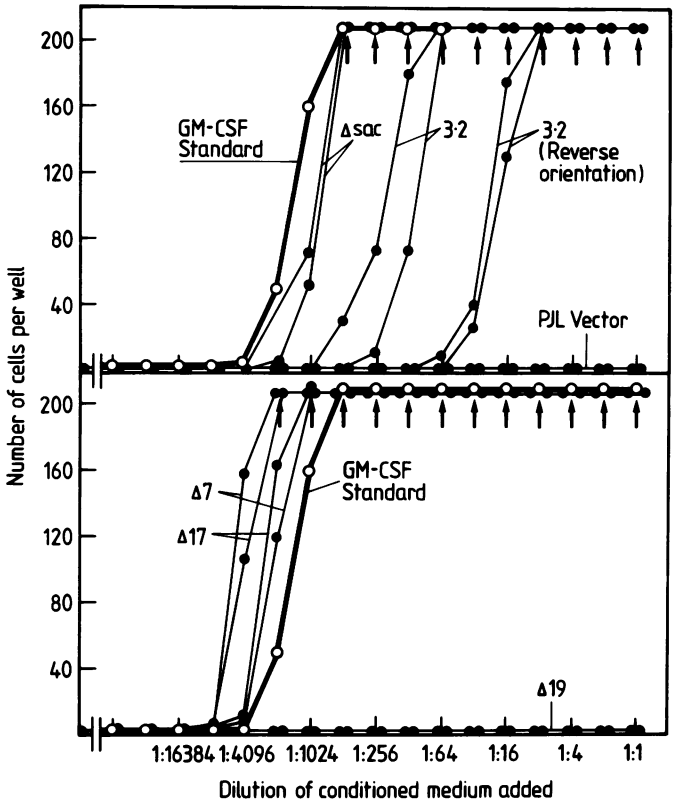

Fig. 5. Proliferation in microwell cultures of FDC P1 cells after stimulation by varying dilutions of duplicate media from COS cells transfected with DNA from clones pGM3.2, a derivative of pGM3.2 in which the cDNA is in the reverse orientation in the expression vector, pGM5′△Sac, pGM5′△7, pGM5′△17 and pGM5′△19. Heavy lines indicate the proliferation stimulated by GM-CSF purified from mouse lung conditioned medium (Burgess *et al.*, 1985). Note that transfection with vector alone did not result in GM-CSF synthesis. Each point represents mean cell counts in duplicate wells scored after 2 days of incubation.

GM-CSF. Surprisingly, medium from COS cells transfected with DNA from a derivative of pGM3.2 in which the insert is in the 'reverse' orientation, also contained low but detectable levels of GM-CSF (Figure 5A). Synthesis of GM-CSF directed by this recombinant is possibly due to transcripts of the cDNA which had initiated at pseudo-promoter sequences in the bacterial plasmid; it is unlikely to be due to transcription being promoted from within the cDNA, since we have observed the same phenomenon with a different cDNA, which specifies the murine Multi-CSF molecule (unpublished observations).

Our previous experience with expression of the hepatitis B viral surface antigen gene in pJL3 indicated that the presence of a tail of dG-dC residues between the SV40 late promoter and the initiation codon of the surface antigen gene markedly depressed the level of HBsAg synthesized. Therefore, to increase the level of GM-CSF synthesis in COS cells, we constructed a derivative of clone pGM3.2 (pGM5′△Sac) in which the fragment to the right of the *Sac*I site in Figure 1A, which contains the entire protein coding sequence, was inserted into pJL4, thereby deleting the dG-dC tails at the 5′ end of the cDNA (see Materials and methods). As shown in Figure 5A, this recombinant directed the synthesis of several-fold more GM-CSF than pGM3.2. In Table I the relative activities of these clones are quantified, and it can be seen that in this experiment pGM5′△Sac directed the synthesis of between 2.5- and 13-fold more GM-CSF than pGM3.2.

Can translation initiate at the second AUG?

To determine whether the second of the two potential initiation codons (Figure 2) can initiate translation of secreted, biologically active GM-CSF, we constructed a set of mutants with deletions spanning this region of the cDNA using the exonuclease *Bal*31 (Gray *et al.*, 1975). The regions of the cDNA contained within each mutant are shown in Figure 1B; the exact 5′ terminus of the GM-CSF sequence in each of the seven longest was determined by nucleotide sequence analysis and is indicated on Figures 2 and 3. Clones pGm5′△3, 7 and 14 have both initiation codons, clones pGm5′△21, 17 and 13 have only the second and pGM5′△19 and all shorter clones have neither. COS cells were transfected with DNA from each of the 14 clones and the conditioned media assayed for GM-CSF activity 72 h post-transfection using the FDC P1 cell line. Figure 5B shows the detailed results for clones pGM5′△7 and pGM5′△17 (the most active clones) and for pGM5′△19, which is inactive. In Table I the relative efficiency of all clones, in duplicate transfections,

