Identification of two elements involved in regulating expression of the murine leukaemia inhibitory factor gene

Li-Wei HSU* and John K. HEATH†

CRC Growth Factor Group, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

Mouse leukaemia inhibitory factor (LIF) is a polyfunctional cytokine which exhibits multiple functions *in vitro* and *in vivo*. Two forms of LIF cDNA, differing at their 5′ ends, have been described encoding either diffusible (D-LIF) or matrix -associated (M-LIF) forms of the protein [Rathjen, Toth, Willis, Heath and Smith (1990) (Cell **62**, 1105–1114]. The present report describes the DNA sequence and functional characterization of the murine LIF gene and its surrounding transcriptional regulatory elements. Transient transfection of constructs containing the LIF gene and various amounts of 5′-non-coding sequence failed to give detectable levels of expression, suggesting the presence of inhibitory sequences within the LIF gene. Stable cell lines were produced by transfection of experimental constructs containing various lengths of 5′-non-coding sequence of the LIF gene, or the heterologous phosphoglycerate kinase promoter, linked to an LIF/neomycin-resistance-hybrid-coding sequence. The frequency of recovery of stable clones indicated that sequences located in the first intron between the transcriptional start sites for D-LIF and M-LIF act to suppress expression of the gene in most genomic locations. This region is rich in GC residues and has been shown to be hypomethylated *in vitro* [Kaspar, Dvorak and Bartunek (1993) FEBS Lett. **319**, 159–162]. Analysis of the LIF/neomycin-resistance transgene expression in these stable cell clones demonstrated that transcripts containing the M-LIF or D-LIF exons required the presence of sequences located between −1200 and −3200 in the LIF gene. In the absence of these sequences, transcription is initiated elsewhere within the first intron. These sequences can be replaced by the heterologous phosphoglycerate kinase promoter. Deletion of the GC-rich region between the D-LIF and M-LIF transcriptional start sites results in the appearance of transcripts that do not splice out the first intron of the LIF gene. These may result from gene or promoter trapping of the LIF gene. Sequence analysis of the region between −1200 and −3200 revealed a number of minimal steroid-response elements, regions of similarity to DNAase I-hypersensitive sites in the uteroglobin gene and a region of alternating purine/pyrimidine sequence. This study therefore defines two important regulatory regions in the LIF gene: a GC-rich region in the first intron and a distal 'enhancer' located between −3200 and −1200.

INTRODUCTION

Leukaemia inhibitory factor (LIF) is a polyfunctional growth factor which exhibits diverse biological activities *in vitro* including maintenance of embryonic stem cell pluripotentiality, induction of proliferation, differentiation and cell survival in certain cell lines of haemopoietic origin, induction of the acute-phase response programme of gene expression in hepatocytes, regulation of neurotransmitter expression in peripheral neurons and induction of bone resorption [for review, see Hilton (1992)]. *In vivo* grafting of haemopoietic cell lines overexpressing LIF into adult mice, or systemic administration of LIF protein, leads to complex physiological effects including bone resorption, irritable behaviour, weight loss, extramedullary haemopoiesis in spleen and liver, thymus atrophy (Metcalf and Gearing, 1989) and increase in circulating megakaryocytes and platelets (Metcalf et al., 1991, 1992). In addition, intravenous injection of LIF can protect mice from lethal lipopolysaccharide-induced toxicity (Alexander et al., 1992). Genetic inactivation of the mouse LIF gene by gene-targeting experiments have furthermore demonstrated that maternal expression of LIF is essential for embryo implantation (Stewart et al., 1992; Escary et al., 1993).

Two different LIF gene transcripts have been demonstrated in the mouse (Rathjen et al., 1990a). The isolation of LIF cDNAs revealed that these two distinct transcripts exhibit distinct 5′ ends, arising from initiation of transcription from distinct genomic locations. Analysis of the expression of the two forms of LIF transcript and localization of LIF bioactivity after transfection of LIF cDNAs into cultured cells (Rathjen et al., 1990a) has revealed that the two forms of LIF transcript encode proteins that are either released into the cell culture medium (D-LIF) or associated with the extracellular matrix (M-LIF). The association of LIF with the extracellular matrix appears to be an inherent property of the mature LIF protein and, in at least some situations, is brought about by interaction with a high-affinity LIF-binding protein, gp140, which is itself matrix bound (Mereau et al., 1993).

Further examination of LIF gene expression *in vitro* has shown that LIF is not constitutively expressed but subject to regulation by both cell differentiation and exogenous agents including other growth factors, such as fibroblast growth factor and transforming growth factor β, and glucocorticoids (Rathjen et al., 1990b). Studies *in vivo* have shown that LIF gene expression in the adult is principally restricted to the uterus in which a remarkable burst occurs on day 4 of gestation in the mouse, coincident with the process of embryo implantation (Bhatt et al., 1991). This uterine expression of LIF *in vivo* appears to be controlled, at least in part, by maternal steroid hormones, as expression is suppressed in mice artificially induced to undergo delayed implantation by ovariectomy and administration of progesterone but induced on release of implantational delay by injection of oestrogen (Bhatt et al., 1991). In addition, the level

Abbreviations used: LIF, leukaemia inhibitory factor; D-LIF, diffusible LIF; M-LIF, matrix-associated LIF; Neo, neomycin-resistance coding sequence; DMEM, Dulbecco's modified Eagle's medium; FCS, fetal calf serum; PGK, phosphogylcerate kinase; hGH, human growth hormone.

* Present address: Chang-Gung Medical College, 259 Wen-Hwa 1 Road, 33332 Kwei-san, Tao-Yuan, Taiwan, Republic of China

† To whom correspondence should be addressed.

of LIF mRNA in the uterus (both M and D forms) fluctuates with the oestrous cycle (Shen and Leder, 1992).

