The EMBO Journal vol.6 no.6 pp.1643-1651, 1987

Multiple c-*myb* transcript cap sites are variously utilized in cells of mouse haemopoietic origin

R.J.Watson, P.J.Dyson[1] and J.McMahon

Imperial Cancer Research Fund Laboratories, Dominion House, St Bartholomew's Hospital, Bartholomew Close, London EC1A 7BE, and [1]Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK

Communicated by N.M.Wilkie

Mouse c-*myb* gene transcripts in various cells of haemopoietic origin were analysed using S1 nuclease and RNase mapping techniques and by Northern blotting. It was found that the prevalent 3.8-kb c-*myb* mRNA present in thymocytes, T cell leukaemias, myelomonocytic leukaemias, erythroleukaemias and myeloid stem cells was initiated at several cap sites mapping within a region 97-244 bp upstream from the protein coding sequence. Utilization of additional cap sites mapping further upstream was also observed in certain cells, most notably thymocytes, and this gave rise to RNA species (4.3-5.6 kb) larger than the presumptive mRNA. In contrast, myeloma cell c-*myb* transcripts, which are much less abundant than those in more immature haemopoietic cells, were found to be initiated at a restricted set of cap sites mapping 244-277 bp upstream of the coding sequence. Hence, these data suggest that the abundance of the c-*myb* mRNA may be regulated by a process involving selective utilization of mRNA cap sites. Sites hypersensitive to DNase I were associated with mRNA cap sites in cells that expressed c-*myb*.

Key words: c-*myb*/haemopoietic cells/5' mRNA heterogeneity/nuclease hypersensitivity

Introduction

The c-*myb* gene has been associated with the induction of myeloid leukaemias, either following transduction by the avian retroviruses AMV and E26 (Bishop and Varmus, 1985) or as a result of insertional mutagenesis by murine retroviruses (Lavu and Reddy, 1986; Shen-Ong *et al.*, 1986). These induction events resulted in expression of novel gene products that were truncated analogues of the 75 000 mol. wt polypeptide specified by c-*myb* (Klempnauer *et al.*, 1983; Boyle *et al.*, 1986). Transcription of c-*myb* has been observed primarily in immature haemopoietic cells of all lineages (Gonda *et al.*, 1982; Westin *et al.*, 1982), and the high level of c-*myb* mRNAs in these cells was invariant through the cell cycle (Thompson *et al.*, 1986). Treatment of such cells with agents that induce terminal differentiation (e.g. to granulocytes or monocytes) resulted in a marked decrease in the levels of c-*myb* mRNAs with kinetics that preceded the suppression of cell proliferation in these cultures (Craig and Bloch, 1984; Gonda and Metcalf, 1984). In contrast, stimulation of mature resting human T cells resulted in a marked elevation of c-*myb* mRNAs as cells progressed through the G1 phase of the cell cycle (Torelli *et al.*, 1985; Reed *et al.*, 1986; Stern and Smith, 1986). These various observations suggested that the c-*myb* gene product may play central roles in differentiation and proliferation of haemopoietic cells, perhaps maintenance of elevated levels of this protein being inimical to the cessation of DNA replication that appears to be co-ordinate with the establishment of end-cell differentiation. We have, therefore, been interested in determining how c-*myb* expression may be regulated at the transcriptional level in response to differentiation stimuli.

A number of groups have reported the cloning of murine c-*myb* cDNAs and genomic DNA fragments (Castle and Sheiness, 1985; Gonda *et al.*, 1985; Bender and Kuehl, 1986). The major c-*myb* transcript, variously reported to be between 3.8 kb and 4.0 kb (Gonda and Metcalf, 1984; Sheiness and Gardinier, 1984; Bender and Kuehl, 1986), is encoded by >10 exons (Lavu and Reddy, 1986) spread over perhaps 50 kb of the haploid mouse genome. Seven of these exons (E1-E7) are homologous to the v-*myb* sequence transduced by AMV and E26. Three exons (UE1-UE3) are located upstream of E1, and an as yet undefined number of exons are located downstream of E7. As it is likely that sequences immediately 5' of UE3 will be of major relevance in the regulation of transcription of c-*myb* we set out to map precisely this 5' exon. We report here that c-*myb* transcripts displayed considerable 5' heterogeneity and that, at least quantitatively, the extent of this heterogeneity is cell-type specific.

Results

Isolation of cDNA and genomic DNA fragments representing the c-myb 5' terminus

Polyadenylated RNA from mouse thymocytes was used to prepare a cDNA library in the plasmid vector pUC9. This library was screened with a ^{32}P-labelled v-*myb* sequence obtained from AMV. One plasmid (designated pMCR1), was obtained initially, the 1420-bp insert of which (Figure 1B) contains the presumptive start of the coding sequence and 148 bp of the 5' non-coding sequence. This plasmid was then used to re-screen the cDNA library, and a number of additional positive clones were obtained. Among these was the plasmid pMCR8, which contains a 1735-bp insert including 724 bp of 5' non-coding sequence.

The v-*myb* probe was also used to isolate a number of recombinant bacteriophages containing overlapping mouse c-*myb* DNA sequences. These included isolates KR4A, KR1A and SR5A, which together contain all exons homologous to v-*myb* (E1-E7) as well as the three upstream exons and additional 5' non-transcribed sequences.

Restriction enzyme and sequence analysis of both cDNA and genomic DNA clones enabled identification of the exon/intron structure of this part of the c-*myb* gene (Figure 1A), and this is identical to that recently published (Lavu and Reddy, 1986). It was found that the 5' non-translated sequences of both pMCR1 and pMCR8 were entirely contiguous with genomic DNA, thus implying that UE3 extends at least 724 bp 5' of the protein coding sequence. To confirm this, and to provide more definitive data on the 5' boundary of UE3, nuclease analyses using both DNA and RNA probes were performed.