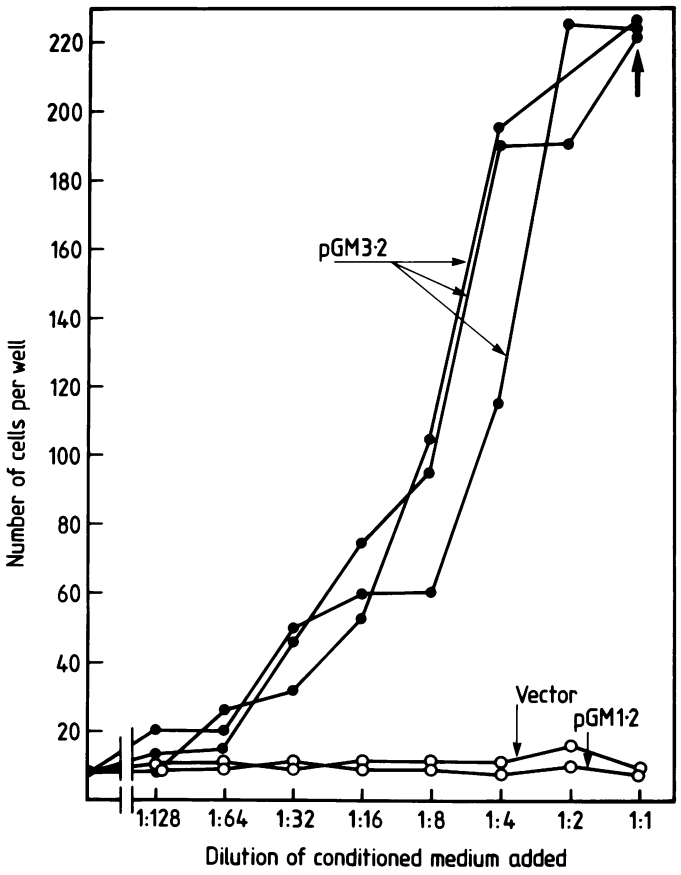

Fig. 6. Proliferation in microwell cultures of purified progenitor cells (Nicola *et al.*, 1979) after stimulation by varying dilutions of three batches of medium conditioned by COS cells after transfection with pGM3.2 DNA. Note that medium from COS cells transfected with vector DNA alone or with a clone containing an incompelete GM-CSF cDNA sequence (pGM1.2) failed to stimulate cell proliferation. Each point represents mean cell counts in duplicate wells scored after 2 days of incubation.

is quantified. All clones that have either both initiation codons (pGM5′△3, 7 and 14) or only the second (pGM5′△21, 17 and 13) directed the synthesis and secretion of GM-CSF, whereas all clones that have neither initiation codon were inactive. Thus it can be concluded that the second AUG can initiate translation of the GM-CSF mRNA and give rise to stable, secreted and biologically active GM-CSF. Furthermore, pGM5′△17 is reproducibly one of the most active clones (along with pGM5′△7, see Table I), indicating that translation of GM-CSF can be efficiently initiated at the second AUG and that secretion of GM-CSF can occur efficiently in the presence of a leader peptide starting with the second methionine residue. In this experiment, clones pGM5′△7 and pGM5′△17 directed the synthesis of 8- to 60-fold more GM-CSF than pGM3.2. The amount of GM-CSF synthesized by pGM5′△7-transfected COS cells was calculated to be 80 ng/ml.

Biological activities of GM-CSF synthesised by COS cells

Direct stimulation of haemopoietic progenitor cells. To determine whether the GM-CSF encoded by pGM3.2 or its derivatives was capable of direct stimulation of granulocyte-macrophage progenitor cells, microwell cultures were prepared each containing 200 progenitor cells (100% blast cells) purified by fluorescence-activated cell sorting from foetal liver populations (Nicola *et al.*, 1980). Assays using three pools of media from COS cells

Table I. Relative concentrations of GM-CSF in media conditioned by COS cells transfected with GM-CSF cDNA

cDNA clone	Initiation codons	GM-CSF concentration	
pGM3.2 (reverse)	Both	30	30
pGM3.2	Both	140	400
pGM5′△Sac	Both	1020	1900
pGM5′△3	Both	1500	4000
pGM5′△7	Both	8200	3000
pGM5′△14	Both	1500	1020
pGM5′△21	2nd	2500	2500
pGM5′△17	2nd	7000	3500
pGM5′△13	2nd	1020	700
pGM5′△19	Neither	0	0
pGM5′△26	Neither	0	0
pGM5′△15	Neither	0	0
pGM5′△32	Neither	0	0
pGM5′△30	Neither	0	0
pGM5′△25	Neither	0	0
pGM5′△8	Neither	0	0
pGM5′△29	Neither	0	0