These findings raise a number of issues regarding both the organization and regulation of the LIF gene. These include identification of the genetic origins of the different LIF gene transcripts as well as the identity, localization and design of transcriptional regulatory elements involved in the regulation of LIF gene expression by both external signals and cell differentiation. In this study we have analysed both the genetic organization and transcriptional regulation of the LIF gene by transfection of hybrid LIF gene transgenes into cultured fibroblast cell lines. These studies reveal that the LIF gene contains two principal genetic regulatory elements: a 5′ distal enhancer region, which is required for expression of both M-LIF and D-LIF transcripts, and a second element, located between the D-LIF and M-LIF transcriptional initiation sites, the major activity of which has been defined as promoting expression of the LIF gene depending on its genomic location. This region has been shown to be subject to hypomethylation in cultured cells *in vitro* (Kaspar et al., 1993). In addition, DNA sequence analysis of the LIF gene and its surrounding DNA sequences has revealed the location of regions in the 5′ distal control element that may be involved in regulation of the LIF gene by maternal steroid hormones.

MATERIALS AND METHODS

Isolation of mouse LIF gene from bacteriophage λ genomic DNA library and DNA sequencing

Recombinant bacteriophage containing the mouse LIF gene and its surrounding regions were isolated by screening a genomic library derived from partial Sau3A-digested CBA/C57B16 F1 mouse DNA cloned into the *Xho*II site of λFIXII vector (Stratagene). The full-length murine D-LIF-coding sequence (Rathjen et al, 1990a) was used as a probe. Library screening was performed according to the manufacturer's recommendations. Positively hybridizing clones were subjected to further rounds of screening, plaque purification and partial characterization by restriction mapping and Southern blotting of digested DNA with murine LIF cDNAs.

One genomic clone (1LIF1) was found to contain a ~ 20 kb insert which included the entire LIF-gene-coding sequence. An 11 kb *Eco*RI fragment cut from 1LIF1, which included the complete LIF gene, was subcloned into pRS313 vector (Christianson et al., 1992) and designated pRSLIF11. The regions of pRSLIF11 that included the LIF gene and 5′ regions were sequenced by shotgun cloning into M13 and DNA sequencing by the dideoxynucleotide-mediated chain-termination sequencing method (Sanger et al., 1977). The DNA sequence was assembled and analysed using the GCG software package (Devereux et al., 1984).

Construction of vectors for transient transfection

Three mouse LIF-gene-expression vectors, pBSLIF4polyA, pBSLIF3polyA and pBSLIF2.5polyA, were constructed for transient-transfection studies. A 4 kb *Xba*I fragment cut from 1LIF1 was subcloned into pBluescriptIIKS vector (pBSLIF4). This 4 kb fragment contains the entire coding sequence and ~ 1.2 kb of 5′ flanking region. The 3′ polyadenylation element of the pEFBOS vector (Mizushima and Nagata, 1990) was ligated to *Eco*RI-linearized pBSLIF4 plasmid. The orientation of insertion was checked by sequencing. This poly(A) tail containing LIF gene was designated pBSLIF4polyA. pBSLIF3polyA was derived from this plasmid by digestion with *Bam*HI which deletes regions 5′ to the D-LIF transcriptional start site. pBSLIF2.5polyA, which excludes all 5′-non-coding elements, the TATA boxes and the D-form-specific first exon, was derived by *Sac*II digestion and religation.

Construction of vectors for stable transfection

Five LIF gene vectors were derived for stable-transfection studies, pBSLIF11Neo, pBSLIF9Neo, pBSLIF7Neo pBSLIF6.5Neo and PGKLIF3Neo (Figure 1). A 9 kb LIF genomic DNA fragment was subcloned into pBluescriptIIKS vector (pBSLIF9) by manipulating pBSLIF4 and the 5 kb 3′ untranslated region obtained from vector pRSLIF11. A promoterless neomycin-resistance coding sequence (Neo) was derived from pSV2Neo by *Bgl*II and *Bam*HI digestion. The plasmid pBSLIF9 was linearized from the middle of the third exon by *Aat*II digestion and blunted with T4 DNA polymerase. The blunted Neo fragment was then ligated with linearized pBSLIF9. Ampicillin-resistant colonies were transferred to filters and screened with ^{32}P-labelled Neo DNA fragment. The Neo-containing clones were picked and the orientation of insertion was checked by *Pst*I digestion. The junction sequence between the LIF third exon and the beginning of the Neo fragment was confirmed by sequencing. The test plasmid, PGKLIF3Neo, was constructed by ligating the 3.8 kb LIF–Neo fragment into the PGKpUC18 expression vector.

To obtain the pBSLIF11Neo vector, the *Sac*II-cut pBSLIF9Neo was ligated to a *Sac*II fragment cut from pRSLIF11 which contains an additional ~ 2kb fragment of the 5′ flanking region of LIF gene. In construct pBSLIF7Neo, the TATA boxes and the D-form-specific first exon were deleted by digesting with *Sac*II and *Eco*RI. The GC-rich region in the first intron was deleted in pBSLIF6.5Neo by double digestion with *Cel*II and *Eco*RI. All plasmids were linearized by digestion with *Eco*RI before electroporation.

Cell culture and transfection

C3H10T1/2 cells were grown in Dulbecco's modified Eagle's medium (DMEM) F12 media supplemented with 10% fetal calf serum (FCS), 0.12% $NaHCO_3$ and 2 mM glutamine (DMEM:F12/FCS). Cell culture reagents were obtained from ICN Flow (UK), and FCS was obtained from Porton Biologicals (UK). Cell culture plastic and disposables were obtained from Costar (UK). For transient-transfection experiments, 5×10^6 cells were electroporated with 20 μg of plasmid using a Bio-Rad Gene Pulser (0.4 kV and 250 μF). Human growth factor expression vector (PGKhGH; 2 μg) (Bornstein and McKay, 1988) was co-transfected as an electroporation control. The transfected cells were harvested 48 h after electroporation and cytoplasmic RNA was isolated for RNAase-protection analysis.

For stable-transfection experiments, 5×10^6 cells were electroporated with 20 μg of linearized construct and plated into ten 10 cm-diameter tissue culture dishes. The culture medium was replaced with DMEM:F12/FCS medium containing G418 (100 μg/ml; Sigma) 24 h after electroporation, and replenished every 2 days thereafter. Macroscopically visible G418-resistant clones were picked and expanded for further analysis.

