Structure and biological activity of the transcriptional initiation sequences of the murine c-myb oncogene

Peter W. Sobieszczuk*, Thomas J. Gonda and Ashley R. Dunn

Ludwig Institute for Cancer Research, Melbourne Branch of Tumor Biology, PO Royal Melbourne Hospital, Parkville, Victoria 3050, Australia

Received September 15, 1989; Accepted November 9, 1989 EMBL accession nos X16389, X16390

ABSTRACT
To study the control mechanism(s) that govern the transcription of c-myb, genomic clones corresponding to the 5’ region of the murine c-myb gene have been isolated and characterized structurally and functionally. Primer extension and nuclease protection analysis have revealed the presence of multiple transcriptional initiation sites, that are utilized in several hemopoietic cell lines (WEHI3B(D+), FDC-P1 and RB22.2). Some of the sites are used in all cell lines but others are unique; all are located in a region of the c-myb gene that is G-C rich, contains a number of potential Spl binding sites and lacks classical promoter consensus sequences.

Experiments in which well characterized promoters controlling expression of a reporter gene have been replaced by fragments of c-myb DNA (including the observed cap sites) were performed in an attempt to demonstrate promoter activity in various cell types. It was shown that a region of the c-myb gene (approximately 1.0 kbp upstream from the splice donor site of the first exon) contains a weak promoter that has a low level of transcriptional activity in hemopoietic as well as in fibroblastic cells. These results support the suggestion that c-myb expression is not regulated primarily at the level of initiation of transcription.

INTRODUCTION
The c-myb gene is the cellular homologue of a gene (v-myb) carried by the replication defective avian myeloblastosis retrovirus (AMV) which causes acute myelomonocytic leukemia in chickens and transforms immature avian myeloid precursors and mature macrophages in vitro [1,2]. The mouse c-myb gene has been shown to be a target of insertional mutagenesis by Moloney murine leukemia virus (Mo-MLV) that leads to the induction of myeloid hemopoietic malignancies [3—6]. The protein encoded by the c-myb gene (p75c-myb) is localized in the nucleus and has been shown to bind to DNA in vitro [8,9]. The structural homology that exists between the c-myb gene and the transcriptional regulatory factor Cl from Zea mays[10] suggests that c-myb may function as transcriptional regulator. This possibility is supported by the observations that myb protein acts as a sequence-specific DNA binding protein [11] with transcriptional trans-activator activity [12]. c-myb is predominantly expressed in immature hemopoietic cells [13—15] and is subject to a down-regulation in a differentiation-specific manner, i.e. induction of hemopoietic progenitor cells to differentiate results in a substantial decrease of steady-state levels of c-myb mRNA [16,17]. Expression of c-myb also appears to be cell cycle specific in that the steady state level of c-myb mRNA increases, in some cases, when cells are induced to proliferate, with the peak in late G1 and S phases [18].

In the present study, we have performed structural and functional analysis of the mouse c-myb gene in an attempt to determine the basis for its pattern of expression. We have
mapped the 5'-terminal c-myb exon and an upstream region containing the transcriptional initiation sites and shown that this region includes a weak promoter that is active in hemopoietic and fibroblastic cells.

**MATERIALS AND METHODS**

**Screening the library and sequencing**

The genomic library from which the λmyb1 and λmyb2 clones were isolated was constructed within the λ vector Charon 4A. Genomic inserts were recloned into pUC8, pGEM3 and M13 vectors, and DNA subfragments of interest were sequenced. Sequencing of M13 clones was performed by the dideoxy method described by Sanger et al., [19]. Double stranded sequencing was performed using GemSeq K/RT Sequencing System kit in accordance with the manufacturer's technical manual (Promega Biotech).