Media from duplicate transfections of COS cells were assayed in serial dilutions in microwell cultures of FDC P1 cells. For the purposes of comparison, concentrations are expressed as the reciprocal of the dilution stimulating the survival and/or formation of 50 cells by 48 h. The initiation codons present in each clone are indicated (see text and Figure 2).

transfected with pGM3.2 showed that each was able to stimulate the proliferation of haemopoietic progenitor cells (Figure 6). Medium from COS cells transfected with vector alone or a cDNA clone containing an incomplete copy of the GM-CSF mRNA (pGM1.2) showed no proliferative activity. Analysis after 6 days of incubation of cultures of progenitor cells stimulated by medium from COS cells transfected with pGM3.2 or pGM5′△Sac showed that the initial progenitor cell population (100% blast cells) had generated maturing populations of granulocytes, macrophages and eosinophils (Table II). The composition of cells in these wells was similar to that in control wells stimulated by purified GM-CSF.

To confirm that the stimulating effects of recombinant GM-CSF were direct, 35 micromanipulated single granulocyte-macrophage precursor cells, from clones of 2−4 cells initiated 24 h previously in GM-CSF-stimulated cultures of foetal liver cells, were washed and each transferred to a separate 1 ml agar culture containing medium from pGM3.2 transfected COS cells. After 72 h of incubation, 29 (83%) of the cells had proliferated and formed clones of up to 90 maturing granulocytes and macrophages. No cells transferred to agar cultures lacking stimulus survived the incubation period and 89% of cells transferred to positive control cultures containing GM-CSF formed granulocyte-macrophage clones of up to 50 cells.

Action on bone marrow cultures. Medium conditioned by COS cells transfected either with pGM3.2 or pGM5′△Sac, when added to cultures of C57BL marrow cells, stimulated the formation of typical colonies containing granulocytes and/or macrophages together with an occasional eosinophil colony. No erythroid, multipotential or megakaryocyte colonies were detected (Table II). The proportions of the various colony types were similar to those in parallel cultures stimulated by purified GM-CSF. Medium from COS cells transfected with vector DNA alone or DNA from a cDNA clone containing only a partial copy of the GM-CSF mRNA (pGM1.2) exhibited no colony stimulating activity in cultures of adult marrow cells.

Table II. Formation of differentiating progeny in cultures of purified progenitor cells stimulated by medium from COS cells transfected with GM-CSF cDNA

Stimulus	Percentage of cells				
	Blasts	Promyelocytes myelocytes	Metamyelocytes polymorphs	Monocytes macrophages	Eosinophils
pGM3.2 1	0	7	83	6	4
2	0	13	79	3	5
3	0	15	72	6	7
pGM5′△Sac 1	0	14	76	10	0
2	0	15	71	14	0
Purified GM-CSF	0	24	57	17	2

Microwell cultures contained 200 progenitor cells (100% blast cells) purified by FACS sorting from 13 day CBA foetal liver (Nicola *et al.*, 1980). The three batches of medium from COS cells transfected with pGM3.2 contained an average of 300 units GM-CSF/ml and the two batches with pGM5′△Sac, 520 units/ml. The purified GM-CSF preparation contained 3000 units/ml. Cells were harvested on day 6 of incubation.

Table III. Stimulation of colony formation in cultures of adult marrow cells by medium from COS cells transfected with GM-CSF cDNA

Stimulus	Number of samples tested	Mean GM-CSF concentration units/ml	Percentage of colonies			
			G	GM	M	Eo
pGM3.2	4	310 ± 160	15 ± 13	20 ± 11	63 ± 22	2 ± 2
pGM5′△Sac	7	520 ± 420	24 ± 5	21 ± 10	55 ± 7	0
pGM1.2	2	0	–	–	–	–
Vector	4	0	–	–	–	–
Purified GM-CSF		3000	33	26	40	1

Serial dilutions of 0.1 ml of test samples were assayed in replicate cultures containing 75 000 C57BL marrow cells. Colonies were counted on day 7 and differential colony counts performed on whole stained cultures. The GM-CSF concentration was calculated from colony numbers. G, granulocyte; GM, granulocyte-macrophage; M, macrophage; Eo, eosinophil.