RNA preparation and RNAase-protection analysis

The procedures for preparation of cytoplasmic RNA and RNAase protection were as described by Rathjen et al. (1990a).

The plasmids pBSmLIFD and pBSmLIFM described by Rathjen et al. (1990a) were used to prepare riboprobes. Plasmid pBSmLIFD was linearized by *Hin*dIII followed by *in vitro* transcription with T7 RNA polymerase. *Eco*RI-linearized

pBSmLIFM was transcribed *in vitro* with T3 RNA polymerase (666 nucleotides).

A *Stu*I–ApaI fragment which contains part of the second intron and the entire second exon was ligated to *Sma*I–*Apa*I-digested pBluescriptIIKS. Antisense riboprobe of this construct was generated by T3 RNA polymerase after *Bam*HI linearization.

RESULTS

DNA sequence of the LIF gene

A genomic clone encoding the complete LIF-gene-coding sequence was obtained by screening a C57Bl6/CBA genomic DNA library. An 11 kb *Eco*RI fragment spanning the complete LIF gene was isolated and the DNA sequence of the LIF-gene-coding sequences, introns and 5′ and 3′ coding sequence was determined. This sequence differs in many respects in regions of overlap from the preliminary sequence of the murine LIF gene reported by Stahl et al. (1990). This includes the identification of unassigned nucleotides, as well as single base-pair differences, insertions and deletions. The nucleotide sequence of LIF protein-coding sequences (and therefore the deduced LIF amino acid sequence) were, however, identical. The differences in regions of non-coding sequence may in part arise from polymorphic differences in the LIF gene between different strains of mouse.

Inspection of the LIF gene DNA sequence (Figure 1) permitted identification of the location of the exons encoding the D-LIF exon and the M-LIF exon. The M-LIF exon is located 463 nucleotides 3′ of the D-LIF exon. The region between the D-LIF exon and the M-LIF exon is characterized by a high concentration of GC nucleotides, significant regions of sequence conservation between the mouse and human LIF genes and a high density of the recognition sites for the methylation-sensitive restriction enzyme *Hpa*II (Stahl et al., 1990; Kaspar et al., 1993; Figure 1). The region between the D-LIF exon and the M-LIF exon also contains a 41 bp polypyrimidine tract. Polypyrimidine tracts located with intronic sequences have been implicated in the control of exon selection in the alternative splicing of a number of genes (Smith and Nadal-Ginard, 1989). Further features of the LIF gene sequence will be discussed below.

Transient-transfection studies

Initial studies of LIF gene regulation *in vitro* focused on analysis of LIF gene transcription after transient transfection, into C3H10T1/2 fibroblast cells or PYS-2 yolk sac carcinoma cells, of constructs containing the entire LIF gene, a heterologous 3′ region containing a polyadenylation sequence and various amounts of 5′-non-coding sequence (see the Materials and methods section). A PGKhGH reporter plasmid was co-transfected with experimental constructs as a transfection control. Analysis of LIF gene expression by RNAase protection in transiently transfected cells failed to reveal detectable levels of expression derived from the transfected LIF-gene-derived constructs. However, both the hGH-reporter-gene-derived transcripts and endogenous LIF transcripts could be detected in these experiments (results not shown). These findings suggest that the transfected LIF-gene-derived DNAs either contained insufficient non-coding sequences to direct detectable levels of expression or contained sequences that suppressed expression in a transient-transfection setting. These findings contrast with a report (Gough et al., 1992) in which transient expression of a chloramphenicol acetyltransferase reporter sequence directed by 5′-non-coding sequences of the LIF gene was observed. An important difference between the two experiments was that the constructs employed in the present study, unlike those of Gough et al. (1992), all included non-coding intron sequences of the LIF gene. This raised the possibility that regions within the LIF gene itself might influence expression in transfected cells.

Expression of LIF transgenes in stable cell lines

In the light of the above results, a series of LIF gene constructs was created designed to examine the expression of the LIF gene in stably transformed cell lines. These constructs not only contained members with additional 5′-non-coding sequences to those used for transient-transfection studies but also included 3′-non-coding sequences of the LIF gene including putative polyadenylation sequences (Figure 2). An additional feature of these constructs was that they contained a promoterless Neo sequence, preceded by a translational stop codon inserted into a unique *AatII* site of the third exon.

The purpose of this design was two-fold: first it permitted isolation of stably transfected cells in which the transfected LIF-gene-derived constructs were transcribed by virtue of acquisition of resistance to G418. Secondly, it permitted simultaneous analysis by both endogenous and transgene-derived LIF gene expression by RNAase protection using probes that spanned the site of insertion of the Neo coding sequence. An additional construct was prepared to test the efficiency of acquisition of G418-resistance using the hybrid LIF–Neo gene. In this case the 5′-non-coding sequences of the LIF gene were replaced by the murine PGK promoter. A control plasmid was also employed in which the murine PGK promoter (Adra et al., 1987) was directly fused to Neo sequences (PGKNeo).

The experimental constructs were transfected into C3H10T1/2 fibroblast cells and subjected to selection for acquisition of neomycin resistance by growth in G418. The results of a typical experiment (from three performed) are shown in Figure 2. These clearly reveal that the number of G418-resistant colonies recovered depended on the identity of the construct transfected. In particular, relatively few colonies were recovered from cells transfected with constructs containing 5′-non-coding sequences derived from the LIF gene but a tenfold increase in the number of colonies obtained was observed in the case of the LIF6.5Neo construct in which all 5′-non-coding sequences, the D-form exon and most of the intron between the D-LIF and M-LIF exons (including the 'GC-rich' region) were deleted. In addition, replacement of the 5′-non-coding sequence of the LIF gene with the potent murine PGK promoter did not give rise to an appreciable increase in the frequency of G418-resistant clones recovered. Finally, in all cases the number of colonies recovered was substantially lower than those derived by transfection of cell with the PGKNeo control plasmid. This suggests that the hybrid LIF–Neo transcripts were less efficient than direct expression of Neo in conferring resistance to G418. The relative inefficiency of bicistronic transcripts in conferring drug resistance has been noticed in other studies (Sedivy and Sharp, 1989).