S1 nuclease analysis of the UE3 5' boundary

Two probes were used in the S1 nuclease analysis of UE3 (Figure

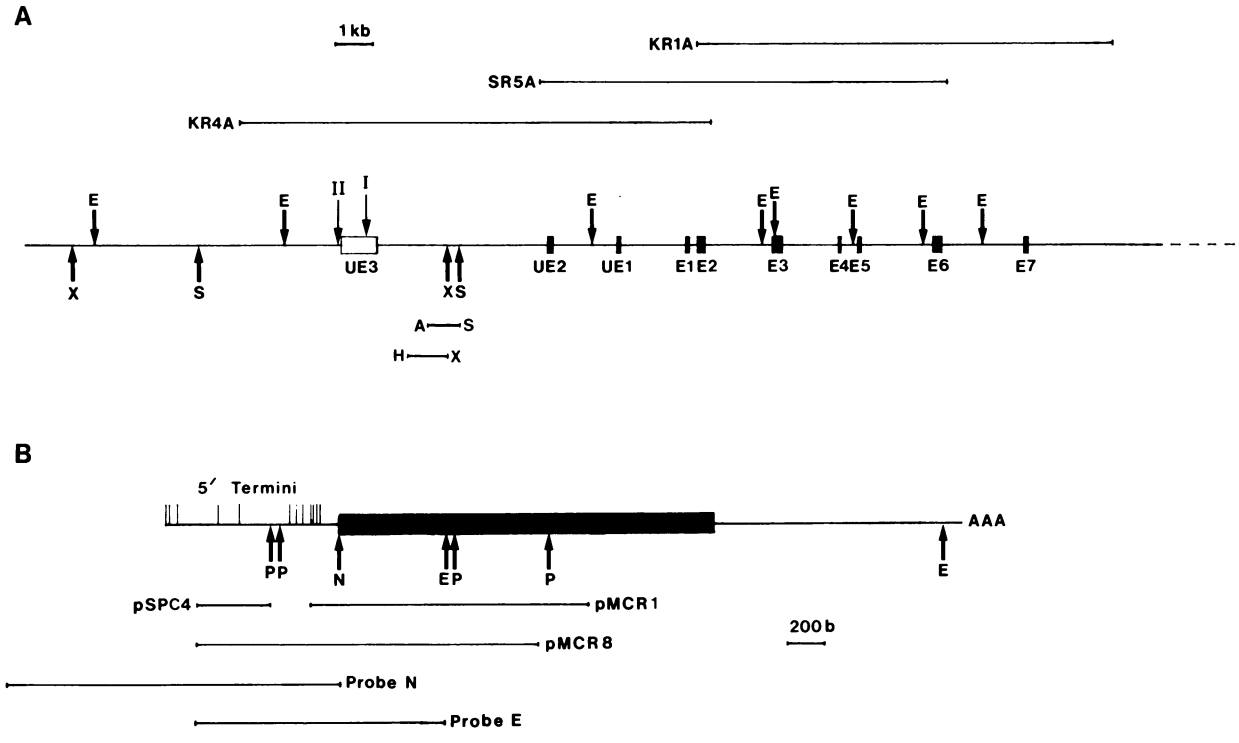

Fig. 1. Physical maps of the c-*myb* gene and cDNA. **(A)** *Eco*RI (E) restriction map of the 5′ region of c-*myb* showing the locations of the upstream exons (UE3−UE1) and those with homology to v-*myb* (E1−E7) based on published data (Castle and Sheiness, 1985; Lavu and Reddy, 1986). Coding sequences are identified by a filled box, non-coding sequences by an open box. Also shown are the *Sac*I (S) and *Xba*I (X) restriction sites used to identify nuclease hypersensitive sites I and II and (below) the *Apa*I (A)/*Sac*I and *Hin*dIII (H)/*Xba*I probes used for indirect end-labelling of the *Sac*I and *Xba*I fragments in this analysis. Above are shown the extents of the c-*myb* inserts contained by the recombinants KR4A, SR5A and KR1A. **(B)** *Eco*RI (E), *Pst*I (P) and *Nco*I (N) restriction map of c-*myb* cDNA based on data obtained here and that published previously (Bender and Kuehl, 1986). Indicated are the most predominant 5′ termini found with thymocyte RNA and the polyadenylation site (AAA). Below are indicated the cDNA inserts of plasmids pSPC4, pMCR1 and pMCR8 and the extent of probes N and E used in the nuclease analyses.

1B). One (probe E) consisted of a 1315-base pMCR8 cDNA sequence extending upstream from the *Eco*RI site (within the *myb* coding sequence) to the 5′-most non-translated sequence. The other (probe N) consisted of a 1.7-kb genomic DNA fragment extending upstream from the *Nco*I site that is co-ordinate with the start of the protein coding sequence. These probes, uniquely 5′-^{32}P-labelled at the *Eco*RI and *Nco*I termini, respectively, were hybridized with polyadenylated RNA isolated from thymuses of C57BL and BALB/c mice and from the myelomonocytic leukaemia cell line WEHI-3B. Control hybridization reactions contained an equivalent amount of yeast tRNA. After annealing, the samples were treated with nuclease S1 and the products were resolved on denaturing polyacrylamide gels. Autoradiography of these gels revealed a heterogeneous display of DNA fragments (Figure 2A and B) the sizes of which correspond to the distance from the labelled 5′ *Eco*RI or *Nco*I termini of the DNA probes to a point of discontinuity (i.e. the 5′ boundary of UE3) with the complementary c-*myb* transcript. For example, S1 protected-DNA fragments with lengths between 660 and 780 bases were obtained upon hybridization of probe E with WEHI-3B RNA (Figure 2A, lane 3).

Similarly, fragments ranging from 100 to 247 bases were obtained with WEHI-3B RNA using probe N (Figure 2B, lane 3). As the cDNA *Eco*RI site is 556 bp downstream from the *Nco*I site, it is apparent that the 5′ boundaries of UE3 revealed with these two probes are co-ordinate. In addition to the heterogeneous species obtained with WEHI-3B RNA, it was notable that a number of prevalent higher mol. wt fragments (ranging from 510 to 880 bases with probe N) were protected in this assay with both C57BL and BALB/c RNA samples (Figure 2B). Thus, additional heterogeneous c-*myb* transcripts initiating some way upstream from those common to WEHI-3B are apparently present in the thymocytes.

To ensure that the heterogeneity demonstrated by this analysis was not due to endolytic cleavage of RNA/DNA hybrids by nuclease S1 at susceptible sites the following control was done. The pMCR8 cDNA sequence was inserted into the pSP65 vector (yielding plasmid pSPC5) so that an RNA copy of this cDNA having the same polarity as c-*myb* mRNA could be made *in vitro* using SP6 RNA polymerase. Different concentrations of the *in vitro* pSPC5 c-*myb* RNA were then hybridized with probe N and analysed as above. In the absence of endolytic cleavage, a single fragment of 727 bases (representing pMCR8 sequences downstream to the *Nco*I site) would be expected to be protected against nuclease digestion, and this fragment was indeed clearly observable in this experiment (Figure 2C). Additional low mol. wt DNA fragments were observed only at high *in vitro* RNA concentration (i.e. 2−10 ng/hybridization) and, with the possible exception of a 210-base fragment, cannot account for the multiple bands observed with thymus and WEHI-3B RNAs. Thus, we conclude that there is considerable heterogeneity of the UE3 5′ boundary and that this heterogeneity is more extensive in thymocyte c-*myb* transcripts than in WEHI-3B RNA. From the sizes of the S1-protected fragments it is apparent that the long 5′ non-coding sequence of pMCR8 is not representative of WEHI-3B c-*myb* transcripts and, indeed, only a subset of thymocyte c-*myb* RNAs