Table IV. Colony stimulating activity on foetal liver haemopoietic cells of medium from COS cells transfected with GM-CSF cDNA

Stimulus	Mean number of colonies	Percentage of colonies					
		G	GM	M	Eo	E/E Mix	Meg
No supplement							
pGM3.2 1	6 ± 2	22	6	50	22	0	0
2	8 ± 3	46	8	46	0	0	0
3	10 ± 2	11	20	60	9	0	0
4	45 ± 6	13	6	76	5	0	0
Purified GM-CSF	20 ± 6	13	33	43	11	0	0
Saline	0	–	–	–	–	–	–
Addition of SCM on day 2							
pGM3.2 1	10 ± 4	6	27	33	18	16	0
2	15 ± 2	8	21	25	21	25	0
3	17 ± 2	21	33	12	22	6	6
4	48 ± 11	5	14	58	4	19	0
Purified GM-CSF	47 ± 5	14	20	29	10	27	0
Saline	4 ± 2	0	0	100	0	0	0

Four batches of media from COS cells transfected with pGM3.2 were each assayed in eight replicate cultures containing 20 000 12 day CBA foetal liver cells. Control cultures contained 300 units of purified GM-CSF or 0.1 ml of saline. In four replicate cultures, colonies were counted on day 7 of incubation and differential counts performed from whole-culture preparations stained with Luxol Fast Blue-hematoxylin ('No supplement'). To the other four replicate cultures, on day 2 of incubation 0.1 ml pokeweed mitogen-stimulated spleen conditioned medium (SCM) was added to sustain proliferation of erythroid and multi-potential progenitor cells initiated by primary the stimulus ('Addition of SCM'). G, granulocyte; GM, granulocyte-macrophage; M, macrophage; Eo, eosinophil; E, pure erythroid; E Mix, multipotential colonies containing erythroid cells; Meg, megakaryocytic.

Action on foetal liver cultures. Medium from COS cells transfected with pGM3.2 cDNA when added to cultures of 20 000 12 day CBA foetal liver cells stimulated the formation of colonies containing granulocyte and/or macrophages as well as typical eosinophil colonies (Table IV). These colonies were similar to those stimulated in control cultures by purified GM-CSF. Although media from COS cells transfected with vector alone did not stimulate cell proliferation in cultures of adult marrow cells, such control media did stimulate the formation by foetal liver cells of low numbers of small macrophage-containing clones, usually of sub-colony size. This macrophage-stimulating material of COS cell origin may have contributed to the somewhat higher proportions of macrophage colonies seen in foetal liver cultures stimulated by pGM3.2-derived material than seen in cultures stimulated by purified GM-CSF.

Delayed addition of spleen conditioned medium (as a source of Multi-CSF) to cultures initiated by purified GM-CSF leads to the additional development of large colonies containing erythroid cells or a mixture of erythroid and other haemopoietic cells (Burgess and Metcalf, 1980; Metcalf *et al.*, 1980). This

was observed also in the present experiments (Table IV). It can be seen that the delayed addition of SCM to cultures initiated with medium conditioned by COS cells transfected with pGM3.2 led to the formation of a large erythroid or mixed erythroid colonies (in equal numbers) and an occasional megakaryocyte colony.

Discussion

GM-CSF synthesised by COS cells is biologically authentic

Medium conditioned by COS cells transfected either with the cDNA clone pGM3.2 or pGM5′△Sac exhibited all the known biological properties of purified GM-CSF. These are (i) selective stimulation of the proliferation of FDC P1 but not 32D cells. (ii) Stimulation of granulocyte and/or macrophage colony formation by adult or foetal progenitor cells, (iii) stimulation of eosinophil colony formation by mouse foetal liver cells, and (iv) the initiation, but not maintenance, of proliferation of foetal erythroid, multipotential and megakaryocyte progenitors. From the experiments in which purified or micromanipulated single progenitor cells generated populations of differentiating granulocytes and macrophages, it can be concluded that the effects of the active material in the COS cell conditioned media were direct and not dependent upon interactions with other cell types in haemopoietic populations.

These observations demonstrate that the clone pGM3.2 does indeed contain a copy of the murine GM-CSF mRNA and verify that the above functions previously attributed to purified GM-CSF (Burgess and Metcalf, 1980; Metcalf *et al.*, 1980; Metcalf and Nicola, 1983), are indeed the properties of the product of a single gene. However, medium from COS cell cultures must contain many macromolecules of COS cell origin; some of these might conceivably act directly on haemopoietic cells or modulate the action of CSF, as possibly was the case with the increased macrophage colony formation in foetal liver cultures. Thus medium from COS cells transfected with cloned GM-CSF cDNA cannot be regarded as containing only GM-CSF and fractionation of such medium would be required if pure preparations of GM-CSF were required. It is likely also that the pattern of glycosylation of GM-CSF synthesised by COS cells may differ from that of GM-CSF from murine cells. However, from the assays performed *in vitro*, COS cell-derived GM-CSF is functionally indistinguishable from GM-CSF purified from murine sources.