Recovery of a G418-resistant clone requires that the incoming DNA construct integrates into a genomic locus which permits expression of the transgene at levels sufficient to confer neomycin resistance. It may therefore be concluded from these observations that DNA constructs containing regions of the LIF gene and 5′-non-coding sequences are significantly more sensitive to features of the chromosomal environment for expression than those containing only the PGK promoter. Moreover a comparison of the frequency of G418-resistant clones recovered in constructs containing the complete LIF gene linked to the PGK promoter and constructs in which parts of the LIF gene coding for the D-form exon and part of intron 1 have been deleted indicates that the inhibitory effect is dependent on a region of the LIF gene

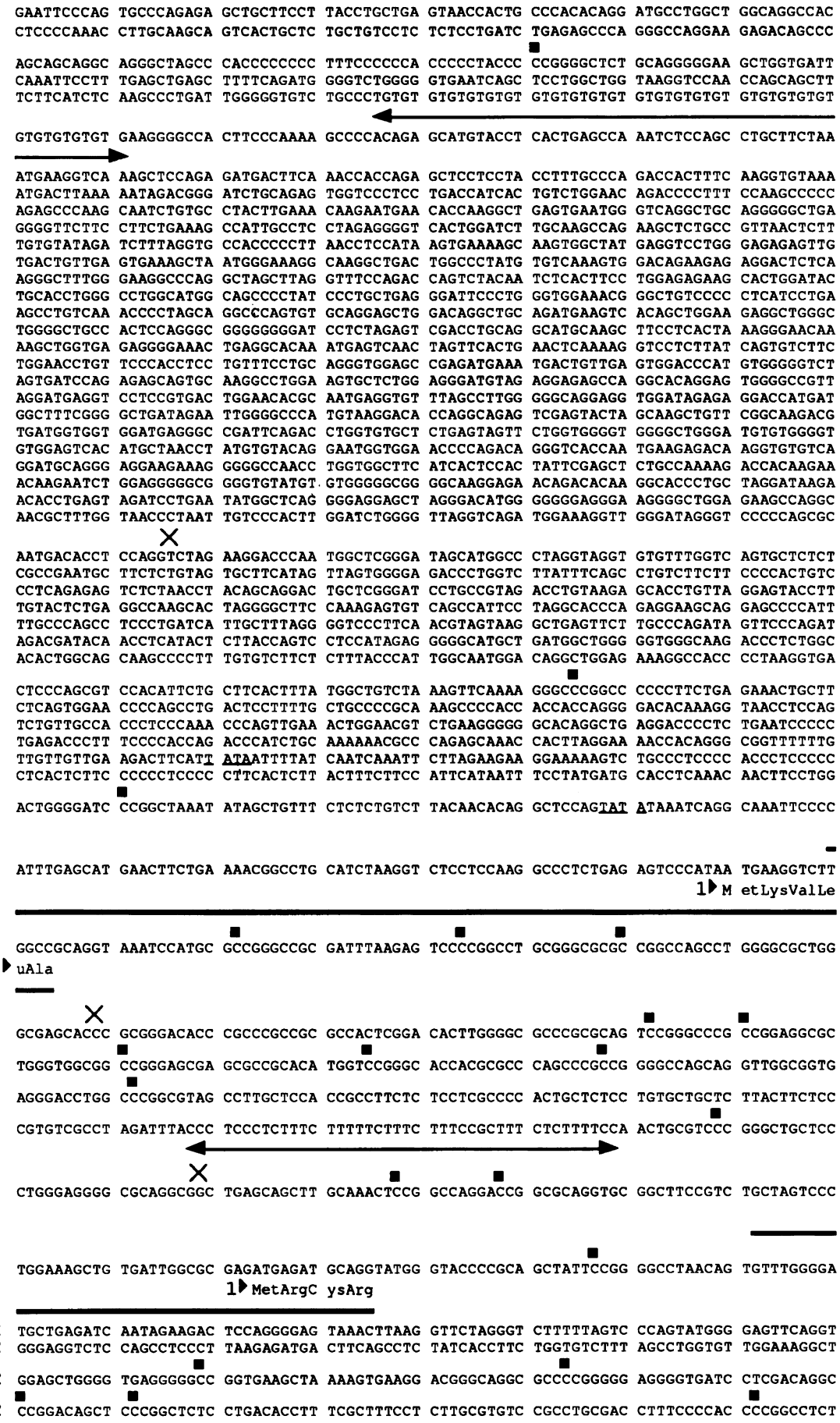

Figure 1 For legend see opposite

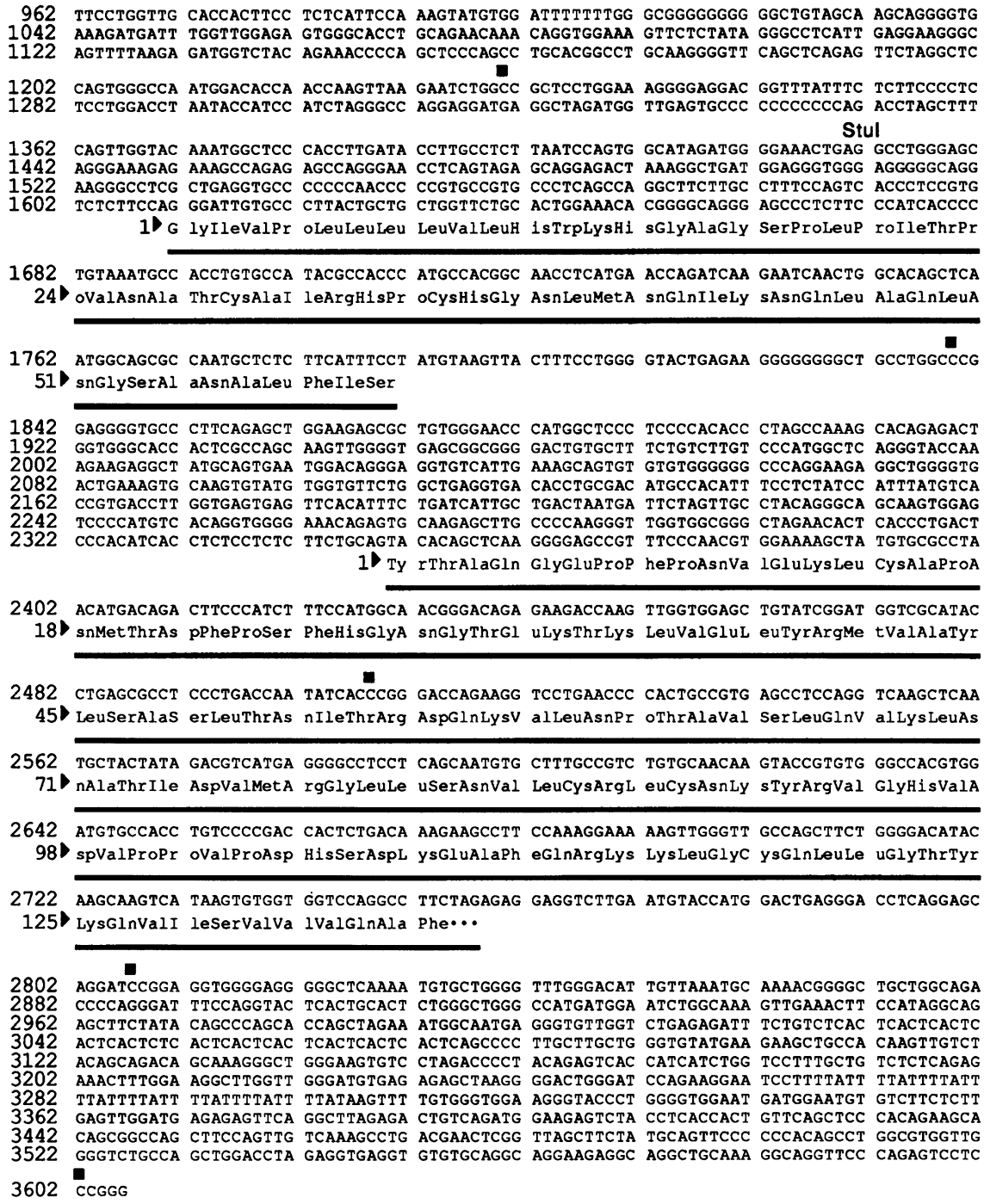

Figure 1 DNA sequence of the LIF gene

Exons are underlined and encoded protein sequence shown. X marks restriction sites in the 5′ end of the gene employed in the generation of LIF gene constructs for analysis of gene transcription (Figure 2). ■ indicates the localization of restriction sites for the methylation-sensitive enzyme *Hpa*II. The 'polypyrimidine tract' in intron 1 and the alternating purine/pyrimidine tract in the 5′-non-transcribed region of the gene are designated by double arrowed lines.