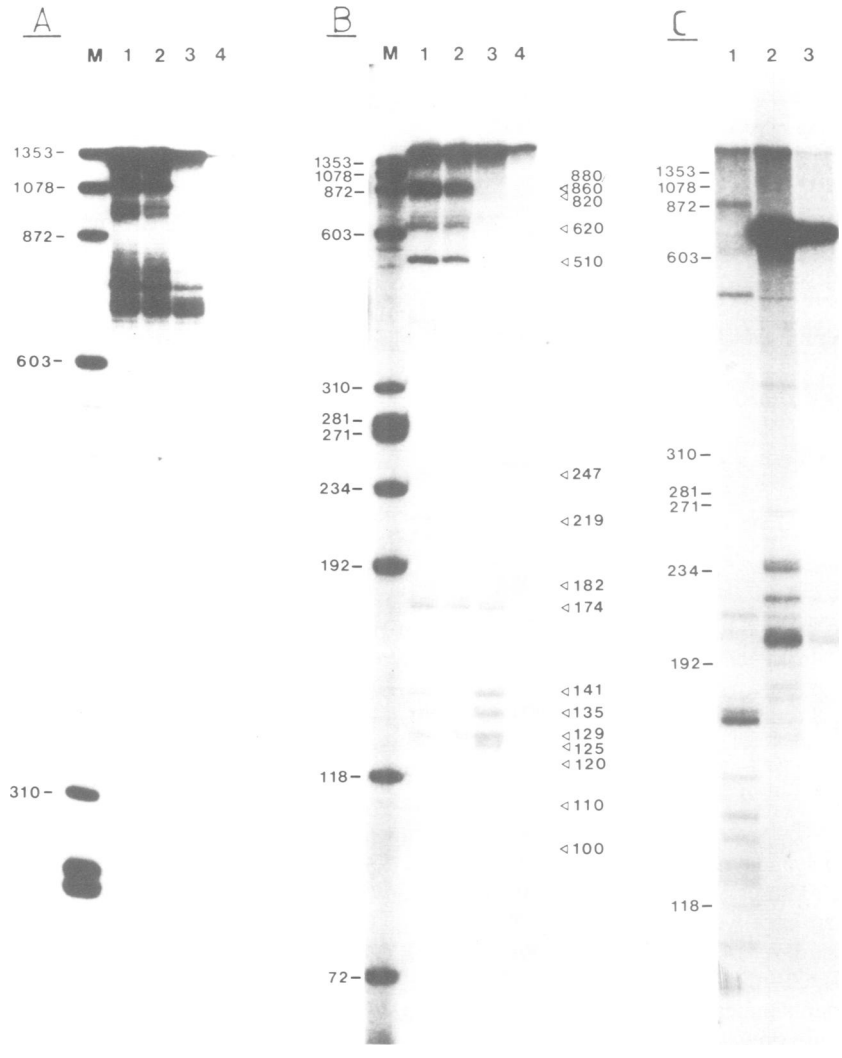

Fig. 2. S1 analyses of c-*myb* transcripts. **(A)** DNA/RNA hybrids formed between 5′-^{32}P-labelled probe E (Figure 1B) and 5 μg polyadenylated RNA isolated from **(1)** C57BL thymocytes, **(2)** BALB/c thymocytes or **(3)** WEHI-3B and **(4)** 10 μg yeast tRNA were treated with S1 nuclease and displayed on a 4.5% polyacrylamide/urea gel. **(B** DNA/RNA hybrids formed with 5′-^{32}P-labelled probe N were analysed as in (A). **(C)** S1 analysis of hybrids formed between 5′-^{32}P-labelled probe N (Figure 1B) and **(1)** polyadenylated R1 RNA and **(2), (3)** different amounts of *in vitro* pSPC5 c-*myb* transcripts (see text). Molecular weight markers (M) were the *Hae*III fragments of φX174 DNA.

contain such an extensive UE3 component.

Riboprobe analysis of c-myb RNA

To substantiate further the results obtained above with nuclease S1, alternative riboprobe assays were performed in which hybrids formed between c-*myb* RNA and *in vitro* labelled anti-sense RNA were delineated by RNase digestion and gel electrophoresis. Moreover, we examined RNAs isolated from a number of different mouse haemopoietic cells to determine whether the phenomenon of c-*myb* transcript 5′ heterogeneity is a feature common to cells of diverse lineages. The origins of the various haemopoietic cells so analysed are outlined in Table I.

The results of an experiment in which a riboprobe equivalent to probe N sequences was hybridized to polyadenylated RNAs isolated from C57BL thymus, WEHI-3B and from the T cell leukaemias AKR1 and R1 are shown in Figure 3A. As for the nuclease S1 analysis, a heterogeneous series of protected species between 100 and 250 bases was obtained with all these RNA samples. Higher mol. wt bands (510, 620 and 880 bases) were observed with thymus RNA, but only at low levels with WEHI-3B. The fragment pattern obtained with AKR1 RNA was similar to that given by WEHI-3B RNA, whereas R1 RNA yielded a pattern more similar to that of thymus RNA, although the proportion of higher mol. wt transcripts was lower. It should be noted that the mol. wt of these larger RNA fragments could only be determined accurately using RNA mol. wt markers on these gels as the mobility of such species tended to be retarded with respect to equivalent length DNA fragments. Control experiments in which a riboprobe was hybridized with *in vitro* pSPC5 c-*myb* RNA (Figure 3B), indicated that the multiplicity of protected fragments observed with cell RNAs could not be accounted for by endolytic cleavage, again with the possible exception of the 210-base fragment. Results obtained with this RNase mapping technique, then, were very similar to those obtained with nuclease S1, although the probes were in one instance uniformly labelled and in the other end-labelled. Hence, no evidence for splicing of UE3 sequences (other than at the normal donor site) was found, as such spliced RNAs should be detectable as additional bands in the RNase analysis.

We also analysed by nuclease mapping cytoplasmic RNA extracted from a number of additional cell lines and from mouse

Table I. Derivations of mouse haemopoietic cell lines

Cell line	Description	Reference
WEHI-3B (D^+)	Differentiation-positive mineral oil-induced BALB/c myelo-monocytic leukaemia	Metcalf and Nicola (1982)
AKR1	Spontaneous AKR/J T-cell lymphoma	Hyman *et al.* (1980)
R1	Spontaneous C58 T-cell lymphoma	Ralph (1973)
S49	Mineral oil-induced BALB/c T-cell lymphoma	Horibata and Harris (1980)
EL4	Chemically-induced C57BL T-cell lymphoma	Ralph (1973)
BW5147	Spontaneous AKR/J T-cell lymphoma	Hyman and Stallings (1974)
Ag-8	Mineral oil-induced BALB/c myeloma	Kohler and Milstein (1975)
MPC-11	IgG secreting myeloma	Laskov and Scharff (1970)
MOPC-31C	IgG secreting plasmacytoma	
M707T	Friend virus-induced erythro-leukaemia	Harrison *et al.* (1982)
FDCP-Mix1	Pluripotent stem cell derived from MoMuLV (*src*)-infected bone marrow culture	Spooncer *et al.* (1986))

spleen without selection for polyadenylated species (Figure 3C). Protected fragments (100−250 bases) indicating transcripts initiated at the downstream cap sites were evident in all T cell leukaemias (S49, EL4, R1 and BW5147), the erythroleukaemia M707T, and the stem cell FDCP-mix-1 as well as with spleen RNA. Higher mol. wt fragments indicating transcripts initiated at upstream cap sites were quite prevalent in the R1 and spleen RNA samples and to lesser extents in S49, EL4 and M707T RNA samples. Hence, the pattern of heterogeneity obtained here with cytoplasmic RNA is very similar to that found previously with total polyadenylated RNA (Figure 3A). Indeed, analyses of thymocytes and WEHI-3B cells fractionated into nuclear and cytoplasmic components yielded no evidence for selective accumulation in these cell sub-compartments of c-*myb* transcripts initiated at specific cap sites (data not shown). The most notable observation made from this analysis was that the only prevalent bands observed with myeloma (Ag-8, MPC-11 and MOPC-31C) RNA samples were of 280 bases and a doublet of 250/247 bases. These bands were clearly evident with other RNA samples, for example R1, in very similar quantity, but transcripts initiating at such cap sites constituted a very small proportion of total c-*myb* transcripts in cells other than these myelomas. Evidence that these predominant myeloma transcripts were contiguous with cloned cDNA and were not alternatively spliced species was obtained by hybridization of a truncated riboprobe complementary to nucleotides −724 to −115 to these RNAs: this resulted in protection of species of 160 and 130 bases (data not shown) which were truncated predictably with respect to those obtained with the riboprobe described above (extending from −724 through +4; Figure 3C). Very faint bands corresponding to transcripts initiated at other downstream cap sites could be detected also with myeloma RNA samples upon prolonged autoradiographic exposure of this gel (Figure 3C). Another observation of note was that no c-*myb* transcripts were detectable with NIH 3T3 fibroblast RNA samples (Figure 3C, lanes 7 and 8). Moreover, cycloheximide treatment of Ag-8 and NIH 3T3 cells did not result in an increase in c-*myb* mRNA levels (Figure 3C, lanes 8 and 10), an observation that is in contrast to that described for chicken bursal lymphocytes and embryo fibroblasts (Thompson *et al.*, 1986).

It may be concluded that in myeloma cells the steady-state levels of c-*myb* mRNAs are much lower than in other more immature haemopoietic cells and that this is correlated with greatly reduced usage of cap sites. Consistent with this conclusion, myeloma c-*myb* transcripts were detected at low level by Northern blot analysis (data not shown) and were of slightly higher mol. wt (3.9 kb) than the 3.8 kb mRNA observed in these other haemopoietic cells. No evidence was found for the presence of c-*myb* mRNA in NIH 3T3 fibroblasts.