Translation of the GM-CSF mRNA

Which of the two AUG codons located close to the start of the major open reading frame of the GM-CSF mRNA (see Figure 2) is used to initiate translation? In 95% of eukaryotic mRNAs translation begins at the AUG codon closest to the 5′ end of the mRNA and Kozak has proposed the so-called 'scanning model' in which the 40S ribosomal subunit 'finds' the initiation codon by scanning the 5′ end of the mRNA (Kozak, 1978, 1984). Moreover, the nucleotide sequences surrounding functional initiation codons tend to conform to the consensus sequence $CC\overset{A}{G}CCAUG(G)$ (Kozak, 1984). Inspection of the sequences surrounding the two potential initiation codons in the GM-CSF mRNA (Figure 2) reveals that the sequence surrounding the first AUG (CUACAAUGG) conforms well to the Kozak consensus sequence, whereas the second AUG (GGAGGAUGU) conforms less well, since it has a uracil residue immediately after and guanosine residues preceding the AUG. Thus the scanning model would predict that the first AUG is used to initiate translation of the GM-CSF mRNA.

However, two observations suggest that the second AUG codon may also initiate translation. Firstly, this AUG codon *can* initiate translation of active, secreted GM-CSF when the first AUG is deleted (e.g., in clones pGM5′△21, pGM5′△17 and pGM5′△13) and indeed the levels of GM-CSF directed by pGM5′△21 and pGM5′△17 are as high as that directed by clones having both AUG codons (Table I). The second observation supporting the idea that the second AUG may be used to initiate translation is its location within an open loop at the end of a region with high degree of secondary structure (Figure 3), in contrast with the location of the first AUG in the base-paired stem of this structure. Since initiation codons in exposed regions appear to be utilised more efficiently than those buried within hydrogen-bonded structures (Kozak, 1980; Pavlakis *et al.*, 1980) it is likely that the second AUG may also be used to initiate translation.

Can GM-CSF exist as an integral membrane protein?

Whilst there is clear evidence for secretion of GM-CSF by many cells types, the occurrence of islands of differentiating granulocytes in contact with adherent bone marrow stromal cells has raised the possibility that some cells may stimulate granulopoiesis by displaying GM-CSF on their membranes (Allen, 1981; Allen and Dexter, 1976; Bentley, 1981; Williams and Burgess, 1980; Zipori, 1981). Indeed, colony stimulating activity has been detected in cell membrane extracts (Price *et al.*, 1975).

Interestingly, the sequence of the GM-CSF leader peptide lends support to the notion that GM-CSF might be able to exist as a membrane bound protein in addition to the secreted form normally observed. Since the GM-CSF mRNA has two different translational initiation codons, two different polypeptides differing at their NH_2 terminus could be specified. Polypeptides that initiate at the second AUG can clearly be precursors of a secreted protein, since we have shown that mutants which must initiate at this codon are secreted from COS cells and indeed the protein specified has the essential characteristics of a secreted protein: a core of hydrophobic amino acids preceding the site at which the leader region is cleaved from the mature protein. If translation initiates instead at the first AUG the polypeptide specified has several features typical of an integral membrane protein and we speculate that these molecules might be presented on the cell surface. On this model, the sequence of 20 uncharged and predominantly hydrophobic amino acids (residues 13 – 32) would represent the transmembrane segment: whilst this region presumably acts as the signal sequence for secretion of the protein it is also long enough to span the lipid bilayer and could conceivably fulfill both roles, as suggested for certain other cell-surface proteins (Holland *et al.* 1984). Whilst the NH_2-terminal 12 residues, which are highly charged, are atypical for a secreted protein (see Results), they are reminiscent of the hydrophilic intracellular sequences of integral membrane proteins. In particular those which are attached to the cell membrane *via* their NH_2 termini are also highly charged in this region (Drickamer, 1981; Drickamer *et al.*, 1984; Schneider *et al.*, 1984; Strubin *et al.*, 1984). Thus we speculate that the NH_2 terminus of the GM-CSF molecule would be intracellular and the major C-terminal portion of the protein extracellular and hence available to interact with neighbouring cells.

It is conceivable that cells might exert control over the nature of the GM-CSF synthesised by manipulating which AUG is used for initiation of translation and it is tempting to speculate that the RNA secondary structure illustrated in Figure 3 might play a role in this process. However, since in the GM-CSF gene the first initiation codon is not specified on the same exon as the se-

cond (E.Stanley, personal communication), it is also possible that differential RNA splicing might be able to generate two different mRNAs, differing in their coding capacity at the 5′ end.