located in intron 1. The function of this region, defined by these experiments, is to confine expression of the LIF gene to a restricted number of genomic loci. This regulatory function can, in addition, be conferred upon the heterologous PGK promoter. These observations therefore define a significant regulatory element of the LIF gene located in the GC-rich region between the D-exon and the M-exon in the first intron.

RNAase-protection analysis of the RNA isolated from G418-resistant clones

Cytoplasmic RNA was prepared from the clones isolated in the experiments described above and further analysed by RNAase-protection techniques using probes derived from either LIF cDNAs or regions of the LIF gene. The purpose of these

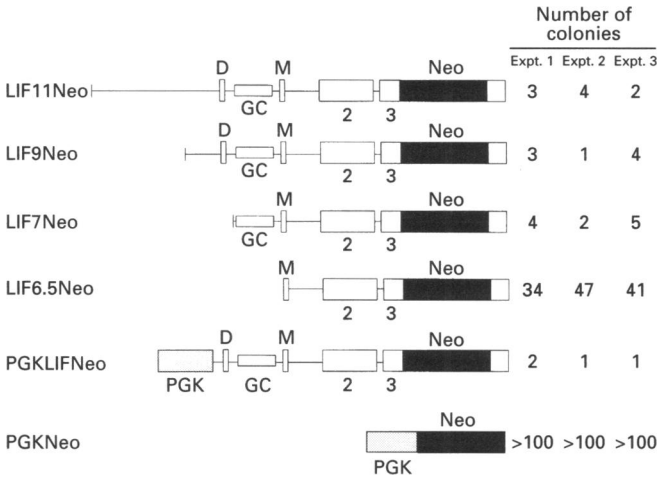

Figure 2 Design of experimental constructs for stable-transfection experiments and the number of stable G418-resistant colonies recovered for each construct (data from one experiment)

D, Location of the D-LIF exon; M, location of M-LIF exon; GC, location of the GC-rich region in the first intron. Shaded region indicates the neomycin-resistance coding region. The results of three independent experiments are shown. Clones isolated in experiment 1 were expanded for further analysis.

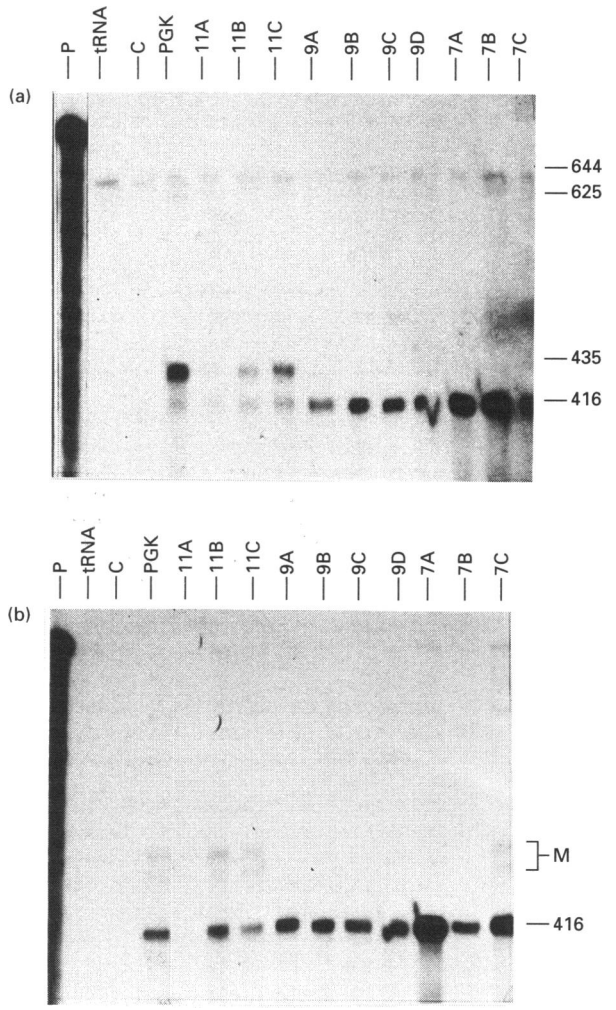

Figure 3 Analysis of LIF transcripts in stably transfected clones using (a) a D-LIF cDNA probe and (b) an M-LIF probe