Map locations of the c-myb transcript cap sites

The major 5′ termini determined by both nuclease techniques were mapped on the c-*myb* gene DNA sequence (Figure 4). The sequence shown is identical to that published by Bender and Kuehl (1986), with the exceptions of the addition of C nucleotides at positions −14 and −100. Termini common to thymus and WEHI-3B RNA (at nucleotides −97/−244) map downstream from a GC-rich sequence that contains three potential SP1 transcription factor binding sites (CCGCCC; positions −447/−442, −423/−418 and −419/414). These downstream 5′ termini map within or around a highly reiterated TC-rich sequence (nucleotides −201/−134). Neither these downstream termini, nor those found upstream with thymus and R1 RNAs (at nucleotides −875, −855, −815, −615 and −505) and the prevalent myeloma cap site (nucleotides −277 and −247/−244), appear to be associated with TATA boxes. Significantly, the termini mapped here do not correlate with canonical splice acceptor sequences, arguing strongly that the heterogeneity found with these nuclease analyses does not reflect a complex splicing pattern within UE3, but rather reflects the use of a number of alternative RNA cap sites.

RNA blot analyses of c-myb transcripts

To assess the significance of these alternative 5′ termini on the genesis of c-*myb* mRNA, an RNA blot hybridization analysis was performed. Here duplicate sets of polyadenylated thymus, WEHI-3B, AKR1 and R1 cell RNAs were resolved on a denaturing agarose gel and transferred to nylon membrane. These blots were hybridized either with a pMCR1 probe (consisting primarily of coding sequence) or with a nick-translated probe consisting entirely of upstream 5′ non-coding sequence (pSPC4, nucleotides −724 to −339). Hence, this latter probe should hybridize solely to those transcripts initiated upstream of −339 while pMCR1 should hybridize to all transcripts. As expected, pMCR1 was found to hybridize to the major 3.8-kb c-*myb* transcript (Figure 5A) in all four RNA samples and additionally to a smear of higher mol. wt transcripts in thymus and to a lesser extent R1 RNA samples. A readily observable band of 4.3 kb was also present in the R1 RNA sample. These higher mol. wt transcripts were not apparent with WEHI-3B and AKR1 RNA at this autoradiographic exposure, and were detectable only upon prolonged exposure of the film (data not shown). The 5′ pSPC4 probe, in contrast, did not demonstrate hybridization to the major 3.8-kb c-*myb* RNA; rather, hybridization was restricted to the higher mol. wt transcripts (including the 4.3-kb R1 species). These latter transcripts, estimated to be 4.3−5.6 kb in size, were also detectable using a pSPC4 single-stranded RNA (anti-sense) probe (data not shown), indicating that these transcripts were of

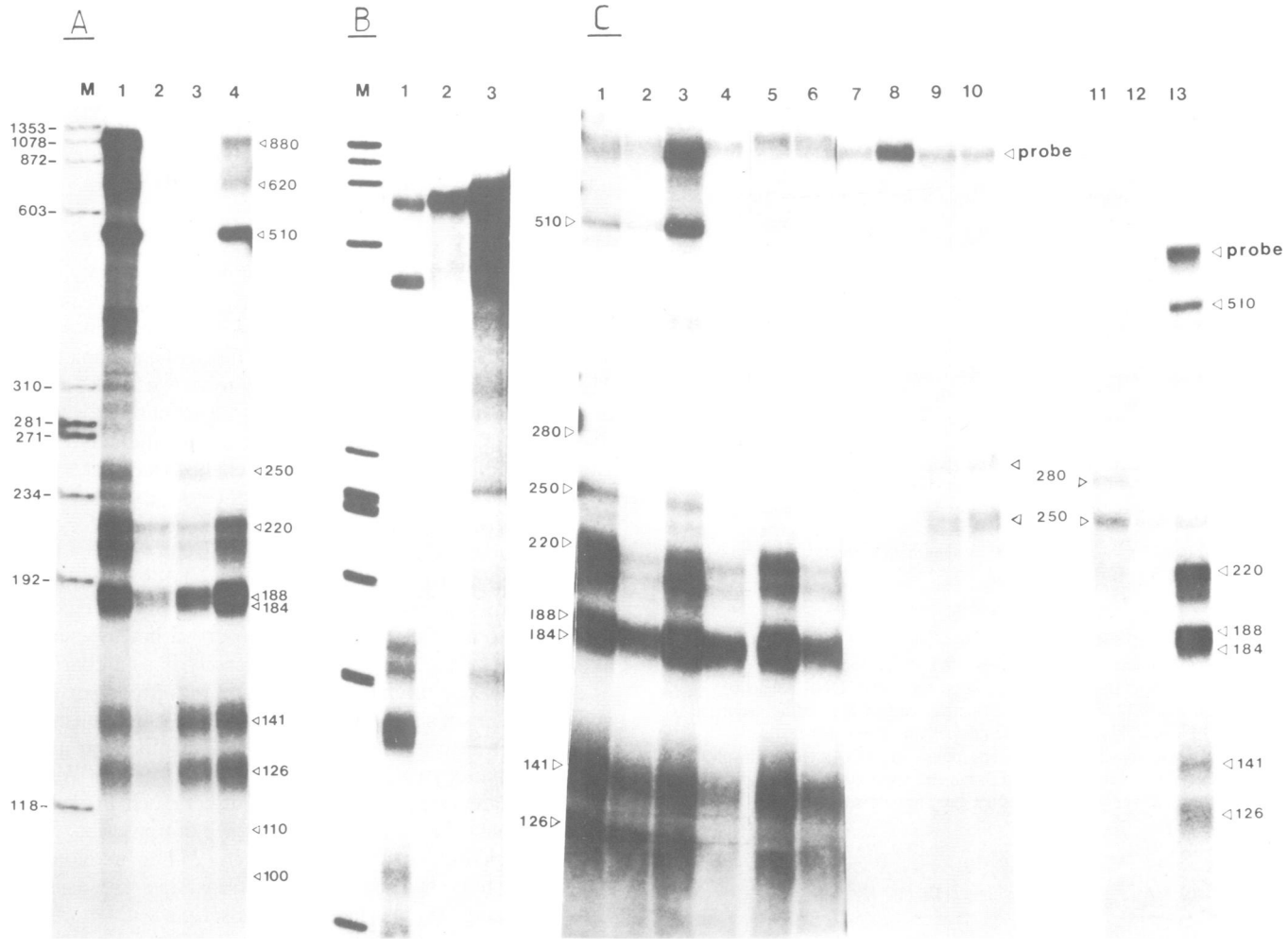

Fig. 3. RNase analyses of c-*myb* transcripts. **(A)** RNA/RNA hybrids formed between a single-stranded RNA probe equivalent to probe N (Figure 1B) and 5 μg polyadenylated RNAs isolated from **(1)** C57BL thymocytes, **(2)** WEHI-3B, **(3)** AKR1 and **(4)** R1 cells were treated with RNases and displayed on a 4.5% polyacrylamide/urea gel. **(B)**RNA/RNA hybrids formed between a riboprobe complementary to nucleotides −724 through +4 and **(1)** polyadenylated R1 cell RNA and **(2)**, **(3)** different amounts of *in vitro* pSPC5 c-*myb* RNA were analysed as in (A). **(C)** Riboprobe analysis of various haemopoietic cell RNAs. RNA/RNA hybrids formed between a riboprobe complementary to nucleotides −724 through +4 and 40 μg cytoplasmic RNA isolated from **(1)** S49, **(2)** EL4, **(3)** R1, **(4)** BW5147, **(5)** M707T, **(6)** FDCP-mix-1, **(7)** NIH 3T3, **(8)** NIH 3T3 treated with cycloheximide for 3 h, **(9)** Ag-8, **(10)** Ag-8 cells treated with cycloheximide for 2 h, **(11)** MPC-11, **(12)** MOPC-31C and **(13)** mouse spleen were analysed as in (A). Indicated are the two predominant 280-base and 250-base fragments observed with myeloma RNA.

the same polarity as c-*myb* mRNA. It may be concluded, therefore, that the generally observed 3.8-kb c-*myb* is initiated only at cap sites between −97 and −244 in the c-*myb* gene. It is likely that the 4.3-kb species is initiated at one or more of the upstream cap sites (between −505 and −875) and is processed similar to the 3.8-kb species. The smear of transcripts up to 5.6 kb suggests the presence of incorrectly processed c-*myb* transcripts both in thymus and R1 cells, as RNAs of this length could be generated, presumably, only by incomplete removal of introns or by splicing between cryptic donor/acceptor sites).