Materials and methods

cDNA library construction and screening

LB3, a cloned T lymphocyte line (Kelso *et al.*, 1984) was maintained and stimulated with concanavalin A as described (Kelso and Metcalf, 1985). Poly(A)-containing RNA (Gough, 1983) and double-stranded cDNA (Gough *et al.*, 1984) were prepared as described. The cDNA, tailed with dC residues, was fractionated by agarose gel electrophoresis and molecules >500 bases recovered and annealed to pJL3 plasmid DNA (Figure 4) cleaved with *Sac*I and tailed with dG residues. *Escherichia coli* MC1061 (Casadaban and Cohen, 1980) was transformed with the cDNA-plasmid mixture to generate a library of ~40 000 independent recombinants. Clones containing GM-CSF recombinants were identified by colony hybridization using a probe derived from clone pGM38 (see Figure 1A) labelled by nick-translation (Rigby *et al.*, 1977).

Construction of 5′ deletion mutants

(i) pGM5′△Sac: the fragment extending from the *Sac*I site at the 5′ end of the pGM3.2 cDNA (Figure 1A) to the *Eco*RI site located in the multi-cloning site 3′ to the cDNA was inserted between the *Sac*I and *Eco*RI sites of pJL4 (Figure 4). (ii) A set of mutants with random deletions at the 5′ end of the cDNA sequence were constructed as follows: plasmid DNA from a subclone containing the cDNA fragment of pGM3.2 in the *Eco*RI site of pUC8 (Vieira and Messing, 1982) was cleaved at the 5′ side of the cDNA with *Bam*HI and digested with exonuclease *Bal*31 (Gray *et al.*, 1975). After addition of *Bam*HI linkers, the resected plasmids were recircularised by ligation and *E. coli* MC1061 (Casadaban and Cohen, 1980) transformed with the recircularised plasmids generating 20 000–30 000 clones. Plasmid DNA prepared from this pool of clones was digested with *Eco*RI and *Bam*HI, fractionated by agarose gel electrophoresis and truncated cDNA molecules ligated to pJL4 DNA (Figure 3) digested with *Bam*HI and *Eco*RI. After transformation of *E. coli* MC1061, individual clones were picked for further analysis.

Nucleotide sequence analysis

DNA fragments subcloned in M13 vectors (Messing and Vieira, 1982) were sequenced by the chain-termination method using dideoxynucleoside triphosphates (Sanger *et al.*, 1977). Sequencing reactions were electrophoresed on thermostatically controlled, 0.2 mm thick, 8% (w:v) polyacrylamide gels (Garoff and Ansorge, 1981).

Transfection of COS cells with DNA

DNA transfections were carried out essentially as described by Danna and Sompayrac (1982) using 10 μg of plasmid DNA in 750 μg/ml DEAE dextran. COS cells (Gluzman, 1981) were seeded in 5 cm dishes 24 h prior to transfection and were 30% confluent on the day of use. Prior to transfection cells were rinsed in warm DME and exposed to the DNA/DEAE dextran mix for 4 h at 37°C. The DNA solution was then removed, replaced with a solution of 25% glycerol in complete growth medium for 1–2 min, washed twice in growth medium and finally incubated at 37°C in DME plus 10% foetal calf serum. Medium was harvested for assay 72 h post-transfection.

Cell lines

Two factor-dependent continuous cell lines were used: (i) 32D clone 3 cells (Greenberger *et al.*, 1982), supplied by Dr. J.Greenberger, Boston, which respond to stimulation by Multi-CSF and cannot survive or proliferate in the presence of GM-CSF, M-CSF or G-CSF (Metcalf, 1985) and (ii) FDC P1 cells originating from Dr. T.M.Dexter, Manchester (Dexter *et al.*, 1980). The present subline responds to proliferative stimulation either by Multi-CSF or GM-CSF (Hapel *et al.*, 1984; Metcalf, 1985) but not by M-CSF or G-CSF.

CSF assays

Mouse bone marrow assays. Assays of 0.1 ml of serial dilutions of test material were performed in 1 ml of agar-medium using 75 000 3 month old C57BL bone marrow cells (Metcalf, 1984). Colonies (clones of >50 cells) were scored on day 7 and colonies typed from fixed cultures stained with Luxol-Fast-Blue haematoxylin. CSF units were calculated from the linear portion of the dose-response curves, assigning 50 units to the concentration stimulating the formation of half-maximal colony numbers.