(**a**) Endogenous D-LIF transcripts yield a protected band of 644 nucleotides and endogenous transcripts lacking the D exon yield a band of 625 nucleotides. Transgene-derived transcripts yield protected species of 435 nucleotides for RNAs containing the D exon and 416 for transcripts lacking the D exon. (**b**) Endogenous transcripts are not visible at this autoradiographic exposure. Transgene-derived transcripts yield protected species of 435 and 439 nucleotides for RNAs containing the M exon and 416 for transcripts lacking the M exon. 11A, 11B and 11C are individual clones from transfectants harbouring LIF11Neo; 9A, 9B and 9C are individual clones from transfectants harbouring LIF9Neo; 7A, 7B and 7C are individual clones from transfectants harbouring LIF7Neo. P, RNA probe; tRNA, tRNA control; C, non-transfected control cells.

experiments was to examine the nature of the transcripts expressed in these clones.

In the first instance, an antisense strand riboprobe derived from the cDNA fragment which contains the first exon specific for the D form but not the M form (Rathjen et al., 1990a) was used. Protection of endogenous D-form transcripts will give rise to a 644-nucleotide species (corresponding to the length of LIF-specific sequence in the probe). Endogenous M-form transcripts will only be protected in the region of the second and third exons giving rise to a 625-nucleotide species. The protected species corresponding to transcripts derived from exogenous DNA constructs will be shorter (439 and 420 nucleotides respectively) than the endogenous LIF transcripts because of introduction of the Neo sequence.

The results of these experiments are shown in Figure 3(a). In all cases the expected protected species derived from endogenous C3H10T1/2 LIF transcripts were detected. In samples derived from cells transfected with experimental DNA constructs, additional protected species were observed arising from expression of the LIF transgenes. In the case of clones derived from cells transfected with LIF11Neo, which contains the entire LIF gene and 3.2 kb of 5′-non-coding sequence, two protected species were observed of 439 and 420 nucleotides. Identical species were also observed in clones derived from cells transfected with the PGKLIFNeo construct. These species correspond to transcripts that either contain (439) or lack (420) the D-form exon. In all other cases (i.e. LIF9Neo, LIF7Neo and LIF6.5Neo), only the 420-nucleotide (D form-deficient) species was detected. These findings therefore define a second regulatory element in the LIF gene located between −1200 and −3120 nucleotides from the D-form exon. Sequences present in this region are required for initiation and expression of transcripts which contain the D-form exon. This regulatory element can be replaced by the heterologous murine PGK promoter.

Although the 420-nucleotide species was of a size predicted to arise from transcripts initiated from the M-form exon, it was necessary to confirm this by analysis of these RNAs using the M-LIF cDNA probe (Rathjen et al., 1990a) containing the M-form exon. In this instance, endogenous D-LIF transcripts will give rise to a protected species of 625 nucleotides and M-form transcripts a species of 666 nucleotides. In the case of the transfected constructs, transcripts initiated at the M-LIF exon will give rise to a species of 457 nucleotides and transcripts lacking the M-LIF exon will give rise to a species of 416 nucleotides. Analysis of RNAs from the transfected clones is shown in Figure 3(b). In all cases endogenous LIF transcripts of the predicted size are detected as well as the 416-nucleotide species, corresponding to transcripts lacking the M-LIF exon, derived from the transfected DNAs. The 457-nucleotide protected species, containing the M-LIF exon is, however, only detected in transfectants containing either the 3.12 kb 5′-non-

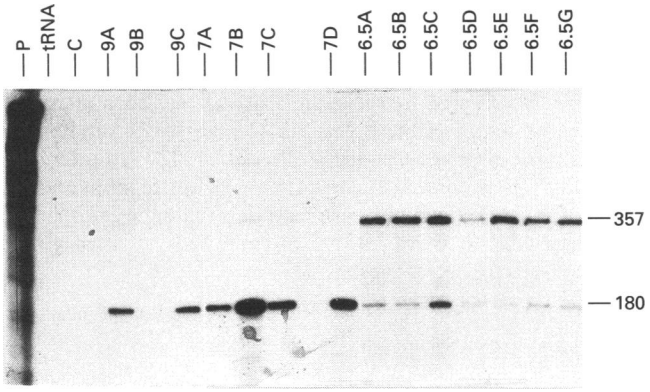

Figure 4 Analysis of LIF transcripts in stably transfected clones using a *Stu*I–*Apa*I LIF genomic fragment containing the second exon and part of the intron between the second exon and the M-LIF exon

Transcripts containing both intron and exon 2 sequences yield a protected species of 357 nucleotides. Transcripts second-exon sequences alone yield a species of 180 nucleotides. Abbreviations as in Figure 3.

coding sequence of the LIF gene (LIF11Neo) or the heterologous PGK promoter. These results therefore reveal that, as is the case for D-LIF transcripts, either sequences between −1200 and −3120 of the LIF gene or a heterologous promoter such as PGK are required for the expression of transcripts containing the M-LIF exon. The regulatory region defined by these experiments, lying between −3120 and −1200, is therefore required for expression of both M-LIF and D-LIF transcripts. Deletion of this element leads to the production of transcripts that contain both the second and third exons of the LIF gene (but neither the D-LIF nor M-LIF exons) and are presumably initiated elsewhere.