Nuclease-sensitive sites mapping near the 5′ end of c-myb

To establish whether active transcription of c-*myb* may be correlated with the presence of nuclease-sensitive sites in chromatin in the vicinity of the RNA cap sites, nuclei isolated from WEHI-3B cells, NIH 3T3 fibroblasts Ag-8 cells and from thymocytes were treated with increasing concentrations of DNase I. DNA extracted from these nuclei was then cleaved with restriction endonucleases, such as *Sac*I and *Xba*I that yielded fragments encompassing the transcription initiation sites (Figure 1A). To localize sites within these DNA fragments susceptible to DNase I cleavage, Southern blots were hybridized with probes mapping to one end of the relevant 6.6-kb *Sac*I and 9.5-kb *Xba*I fragments. For example, it was found that DNase I treatment of WEHI-3B nuclei cleaved the 6.6-kb *Sac*I fragment at a major site (site I) located 2.4 kb from the rightmost end, and at a prevalent but less intense site (site II) located 3.1 kb from this end (Figures 1 and 6A). Sites in equivalent positions were found also with thymocyte nuclei, as indicated by 2.05-kb and 2.75-kb subfragments of the 9.5-kb *Xba*I fragment (Figure 6D). DNase I-hypersensitive sites I and II can be localized to approximately nucleotides −300 and −1000, respectively, in the c-*myb* DNA sequence, i.e. immediately 5′ of the cap sites used by the predominant 3.8-kb c-*myb* mRNA and those upstream cap sites used most prevalently in thymocytes. Further nuclease-sensitive sites represented by 1.6-kb and 1.2-kb subfragments of the 6.6-kb *Sac*I fragment (Figure 6A) were also apparent both in WEHI-3B and thymocytes nuclei. However, these additional sites, which are located in the first intron of c-*myb*, are clearly less sensitive to DNase I cleavage than either sites I or II in these cells.

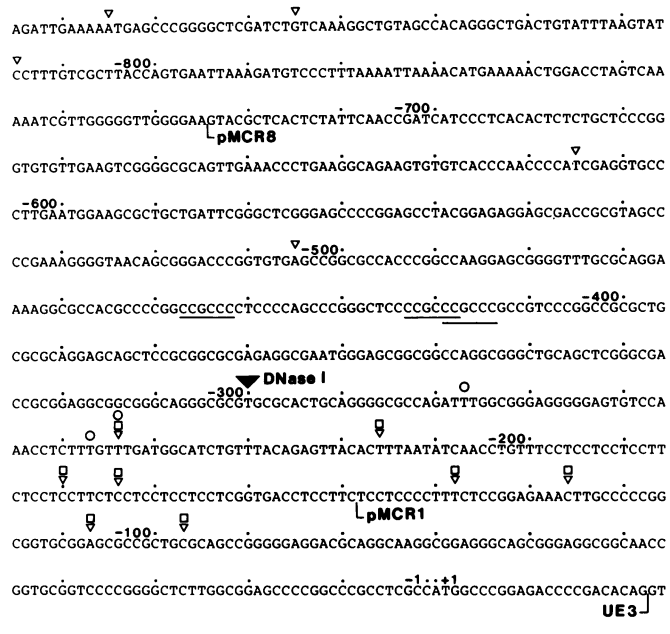

Fig. 4. Cap sites of c-*myb* transcripts. Indicated are the 5′ termini of the predominant thymocyte (open triangles), WEHI-3B (open squares) and Ag-8 (open circles) c-*myb* transcripts mapped on the c-*myb* DNA sequence (Bender and Kuehl, 1986). The presumptive start of the coding sequence is designated +1 and 5′ non-coding sequences are given in negative numeration. Also shown are the 5′ end points of cDNAs cloned in pMCR1 and pMCR8, the 3′ terminus of UE3 and the approximate location of DNase 1-sensitive site I. The core SP1 binding site sequences are underlined.

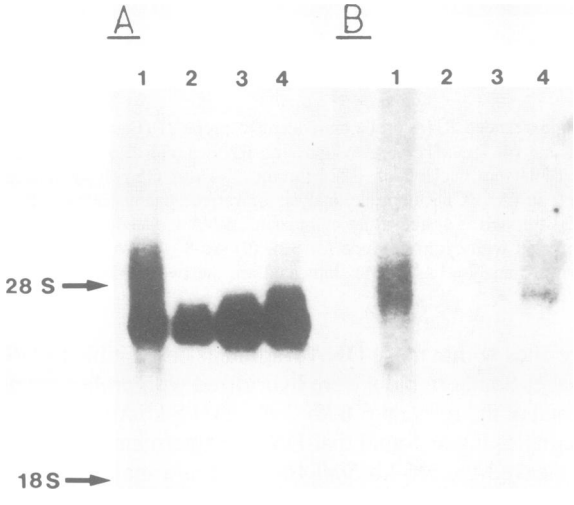

Fig. 5. Northern blot analyses of c-*myb* RNAs. Duplicate 5 μg polyadenylated RNA samples isolated from **(1)** C57BL thymocytes, **(2)** WEHI-3B, **(3)** AKR1 and **(4)** R1 cells were separated on a formaldehyde agarose gel and blotted onto nylon membrane. These blots were then hybridized with **(A)** pMCR1 and **(B)** pSPC4 nick-translated DNA probes.

In contrast to the results above, no nuclease-sensitive sites were found in the 5′ region of c-*myb* in NIH 3T3 nuclei (Figure 6B). However, in Ag-8, myeloma cell nuclei site I (and the two additional sites within the first intron) was clearly detected, whereas site II was not detectable (Figure 6C). It is noteworthy, then, that cells in which c-*myb* transcripts were detectable (regardless of the abundance) contain nuclease-sensitive site I, while this site is completely absent from cells which do not contain c-*myb* mRNA. Site II is absent from myeloma cells in which the steady-state levels of c-*myb* transcripts are considerably less than found in more immature haemopoietic cells, but is present in both WEHI-3B and thymocytes which contain transcripts initiated at upstream cap sites in very different amounts.

Discussion

We have demonstrated that considerable heterogeneity exists in the 5′ boundary of UE3 of c-*myb* transcripts and, at least quantitatively, that there is quite substantial variation between the prevalent mRNA 5′ termini observed in different haemopoietic cells. The proposition that this heterogeneity reflects usage of a number of alternative transcription initiation sites cannot be proven definitively by this type of analysis, yet it is most unlikely that such variation could arise by a splicing mechanism as most 5′ termini are not associated with a potential splice acceptor site. Nevertheless, it is possible that splicing of UE3 sequences, for example using potential acceptor sites at sequences −124/−123 and −434/−433 and donor sites within the UE3 domain, or, of course, from sequences transcribed upstream of this region, does contribute to a small extent to the observed heterogeneity. An alternative explanation of 5′ heterogeneity is as a consequence of exonucleolytic or endolytic digestion of a precursor initiated at some upstream site. If this were indeed so then it must be concluded that transcripts having longer UE3 components are more stable in thymocytes than they are in cells such as WEHI-3B where they are much less prevalent. Further, such an explanation cannot account for the relative prevalence of transcripts with cap sites at nucleotides −277 and −247/−244 in myeloma cells unless it is hypothesized that these transcripts are resistant to further cleavage which generates transcripts with shorter UE3 components in other haemopoietic cells. For these reasons, it seems most likely that 5′ heterogeneity does indeed reflect initiation of c-*myb* transcripts at a plethora of cap sites.