Foetal liver assays. Cultures were performed as above using medium with 20% heat-inactivated human plasma instead of foetal calf serum and 20 000 12 day CBA foetal liver cells. To document the ability of test material to initiate but not to sustain proliferation of erythroid and multipotential precursors, 0.1 ml of pokeweed mitogen stimulated spleen conditioned medium (containing Multi-CSF) was added at day 2 to the surface of replicate cultures. Colony counting and typing were performed as for marrow cultures after a total incubation period of 7 days.

Microwell assays. Replicate wells of Lux 5620 HLA tissue culture plates (Miles Laboratory, Naperville, IL) contained 5 μl of serial dilutions of assay material, 10 μl of Dulbecco's Modified Eagle's Medium containing 20% foetal calf serum and 200 FDC P1 or 32D cells or 200 progenitor cells fractionated by fluorescence-activated cell sorting from 13 day foetal liver cells (Nicola *et al.*, 1980). Cultures were incubated for 48 h at 37°C in a fully humidifed atmosphere of 10% CO_2 in air, then cell counts performed using an inverted microscope. Control wells contained serial dilutions of purified GM-CSF (Burgess *et al.*, 1985) and CSF units were calculated on the basis of these control cultures.

Acknowledgements

We are extremely grateful to Anne Kelso for providing LB3 cells, Nick Nicola for FACS-fractionated progenitor cells and Tony Burgess for purified GM-CSF. We also thank Tony Burgess, Jim Goding, Nick Nicola, Anne Kelso, Tom Gonda and Yasha Gluzman for discussions, Jim Goding and Jerry Adams for useful comments on the manuscript, Cathy Quilici for technical assistance and Sue Blackford for her ever-patient typing. The work at the Walter and Eliza Hall Institute was supported by the Carden Fellowship Fund of the Anti-Cancer Council of Victoria, the National Health and Medical Research Council, Canberra and the National Institutes of Health, Bethesda Grant No. CA22556.

References

Adams,J.M. (1977) in Stewart,P.R. and Letham,D.S. (eds.), *The Ribonucleic Acids*, Springer-Verlag, NY, pp. 81-128.
Allen,T.D. (1981) in Porter,R. and Whelan,J. (eds.) *Microenvironments in Haemopoietic and Lymphoid Differentiation*, CIBA Foundation Symposium **84**, Pitman Medical, London, pp. 38-67.
Allen,T.D and Dexter,T.M. (1976) *Differentiation*, **6**, 191-194.
Bentley,S.A. (1981) *Exp. Hematol.*, **9**, 308-312.
Blackburn,D.E., Hobbs,A.A. and Rosen,J.M. (1982) *Nucleic Acids Res.*, **10**, 2295-2307.
Burgess,A.W. and Metcalf,D. (1977) in Baum,S.J. and Ledney,G.D. (eds.), *Experimental Haematology Today*, Springer, NY, pp. 135-146.
Burgess,A.W. and Metcalf,D. (1980) *Blood*, **56**, 947-958.
Burgess,A.W., Camakaris,J. and Metcalf,D. (1977) *J. Biol. Chem.*, **252**, 1988-2003.
Burgess,A.W., Metcalf,D., Sparrow,L.G. and Nice,E.C. (1985) *J. Biol. Chem.*, in press.
Casadaban,M.J. and Cohen,S.N. (1980) *J. Mol. Biol.*, **138**, 179-207.
Cochet,M., Gannon,F., Hen,R., Maroteaux,L., Perrin,F. and Chambon,P. (1979) *Nature*, **282**, 567-574.
Danna,K.J. and Sompayrac,L.M. (1982) *J. Virol. Methods*, **5**, 335-341.
Dexter,T.M., Garland,J., Scott,D., Scolnick,E. and Metcalf,D. (1980) *J. Exp. Med.*, **152**, 1036-1041.
Drickamer,K. (1981) *J. Biol. Chem.*, **256**, 5827-5839.
Drickamer,K., Mamon,J.F., Binns,G. and Leung,J.O. (1984) *J. Biol. Chem.*, **259**, 770-778.
Efstratiadis,A., Kafatos,F.C. and Maniatis,T. (1977) *Cell*, **10**, 571-585.
Garoff,H. and Ansorge,W. (1981) *Anal. Biochem.*, **115**, 450-457.
Ghosh,P.K., Reddy,V.B., Swinscoe,J., Choudary,P.V., Lebowitz,P. and Weissman,S.M. (1978) *J. Biol. Chem.*, **253**, 3643-3647.
Gluzman,Y. (1981) *Cell*, **23** 175-182.
Gough,N.M. (1983) *J. Mol. Biol.*, **165**, 683-699.
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.
Gray,H.B.,Jr., Ostrander,D.A., Hodnett,J.L., Legerski,R.J. and Robberson,D.L. (1975) *Nucleic Acids Res.*, **2**, 1459-1492.
Greenberger,J.S., Hapel,A., Nabel,G., Eckner,R.J., Newburger,P.E., Ihle,J., Denburg,J., Moloney,W.C., Sakakeeny,M. and Humphries,K. (1982) in Baum,S.J., Ledney,G.D. and Thierfelder,S. (eds.), *Experimental Haematology Today*, Karger, Basel, pp. 195-209.
Hapel,A.J., Warren,H.S. and Hume,D.A. (1984) *Blood*, **64**, 786-790.
Holland,E.C., Leung,J.O. and Drickamer,K. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7338-7342.
Jeffery,W.R. and Brawerman,G. (1974) *Biochemistry (Wash.)*, **13**, 4633-4637.
Kelso,A. and Metcalf,D. (1985) *Exp. Hematol.*, in press.
Kelso,A., MacDonald,H.R., Smith;K.A., Cerottini,J.-C. and Brunner,K.T. (1984) *J. Immunol.*, **132**, 2932-2938.
Kozak,M. (1978) *Cell*, **15**, 1109-1123.
Kozak,M. (1980) *Cell*, **19**, 79-90.
Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857-872.
Lusky,M. and Botchan,M. (1981) *Nature*, **293**, 79-81.
McReynolds,L., O'Malley,B.W., Nisbet,A.D., Fothergill,J.E., Givol,D., Fields,S., Robertson,M. and Brownlee,G.G. (1978) *Nature*, **273**, 723-728.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269-276.