Cytoplasmic RNAs from transfected cell lines were further examined employing a probe derived from LIF genomic DNA which included the second exon and 167 bp of intron sequence lying between the M-form exon and exon 2 (Figure 4). In all cases a protected species of 180 nucleotides was observed corresponding to spliced transcripts containing the second exon. However, in the case of clones containing the shortest construct, LIF6.5Neo, a second protected species of 357 nucleotides, was also observed. This corresponds to transcripts that contain both exon 2 and 5′ intron sequences, i.e. transcripts that were initiated 5′ to the region defined by the probe and in which the region between the M-LIF exon and the second exon was not spliced out. Since G418-resistant clones were obtained with an approximately tenfold higher frequency in cells transfected with LIF6.5Neo compared with other constructs, a potential explanation of this result is that these transcripts arise from promoter- or gene-trapping events in which transcription is initiated from regions outside the LIF6.5Neo transgene and proceeds through the M-form exon and succeeding intron without splicing. An additional function of the GC-rich region may therefore be to either suppress or terminate transcription controlled by exogenous regulatory elements. The appearance of both spliced and unspliced transcripts may result from the deletion, in LIF6.5Neo, of the polypyrimidine tract located between +419 and +450.

DNA sequence of the distal 5′ flanking region of the murine LIF gene

The above results demonstrated that elements present in the region −3127 to −1200 are required for expression of both M-LIF- and D-LIF-containing transcripts of the type found by expression of the LIF gene in normal tissues. The total 3.3 kb 5′ flanking sequence is shown in Figure 1. This region contains a number of potentially significant regions of DNA sequence. First, there is a [$d(TG)_{28} \cdot d(CA)_{28}$] repeat sequence located between −2924 and −2969. Alternative purine/pyrimidine repeated sequences have been found in the 5′ flanking region of many other genes (Naylor and Clark, 1990; Gubbay et al., 1992) and may play an important role in regulation of gene expression. The $d(TG)_n \cdot d(CA)_n$ has the potential to adopt left-handed Z-DNA and change the chromatin structure *in vitro* and *in vivo* (Hanniford and Pulleybank, 1983).

Second, the LIF distal sequence shares some similarity in the promoter region with the rabbit uteroglobin gene which is also expressed under hormonal control in the uterine endometrium. Jantzen et al. (1987) identified a series of progesterone-dependent DNAase I-hypersensitive sites in the uteroglobin promoter. A particular feature of these sites was the presence of divergent, but related, octanucleotide boxes (HS −2.4), found in both orientations and overlapping binding sites for the progesterone receptor. It was argued that these boxes might represent binding sites for endometrium-specific proteins. The 5′ region of the LIF gene contains three copies of the HS −2.4 sequence. Minimal recognition sequences for progesterone/glucocorticoid receptor binding (Schreidereit et al., 1983; Jantzen et al., 1987; Beato et al., 1989) can also be identified at −3167 to ~ −3160 and −1382 to ~ −1376 in the distal 5′ flanking region of mouse LIF gene. At least in the case of the human metallothionein-IIA promoter these sequences are insufficient, by themselves, to confer steroid responsiveness and their biological function in the context of the LIF gene remains to be elucidated.

DISCUSSION

The studies reported here have led to the functional definition of two important regulatory regions in the LIF gene. The first region is located between the D-form axon and the M-form exon and is characterized by a high-GC content and regions of sequencing similarity to the human LIF gene. In addition, Kaspar et al. (1993) have demonstrated that this region of the LIF gene is hypomethylated in many types of cells and associates with specific nuclear proteins. The authors have argued that the GC-rich region has the properties expected of a CpG 'island' found in the promoter region of many genes. The function of this region, defined by the experiments reported here, is complex. Firstly inclusion of this region in experimental constructs leads to a significant reduction in the ability to recover stable clones expressing the transgene as well as suppression of gene expression in transient-transfection experiments. The region therefore confers an important requirement for specific genomic locations to permit expression of the LIF gene. The identification of hypomethylated regions in the endogenous LIF gene in this region and binding sites for nuclear proteins (Kaspar et al., 1993) suggest that the action of these proteins may be to repress transcription. It is of interest that the nuclear protein factor MeCP-1 has been demonstrated to bind to hypomethylated regions in the GC islands and act to repress transcription (Boyes and Bird, 1991, 1992). The dependence on genomic environment for expression observed with constructs containing this region may accordingly reflect some feature of the methylation status of the integrated transgene.

The experiments reported here also demonstrate that it is possible to direct expression of the LIF gene in a construct that is entirely devoid of sequences 5′ to the GC-rich region, including

TATA and CAAT elements. The region between the D-exon and the second exon therefore also has the properties of a promoter region of the TATA-less housekeeping type. Mapping transcripts directed by this element has revealed that they lack both the D-exon and the M-exon. This suggests that there is a cryptic transcriptional start site within this region of the LIF gene whose action becomes evident on deletion of the 5′ enhancer. It is of interest that Escary et al. (1993), in their study of LIF function by gene-targeting techniques, introduced a β-galactosidase reporter gene into the LIF gene by deletion of part of the GC-rich region. The reporter gene, unlike the endogenous wild-type LIF gene, was not expressed in the resulting transgenic mice, suggesting that the GC-rich region plays an essential role in the expression of the normal LIF gene.

A final feature of the GC-rich region is that, on deletion, not only does the number of 'successful' (defined by active transcription) integration events rise, but that a significant proportion of the transcripts in these clones fail to splice out the intron between the M-LIF exon and the second exon. This suggests that the GC-rich element has some influence over the pattern of splicing of the LIF gene. It may be significant therefore that the GC-rich region contains a polypyrimidine tract of the type implicated in alternative splicing mechanisms in other genes (Smith and Nadal-Ginard, 1989). An interesting possibility is that these unspliced transcripts are due to gene-trapping events in which the experimental transgene has integrated into the coding sequence of a transcriptionally active gene. It follows that it would be of interest to examine the effect of the GC-rich region on the frequency and identity of enhancer- and gene-trapping events in a heterologous construct.

The second element in the LIF gene is located in the 5′ region of the gene between −3120 and −1200. This element can be defined as a region required for expression of the LIF transcripts initiated at both the D exon and the M exon. The action of this element can be emulated by the heterologous PGK promoter and so it, in some respects, can be considered to resemble a classical enhancer element in its function.