Data suggesting 5′ heterogeneity of c-*myb* transcripts isolated from the murine pre-B cell leukaemia 70 Z/3B have recently been published (Bender and Kuehl, 1986). Curiously, the 5′ termini so mapped were in part different from those reported here, although they encompassed broadly the same c-*myb* upstream region (nucleotides −161 to −947). Most notably the prominent initiation site at nucleotide −505 was not observed with 70 Z/3B. By extrapolation from human leukaemias, pre-B cells may represent the cell in which the highest levels of c-*myb* transcripts are found (Mavilio *et al.*, 1986), and it is then possible, though by no means certain, that these apparently different initiation sites may reflect real differences between cells that transcribe c-*myb* at higher and lower levels.

5′ mRNA heterogeneity has been reported for a number of other genes specifying, for example, human epidermal growth factor (Ishii *et al.*, 1985), hamster hydroxymethylglutaryl-CoA reductase (Reynolds *et al.*, 1984), mouse hypoxanthine phosphoribosyltransferase (HPRT) (Melton *et al.*, 1986) and mouse dihydrofolate reductase (dhfr) (McGrogan *et al.*, 1985). A common feature of such genes is the absence of a TATA box with associated CAAT box sequence in the promoter region and the presence of a GC-rich sequence containing multiple copies of the core sequence (CCGCCC) of the SP1 transcription factor binding site. The CCGCCC motifs of mouse HPRT and dhfr genes are associated with repeated sequences that demonstrate promoter activity *in vivo* and which lie close to the predominant 5′ mRNA termini (McGrogan *et al.*, 1985; Melton *et al.*, 1986). Although three

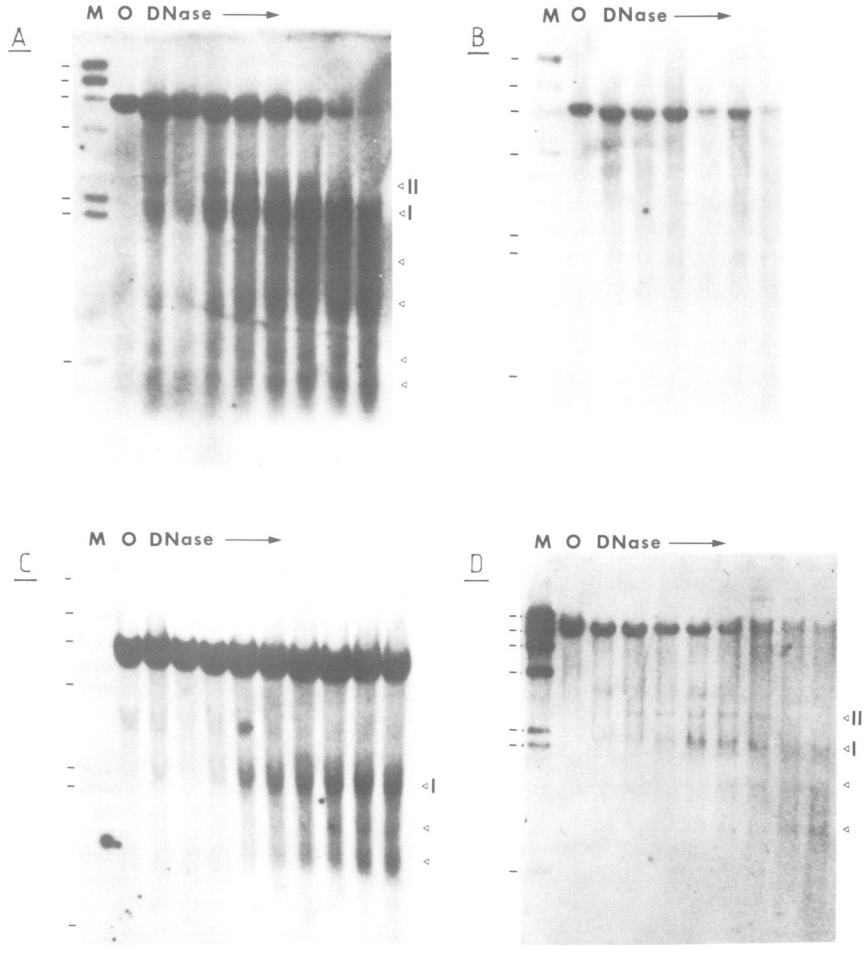

Fig. 6. Mapping c-*myb* nuclease hypersensitive sites. Nuclei prepared from **(A)** WEHI-3B, **(B)** NIH 3T3, **(C)** Ag-8 and **(D)** mouse thymocytes were either untreated (○) or treated with increasing concentrations of DNase I. Following limited DNase I digestion, DNA was isolated from nuclei and recut with *Sac*I (A, B and C) or *Xba*I (D), separated by agarose gel electrophoresis and blotted onto nitrocellulose. These blots were hybridized with nick-translated probes, either a 0.85-kb *Apa*I/*Sac*I or a 1.05-kb *Hin*dIII/*Xba* fragment (Figure 1A), homologous to one end of the 6.6-kb *Sac*I and 9.5-kb *Xba*I fragments that encompass the c-*myb* promoter. Indicated by triangles are the two major fragments (3.1-kb and 2.4-kb with *Sac*I; 2.75-kb and 2.05-kb with *Xba*I) that result from cleavage at nuclease hypersensitive sites I and II and the minor sites that are apparent upon increased DNase I digestion. Also indicated are the locations of the *Hin*dIII fragments of lambda DNA used as markers.

such CCGCCC motifs are present in the c-*myb* 5′ sequence, these do not form parts of larger repeated units and lie >200 bp 5′ to the most predominant mRNA termini. Thus there is no *a priori* indication that these GC sequences have a role in transcription of the c-*myb* gene.

The use (e.g. in thymocytes) of widely separated cap sites for initiation of c-*myb* transcripts suggests that discrete promoters may be active in these cells in transcription of this gene. An analogous situation has been observed with the mouse dhfr gene where a sequence upstream of the GC-rich promoter appears to regulate expression of alternative, less abundant, and rather variable, dhfr mRNAs (McGrogan *et al.*, 1985). The presence in thymocytes of an additional upstream nuclease-sensitive site may possibly reflect usage of this putative alternative promoter. Such a promoter may not be entirely silent in cells such as WEHI-3B, as lesser amounts of transcripts initiated at upstream cap sites were observed in these cells also. The relationship between the predominant myeloma c-*myb* transcripts and those initiating immediately downstream is further open to conjecture. Transcripts initiated at −277 and −247/−244 are of similar abundance, for example, in both Ag-8 and S49, whereas those initiated downstream are vastly more abundant in S49 (Figure 3C). Initiation at downstream cap sites may then be a process by which the abundance of c-*myb* mRNAs is regulated. It will therefore be interesting to determine whether initiation at these downstream cap sites involves the use of an additional promoter to that operative in myeloma cells, or whether additional factors operate to initiate transcription at downstream cap sites by polymerase molecules that entered the transcriptional complex at a common site some way upstream (perhaps site I identified by nuclease hypersensitivity). Clearly, maturation of pre-B cells (Bender and Kuehl, 1986) to mature antibody secreting B cells (represented by myelomas) results in a considerable reduction in c-*myb* steady-state levels, and it will be interesting to determine whether this effect can be accounted for by a mechanism involving selective or preferential cap site usage.

The majority of 3.8-kb c-*myb* mRNAs, having 5′ non-coding sequences of 97−217 bases, contain no ATG triplet upstream of that which putatively initiates translation at nucleotide +1. A further ATG (nucleotides −241/−239) present in the predominant myeloma c-*myb* mRNAs is quickly followed by an in-frame terminator (TAA, nucleotides −214/−212). c-*myb* transcripts

having the longest 5′ non-coding sequences contain several additional ATGs, all but that at nucleotides −362/−360 are followed by an open reading frame of 510 nucleotides in cDNA (which is out of frame to the c-*myb* protein coding sequence) which may encode alternative proteins, possibly by a process involving differential splicing. Alternatively, transcripts initiated at upstream cap sites may merely be adventitious species that do not encode proteins and are in general not correctly processed, as may be inferred from the smear of high mol. wt transcripts observed with thymus RNA using the 5′ pSPC4 probe in Northern blot analysis (Figure 5B).