**Primer extension and riboprotection analysis of RNA**

Primer extension analysis was performed on polyadenylated RNA as follows. A 30 base oligonucleotides complementary to the 5' region of the c-myb gene sequence were radiolabelled using polynucleotide kinase, followed by heating to 65°C for 15 minutes to inactivate the enzyme. 50 ng of the primer was hybridized with 1 μg of poly(A+ ) RNA in a solution of 0.3 M NaCl, 10 mM Tris·Cl (pH = 7.5), 1 mM EDTA at 65°C overnight. The nucleic acids were precipitated with 0.3 volumes of 3 M sodium acetate (pH = 7.0) and 2.5 volumes of ethanol. Following washing with 70% ethanol, the pellet was redissolved in 50 μl of reverse transcription buffer (50 mM Tris·Cl (pH = 8.3), 20 mM KCl, 10 mM MgCl2, 10 mM DTT and 1 mM of all four dNTPs) to which was added 1 μl of AMV reverse transcriptase (9.5 units/μl). Primer extension was allowed to proceed for 1 hour at 40°C after which the reaction products were precipitated with 0.3 volumes of 3M sodium acetate and 2.5 volumes of ethanol. Following washing with 70% ethanol the pellet was dissolved in sequencing loading solution. The products of the reaction were analyzed on a sequencing gel.

Nuclease protection mapping was performed with antisense [32]P-RNA synthesized in vitro [20] from the SP6 and T7 promoters in a linearized pGEM3 subclone of the appropriate genomic clone (see fig.1 and 5). A transcription kit from Promega Biotech was utilized, using 50 to 100 μCi of α ([32P]UTP and 2 μg of linearized template. Between 5×10⁷ and 1×10⁸ counts were incorporated into ethanol precipitable material. The labelling protocol was carried out according to the manufacturers' recommendations. Following hybridization with poly(A+) RNA from the appropriate cell type, unprotected RNA strands were digested with RNase A and RNase T₁ according to the Riboprobe System protocol (Promega Biotech).

**Tissue culture and transfection**

NIH3T3 cells were maintained in DME (Dulbecco's modified Eagle's medium) and 10% NCS (Newborn Calf Serum) at 37°C in an atmosphere of 10% CO₂. WEHI3B(D+), P3K and K562 cells were maintained in DME supplemented with 10% FCS (Foetal bovine serum) at 37°C and 10% CO₂. FDC-P1 media was supplemented with WEHI-3B [21] conditioned media and the media for RB22.2 [22] was supplemented with 50 μM 2-mercaptoethanol.

**Calcium phosphate transfection**

Transfections were performed essentially by the method of Graham and van der Eb [23] with modifications [24]. Cells were seeded on 5 or 10 cm Nunc tissue culture dishes such that they were approximately 30% confluent at the time of transfection. Transfection solution
(500 ng of 'CAT' plasmid DNA, 31.25 μl 2M CaCl₂, 5 μg of carrier DNA (sheared murine genomic DNA) in a total volume of 250 μl) was added dropwise to 250 μl of 2×HEBS(PO₄) (HEBS(PO₄) is 140 mM NaCl, 50 mM Hepes (pH = 7.0), 1.4 mM Na₂HPO₄), with continuous bubbling. The solution was then left to stand for 20–30 minutes and then slowly added to cells in DME with 10% FCS. The cells were incubated for 12–16 hrs in the presence of the precipitate at 37°C in an atmosphere of 10% CO₂.

**Electroporation**

Electroporation was performed using BioRad Gene Pulser basically in accordance with manufacturer's application manual.

Briefly, 40 μg of superhelical DNA was used per single electroporation and approximately 10⁷ cells suspended in 0.5 ml of PBS. Voltage and capacitance condition were established for every cell line based on survival kinetics as a function of increasing dosages of electric pulses. These condition were found to be optimal when 50% to 90% of the cells were dead 12 h after the electric shock. For example hemopoietic RB22.2 cells were electroporated with single shock at 450 V and 500 μF. All experiments were performed in triplicate at room temperature. Immediately following the electric discharge, cells were transferred to the tissue culture dishes and growth medium was added. 12 h later cell survival was counted using the vital dye, trypan blue (0.2 g/l in PBS).