Metcalf,D. (1984) *The Haemopoietic Colony Stimulating Factors*, published by Elsevier, Amsterdam.
Metcalf,D. (1985) in Stomatoyannopoulos,G. and Neinhuis,A. (eds.), *Hameoglobin Switching*, Academic Press, NY, in press.
Metcalf,D. and Nicola,N.A. (1983) *J. Cell. Physiol.*, **116**, 198-206.
Metcalf,D., Johnson,G.R. and Burgess,A.W. (1980) *Blood*, **55**, 138-147.
Nicola,N.A., Burgess,A.W., Staber,F.G., Johnson,G.R., Metcalf,D. and Battye,F. (1980) *J. Cell. Physiol.*, **103**, 217-237.
Pasek,M., Goto,T., Gilbert,W., Zink,B., Schaller,H., McKay,P., Leadbetter,G. and Murray,K. (1979) *Nature*, **282**, 575-579.
Pavlakis,G.N., Lockard,N.E., Vamvakopoulos,N., Rieser,L., RajBhandary,V.L. and Vournakis,J.N. (1980) *Cell*, **19**, 91-102.
Perlman,D. and Halvorson,H.O. (1983) *J. Mol. Biol.*, **167**, 391-409.
Price,G.B., McCulloch,E.A. and Till,J.E. (1975) *Exp. Hematol.*, **3**, 227-233.
Reynolds,G.A., Basu,S.K., Osborne,T.F., Chin,D.J., Gil,G., Brown,M.S., Goldstein,J.L. and Luskey,K.L. (1984) *Cell*, **38**, 275-285.
Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237-251.
Salser,W. (1977) *Cold Spring Harbor Symp. Quant. Biol.*, **42**, 985-102.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Schneider,R.C., Owen,M.J., Banville,D. and Williams,J.G. (1984) *Nature*, **311**, 675-678.
Sparrow,L.G., Metcalf,D., Hunkapiller,M.W., Hood,L.E. and Burgess,A.W. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 292-296.
Strubin,M., Mach,B. and Long,E.O. (1984) *EMBO J.*, **3**, 869-872.
Tinoco,I.,Jr., Unlenbeck,O.C. and Levine,M.D. (1971) *Nature*, **230**, 362-367.
Tinoco,I.,Jr., Borer,P.N., Dengler,B., Levine,M.D., Unlenbeck,O.C., Crothers,D.M. and Gralla,J. (1973) *Nature New Biol.*, **246**, 40-41.
Tooze,J., ed. (1980) in *Molecular Biology of Tumour Viruses, Part 2: DNA Tumour Viruses*, published by Cold Spring Harbor Laboratory Press, NY.
Twigg,A.J. and Sherratt,D. (1980) *Nature*, **283**, 216-218.
Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259-268.
von Heijne,G. (1983) *Eur. J. Biochem.*, **133**, 17-21.
Williams,N. and Burgess,A.W. (1980) *J. Cell. Physiol.*, **102**, 287-295.
Wong,G.G., Witek,J.A., 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., Kaufman,R.J., Hewick,R.M., Wang,E.A. and Clark,S.C. (1985) *Science (Wash.)*, in press.
Zipori,D. (1981) *J. Supramol. Struct.*, **17**, 347-357.

Received on 20 December 1984; revised on 17 January 1985