We were unable to detect the 3.8-kb c-*myb* mRNA in NIH 3T3 fibroblasts, even in cells treated for 3 h with cycloheximide, conditions under which the cognate chicken mRNA was readily detectable in chick embryo fibroblasts (Thompson *et al.*, 1986). Analyses of mouse embryo fibroblasts (data not shown) showed that whereas c-*myc* mRNA was inducible with either cycloheximide treatment or serum stimulation, c-*myb* mRNA again could not be detected under these conditions. Homology between mouse and chicken c-*myb* terminates abruptly immediately upstream of the +1 position of the murine gene. Indeed, the chicken c-*myb* gene appears to contain an additional amino-terminal coding sequence of 174 bp and does not contain a TC-rich cap site region as does the murine gene (Rosson and Reddy, 1986), implying that chicken c-*myb* transcripts are not initiated using a promoter equivalent to that predominant in the mouse. It will be interesting to see whether the upstream murine cap sites are vestiges of that still utilized in chickens.

The concepts that the murine c-*myb* gene may be transcribed from alternative promoters or that specific nuclear factors influence cap site usage, that this results in regulation of the abundance of c-*myb* mRNA and that transcripts initiating at the downstream cap sites only appear to give rise to the recognized 3.8-kb c-*myb* mRNA, have important bearing upon studies of the regulation of this gene. We intend to study, by nuclear run-off assays, the utilization of these promoters in various cells, to determine how this relates to the steady-state levels of these transcripts and to study the effect of differentiation stimuli on their respective usage.

Materials and methods

Cell lines

Thymuses were prepared from young (<6 week) C57BL and BALB/c mice. The origins of the various cell lines utilized are given in Table I. The WEHI-3B myelomonocytic leukaemia was the differentiation-positive (D^+) subline and was obtained from Dr Herman Waldmann, University of Cambridge. T cell leukaemias S49, EL4, BW5147, R1 and AKR1, the myeloma Ag-8 and the erythroleukaemia M707T were obtained from Dr Michael Horton, St Bartholomew's Hospital. Myelomas MPC-11 and MOPC-31C were obtained from Flow Laboratories. These cells were maintained in RPMI 1640 medium supplemented with 10% fetal calf serum. FDCP-mix-1 cells are pluripotent stem cells and were obtained from Dr Michael Dexter. These cells are dependent on IL-3 and were maintained in Fischer's medium supplemented with 20% horse serum and 10% WEHI-3B D^- conditioned medium.

Molecular cloning of c-myb genomic DNA and cDNA

Methods used were, unless otherwise indicated, those published previously (Maniatis *et al.*, 1982). DNA isolated from C57BL mouse spleen was partially digested with limited amounts of the restriction endonuclease *Sau*3A. DNA fragments in the size range 10−20 kb were isolated from an agarose gel, eluted and ligated with *Bam*HI cut arms of the bacteriophage vector L47 (Loenen and Brammer, 1980). Recombinant bacteriophage were packaged *in vitro* and plated on agar with *Escherichia coli* LE392. Approximately 2×10^6 plaques were screened in total by replica plating onto nitrocellulose and probing with the 1.3-kb v-*myb* containing *Kpn*I/*Sac*I fragment of pVM2 (Klempnauer *et al.*, 1982) labelled by nick-translation. Eight recombinants with homology to v-*myb* were obtained, of which KR1A, SR5A and KR4A are representative, and these were further characterized by restriction analysis and blot hybridization.

cDNA was synthesized using AMV reverse transcriptase and 5 μg of C57 mouse thymocyte polyadenylated RNA using the method of Gubler and Hoffman (1983). Double-stranded DNA was G-tailed using terminal transferase and annealed with pUC9 (Vieira and Messing, 1982) C-tailed at *Pst*I termini. Transformed *E. coli* (~50 000 colonies) were screened by replica plating on nitrocellulose and probing with v-*myb* as above. A single plasmid, pMCR1, was thus isolated, and this was used to screen additional transformants, yielding several more c-*myb* cDNA-containing plasmids including pMCR8.

RNA extraction

Total cell RNA was extracted from tissue and cells using the guanidinium/hot phenol method (Maniatis *et al.*, 1982). Polyadenylated RNA was selected by chromatography on oligo(dT)−cellulose columns as described previously (Watson *et al.*, 1979). Cytoplasmic RNA was isolated by lysis of cells with NP40 and subsequent phenol/chloroform extraction as described previously (Watson *et al.*, 1979). When applicable, nuclear RNA preparations were made by extraction of the nuclear pellet using the guanidinium/hot phenol method.

S1 nuclease and RNase analyses

DNA restriction fragments were 5′-^{32}P-labelled using [γ-^{32}P]ATP and polynucleotide kinase to a specific activity of 10^6-10^7 c.p.m./μg. Approximately 10^5 c.p.m. of the labelled DNA fragment was then co-precipitated with 5 μg polyadenylated RNA and 15 μg yeast tRNA using 70% ethanol. Controls contained the labelled probe plus 20 μg yeast tRNA. The nucleic acids were pelleted, dried briefly and taken up in 20 μl 80% formamide containing 0.4 M NaCl, 1 mM EDTA and 40 mM PIPES, pH 6.5. After overlaying with paraffin oil, the samples were heated for 10 min at 70°C, then incubated overnight at 52°C. S1 nuclease digestions were performed by adding 250 μl of buffer (250 mM NaCl, 1 mM $ZnCl_2$, 5% glycerol, 30 mM NaOAc, pH 4.6) and 75 U of the enzyme (Bethesda Research Laboratories, Inc.), and incubating at 37°C for 30 min. Samples were phenol/ chloroform extracted, ethanol precipitated and taken up in loading buffer (95% formamide, 1 mM EDTA, 10 mM Tris−HCl pH 7.5 containing bromophenol blue and xylene cyanol dyes). After denaturation at 90°C for 2 min, samples were applied to a 4.5% acrylamide/8 M urea gel and electrophoresed using 2 × TBE (1 × TBE is 50 mM Tris, 50 mM boric acid, 1 mM EDTA).

Riboprobe analyses were performed essentially as described by Melton *et al.* (1984). Genomic and cDNA fragments were cloned in the vector pSP65 to yield the following plasmids: pSPC1 contains a cDNA sequence of nucleotides −724/+4 inserted in the *Hin*dIII/*Sma*I sites of pSP65; pSPC4 contains cDNA nucleotides −724/−339 inserted in the *Hin*dIII/*Pst*I sites; pSP1.7NH contains a 1.7-kb *Nco*I/*Hin*dIII genomic DNA fragment extending upstream from nucleotide +4 inserted in the *Hin*dIII/*Sma*I sites. These plasmids were linearized by digestion with *Hin*dIII and 0.5-μg portions were used to direct synthesis of ^{32}P-labelled anti-sense RNA by SP6 RNA polymerase in a reaction volume of 20 μl containing 40 mM Tris, pH 7.5, 6 mM $MgCl_2$, 2 mM spermidine, 10 mM dithiothreitol, 500 μM ATP, 500 μM CTP, 500 μM UTP, 26 μM GTP (50 μCi total), 10 U RNasin (BRL) and 5 U of the polymerase. Incubation was for 1 h at 37°C, after which RNase-free DNase to a concentration of 20 μg/ml and a further 10 U RNasin were added and incubation was continued for 20 min at 37°C. The preparations were then phenol/ chloroform extracted once, and ethanol precipitated twice in the presence of 2 M ammonium acetate. Labelled RNA was dissolved in 100 μl water and stored at −70°C. Denaturing gel analysis showed that riboprobes made under these conditions were almost entirely of full-length and they were used without further purification as hybridization probes. Hybridizations were done with 5 μg polyadenylated or 20−40 μg cytoplasmic RNA and ~10^5 c.p.m. of the riboprobe in a volume of 20 μl of the 80% formamide solution described above. After denaturing at 85°C for 5 min, samples were incubated overnight at 50°C. Samples were then diluted with 300 μl of RNase digestion buffer (300 mM NaCl, 5 mM EDTA, 10 mM Tris−HCl pH 7.5) containing 40 μg/ml RNase A and 600 U/ml RNase T1. Digestions were done at room temperature for 1 h, after which 25 μl 10% SDS and 25 μl 5 mg/ml proteinase K were added and incubation was continued for a further 20 min at 37°C. After phenol/chloroform extraction, 10 μg carrier tRNA was added and nucleic acids were ethanol precipitated. Pellets were taken up in loading buffer and run on denaturing polyacrylamide gels as described above.