The sequence of this region (−3200 to −1200) of the LIF gene exhibits a number of interesting features. Firstly it contains a poly $d(TG)_n \cdot d(CA)_n$ tract, which is thought to adopt a left-handed Z-DNA structure. Such tracts have been identified in a number of different mammalian genes and have been argued to play a role in the control of gene expression (Jaworski et al., 1987). Of particular interest is the finding that the rat prolactin gene contains two such regions whose action appears to inhibit expression (Naylor and Clark, 1990); this repressive effect can, however, be inhibited in the presence of progesterone.

A second feature of this region of the LIF gene is that it contains a number of sites that exhibit sequence similarity to steroid regulatory elements of the uteroglobin gene. The uteroglobin gene is expressed in the uterus during the preimplantation phase in which its expression is, in a similar fashion to LIF, controlled by maternal steroid hormones (Jantzen et al., 1987; Wolf et al., 1992). In the endometrium, both progesterone and oestradiol can enhance the transcription of the uteroglobin gene (Muller and Beato, 1980). Minimal consensus steroid-hormone recognition sequences can also be found in the distal enhancer region of murine LIF gene, although their function as *bona fide* steroid-response elements remains to be demonstrated. Furthermore there are three copies of a sequence motif identified in the rabbit uteroglobin gene in regions of progesterone-dependent DNAase I hypersensitivity. This suggests that the mechanisms of LIF and uteroglobin gene regulation may share some common features with respect to the action of steroid hormones. The biological function of uteroglobin is thought to be binding and sequestration of progesterone (Wolf et al., 1992). An interesting possibility is therefore that co-regulation of uteroglobin and LIF by maternal steroids might partly explain the depression in LIF expression in the uterus after implantation by uteroglobin-mediated removal of progesterone.

This work was supported by the Cancer Research Campaign (U.K.). L.-W.H. was supported by a postgraduate studentship from the government of Taiwan (R.O.C.).

REFERENCES

Adra, C. N., Boer, P. H. and McBurney, M. W. (1987) Gene **60**, 65–74
Alexander, H. R., Wong, G. G., Doherty, G. M., Venzon, D. J., Fraker, D. L. and Norton, J. A. (1992) J. Exp. Med. **175**, 1139–1142
Beato, M., Chalepakis, G., Schauer, M. and Slater, E. P. (1989) J. Steroid Biochem. **32**, 737–747
Bhatt, H., Brunet, L. J. and Stewart, C. L. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 11408–11412
Bornstein, P. and McKay, J. (1988) J. Biol. Chem. **263**, 1603–1606
Boyes, J. and Bird, A. (1991) Cell **64**, 1123–1134
Boyes, J. and Bird, A. (1992) EMBO J. **11**, 327–333
Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H. and Hieter, P. (1992) Gene **110**, 119–122
Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. **12**, 387–395
Escary, J. L., Perreau, J., Dumenil, D., Ezine, S. and Brulet, P. (1993) Nature (London) **363**, 361–364
Gough, N. M., Wilson, T. A., Stahl, J. and Brown, M. A. (1992) Ciba Found. Symp. **167**, 24–38
Gubbay, J., Vivian, N., Economou, A., Jackson, D., Goodfellow, P. and Lovell Badge, R. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 7953–7957
Hanniford, D. B. and Pulleybank, D. E. (1983) Nature (London) **302**, 632–634
Hilton, D. J. (1992) Trends. Biochem. Sci. **17**, 72–76
Jantzen, K., Fritton, H. P., Igo Kemenes, T., Espel, E., Janich, S., Cato, A. C., Mugele, K. and Beato, M. (1987) Nucleic Acids Res. **15**, 4535–4552
Jaworski, A., Hsieh, W., Blaho, J. A., Larson, J. E. and Wells, R. D. (1987) Science **238**, 773–777
Kaspar, P., Dvorak, M. and Bartunek, P. (1993) FEBS Lett. **319**, 159–162
Mereau, A., Grey, L., Piquet-Pellorce, C. and Heath, J. K. (1993) J. Cell Biol. **122**, 713–719
Metcalf, D. and Gearing, D. P. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 5948–5952
Metcalf, D., Hilton, D. and Nicola, N. A. (1991) Blood **77**, 2150–2153
Metcalf, D., Warning, P. and Nicola, N. A. (1992) Ciba Found. Symp. **167**, 174–182
Mizushima, S. and Nagata, S. (1990) Nucleic Acids Res. **18**, 5322
Muller, H. and Beato, M. (1980) Eur. J. Biochem. **112**, 235–241
Naylor, L. H. and Clark, E. M. (1990) Nucleic Acids Res. **18**, 1595–1601
Rathjen, P. D., Toth, S., Willis, A., Heath, J. K. and Smith, A. G. (1990a) Cell **62**, 1105–1114
Rathjen, P. D., Nichols, J., Toth, S., Edwards, D. R., Heath, J. K. and Smith, A. G. (1990b) Genes Dev. **4**, 2308–2318
Sanger, F., Nicklen, S. and Coulsen, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
Schreidereit, C., Geisse, S., Westphal, H. M. and Beato, M. (1983). Nature (London) **304**, 749–752
Sedivy, J. M. and Sharp, P. A. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 227–231
Shen, M. M. and Leder, P. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 8240–8244
Smith, C. W. J. and Nadal-Ginard, B. (1989) Cell **56**, 749–758
Stahl, J., Gearing, D. P., Willson, T. A., Brown, M. A., King, J. A. and Gough, N. M. (1990) J. Biol. Chem. **265**, 8833–8841
Stewart, C. L., Kaspar, P., Brunet, L. J., Bhatt, H., Gadi, I., Kontgen, F. and Abbondanzo, S. J. (1992) Nature (London) **359**, 76–79
Wolf, M., Klug, J., Hackenberg, R., Gessler, M., Grzeschik, K. H., Beato, M. and Suske, G. (1992) Hum. Mol. Genet. **1**, 371–378

Received 6 January 1994/9 March 1994; accepted 15 March 1994