Unlabelled pSPC65 c-*myb* sense RNA was made *in vitro* as above, except that the concentration of GTP was 100 μM. It was necessary to purify the full-length transcript by electrophoresis on a 5% polyacrylamide/8 M urea gel and elution as described (Maxam and Gilbert, 1980). Approximately 0.1-ng to 10-ng portions of this purified RNA were hybridized with 10^5 c.p.m. of the relevant DNA or RNA probe, and hybrids were nuclease treated as above.

Northern blot analysis

Five micrograms of polyadenylated RNA were applied to 1% agarose gels containing 2.2 M formaldehyde as described by Maniatis *et al.* (1982). Following electrophoresis, RNAs were blotted onto Biodyne nylon membrane (Pall Corporation) in 20 × SSC. After baking at 80°C for 2 h in a vacuum oven, blots were pre-hybridized in 50% formamide containing 3 × SSC, 0.1% SDS, 20 mM sodium phosphate, pH 7.0, 100 μg/ml denatured calf thymus DNA and

5 × Denhardt's (Maniatis *et al.*, 1982) for several hours at 50°C. Fresh hybridization buffer (as above) containing 10^6 c.p.m./ml of the nick-translated probe was then added and hybridization was carried out at 50°C for 2 days. Blots were washed three times for 30 min in 0.1 SSC/0.1% SDS at 65°C.

Mapping DNase I hypersensitive sites

Nuclei were prepared and treated with pancreatic DNase I as described previously (Dyson *et al.*, 1985). Each DNase I digestion contained ~10^7 nuclei in 1 ml volume and between 0.1 and 5 μg of the enzyme. Approximately 20 μg of each sample was subsequently digested with *Sac*I or *Xba*I and applied to a 0.8% agarose gel. After blotting to nitrocellulose, hybridizations were performed at 42°C in the hybridization buffer described above.

Acknowledgements

We thank John Wyke and Gordon Peters for their constructive criticisms, Andrea Kessler for preparation of this manuscript and Hermann Waldmann, Mike Dexter and Mike Horton for providing cell lines.

References

Bender,P. and Kuehl,W.M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3204–3208.
Bishop,J.M. and Varmus,H. (1985) In Weiss,R., Teich,N., Varmus,H. and Coffin,J. (eds). *The Molecular Biology of Tumor Viruses, Part IV, RNA Tumor Viruses*. Cold Spring Harbor Laboratory Press, NY, pp. 351–356.
Boyle,W.J., Lipsick,J.S. and Baluda,M.A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4685–4689.
Castle,S. and Sheiness,D. (1985) *Biochem. Biophys. Res. Commun.*, **132**, 688–695.
Craig,R.W. and Bloch,A. (1984) *Cancer Res.*, **44**, 442–446.
Dyson,P.J., Littlewood,T.D., Forster,A. and Rabbitts,T.H. (1985) *EMBO J.*, **4**, 2885–2892.
Gonda,T.J. and Metcalf,D. (1984) *Nature*, **310**, 249–251.
Gonda,T.J., Sheiness,D.K. and Bishop,J.M. (1982) *Mol. Cell. Biol.*, **2**, 617–624.
Gonda,T.J., Gough,N.M., Dunn,A.R. and de Blaquiere,J. (1985) *EMBO J.*, **4**, 2003–2008.
Gubler,U. and Hoffman,B.J. (1983) *Gene*, **25**, 263–269.
Harrison,P.R., Affara,N., Goldfarb,P.S., Kasturi,K., Yang,Q.S., Lyons,A., O'Prey,J., Fleming,J., Black,E. and Nichols,R. (1982) *Adv. Exp. Med. Biol.*, **158**, 81–88.
Horibata,J. and Harris,A. (1970) *Exp. Cell Res.*, **60**, 61–77.
Hyman,R. and Stallings,V. (1974) *J. Natl. Cancer Inst.*, **52**, 429–435.
Hyman,R., Cunningham,K. and Stallings,V. (1980) *Immunogenetics*, **10**, 261–271.
Ishii,S., Xu,Y.-H., Stratton,R.H., Roe,B.A., Merlino,G.T. and Pastan,I. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4920–4924.
Klempnauer,K.-H., Gonda,T.J. and Bishop,J.M. (1982) *Cell*, **31**, 453–463.
Klempnauer,K.-H., Ramsay,G., Bishop,J.M., Moscovici,G.M., Moscovici,C., McGrath,J.P. and Levinson,A.D. (1983) *Cell*, **33**, 345–355.
Kohler,G. and Milstein,C. (1975) *Nature*, **256**, 495–497.
Laskov,R. and Scharff,M.D. (1970) *J. Exp. Med.*, **131**, 515–541.
Lavu,S. and Reddy,E.P. (1986) *Nucleic Acids Res.*, **14**, 5309–5320.
Loenen,W.H.M. and Brammar,W.J. (1980) *Gene*, **20**, 249–259.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
Mavilio,F., Sposi,N.M., Petrini,M., Bottero,L., Marinucci,M., De Rossi,G., Amadori,S., Mandelli,F. and Peschle,C. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4394–4398.
Maxam,A. and Gilbert,W. (1980) In Grossman,L. Moldave,K. (eds), *Methods in Enzymology*. Academic Press, New York, Vol. 65, 43–62.
McGrogan,M., Simonsen,C.C., Smouse,D.T., Farnham,P.J. and Schimke,R.T. (1985) *J. Biol. Chem.*, **260**, 2307–2314.
Melton,D.A., Krieg,P.A., Rebagliati,M.R., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.
Melton,D.W., McEwan,C., McKie,A.B. and Reid,A.M. (1986) *Cell*, **44**, 319–328.
Metcalf,D. and Nicola,N.A. (1982) *Int. J. Cancer*, **30**, 773–780.
Ralph,P. (1973) *J. Immunol.*, **110**, 1470–1475.
Reed,J.C., Alpers,J.D., Nowell,P.C. and Hoover,R.G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3982–3986.
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.
Rosson,D. and Reddy,E.P. (1986) *Nature*, **319**, 604–606.
Sheiness,D. and Gardinier,M. (1984) *Mol. Cell. Biol.*, **4**, 1206–1212.
Shen-Ong,G.L.C., Morse,H.C., Potter,M. and Mushinski,J.F. (1986) *Mol. Cell. Biol.*, **6**, 380–392.
Spooncer,E., Heyworth,C.M. and Dexter,T.M. (1986) *Differentiation*, **31**, 111–118.
Stern,J.B. and Smith,K.A. (1986) *Science*, **233**, 203–206.
Thompson,C.B., Challoner,P.B., Neiman,P.E. and Groudine,M. (1986) *Nature*, **319**, 374–380.
Torelli,G., Selleri,L., Donnelli,A., Ferrari,S., Emilia,G., Venturelli,D., Moretti, L. and Torelli,U. (1985) *Mol. Cell. Biol.*, **5**, 2874–2877.
Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259–268.
Watson,R.J., Preston,C.M. and Clements,J.B. (1979) *J. Virol.*, **31**, 42–52.
Westin,E.H., Gallo,R.C., Arya,S.K., Souza,L.M., Baluda,M.A., Aaronson,S.A. and Wong-Staal,F. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 2194–2198.

Received on October 29, 1986; revised on March 23, 1987