**CAT assay**

Cell extracts were assayed for CAT activity as described by Sleigh [25]. Fibroblast cell lines growing on 5 cm tissue culture dishes were washed with PBS, trypsinized and then suspended in 5 ml of DME/10% NCS before being pelleted by centrifugation at 2000 rpm for 10 mins (Heraeus Christ minifuge 2). The cells were then resuspended in 250 μl of 0.25 M Tris·Cl (pH = 7.5) and frozen at -20°C until required. The cell suspension was then subjected to 3 rounds of freeze thawing (liquid N₂ to 37°C), heated to 65°C for 10 mins prior to centrifugation for 10 mins in an eppendorf minifuge (to pellet cellular debris). The supernatants were sometimes stored at -20°C for a number of months without any apparent loss of CAT activity. The supernatant was assayed for CAT activity as follows: 40 μl supernatant, plus 20 μl Acetyl-CoA (0.5 mM Acetyl-CoA, 0.1 mM[¹⁴C]-Acetyl-CoA, 58 mCi/mM), plus 20 μl chloramphenicol (8 mM chloramphenicol), plus 20 μl 0.25 M Tris·Cl (pH = 7.5). Reactions were allowed to proceed for between 2 to 16 hours depending on the anticipated range of activities. The reaction mixture was then extracted once with an equal volume of ethyl acetate. The organic phase was transferred to a scintillation vial containing 2 mls of scintillation fluid. The radioactivity of the samples was determined by liquid scintillation counting (2000CA Tri-Carb Liquid Scintillation analyzer). Purified CAT enzyme was included as a control (1 unit per reaction).

**RESULTS**

**Molecular Cloning and Sequencing of the 5' End of the Murine c-myb Gene**

A bacteriophage λ library of murine genomic DNA was screened using as a probe a radiolabelled fragment of DNA derived from the 5' end of the c-myb c-DNA clone MM49 [26]. Two positive clones λmyb1 and λmyb2 were isolated and characterized using a panel of restriction enzymes. On the basis of an indistinguishable pattern of restriction fragments the two clones were judged to be identical, probably originating from the same cloning event. In addition to restriction enzyme analysis of the genomic insert, a Southern blot characterization was performed using five synthetic oligonucleotides spanning the 5' end of the c-myb cDNA. As a result of such analysis a physical map of the 13.0 kbp (5.0
kbp and 8.0 kbp EcoRI fragments) murine genomic insert was created (fig.1, A and B).

In order to identify the 5' boundary of the transcribed myb sequences a number of restriction fragments were isolated from λmyb1, subcloned into plasmid vectors, and subsequently used as radiolabelled probes on Northern blots containing mRNA from myb expressing cell lines. The two genomic HindIII fragments of 2.4 kbp and 1.4 kbp each hybridized to a 3.9 kbp mRNA while the 3.5 kbp and 0.6 kbp fragments did not (data not shown). Thus, the two fragments sharing homology with the 3.9 kbp c-myb mRNA contain exons or parts of the exons of the c-myb gene. Accordingly, sequencing efforts were concentrated on those two HindIII fragments. The nucleotide sequence of the entire 2.4 kbp HindIII fragment was determined, with the exception of approximately 150 bp at the 5' terminus. Comparison of the genomic DNA sequence with the c-myb cDNA revealed the presence of a 59 nucleotide stretch of homology at position 1171 to 1229 (fig.2). The DNA sequence around position 1229 conforms to the consensus sequence of a splice donor site; splicing from this site onto the acceptor site in the 1.4 kbp HindIII/EcoRI fragment (see fig.1) would generate a sequence contiguous with the cDNA. The uninterrupted homology between gene and cDNA sequences shows that the first 59 nucleotides of the c-myb cDNA [26] are present on a single exon suggesting, that this corresponds to the first exon of the c-myb gene.

The nucleotide sequence of the 1.4 kbp HindIII/EcoRI fragment revealed that the next exon was located 80 bp downstream from the HindIII site, allowing both donor and acceptor splice sites to be mapped by comparison to the cDNA sequence, and is shown in fig.2.

Identification of Transcription Initiation Sites

In order to provide unambiguous evidence that the exon identified in the 2.4 kbp genomic HindIII fragment, corresponding to the most 5' region of the c-myb cDNA [26,27] is in fact the first exon of the gene, the location of the c-myb mRNA cap site(s) was established by primer extension and nuclease protection mapping.

Primer extension. Four radiolabelled oligonucleotides corresponding to the first ('jul # 1', 'jul # 2'), and second ('myb # 2', 'myb # 3') exons (figs. 1 and 2) were annealed to poly(A)+ mRNA from two hemopoietic cell lines WEHI3B(D+) [28] and RB22.2 [22] known to express high levels of myb transcript. To exclude the possibility of misinterpretation of the primer extension results due to nonspecific priming of the oligonucleotides, RNAs from the NIH3T3 and P3K cell lines, which do not express detectable amounts of 5-myb [29], were used in parallel reactions.

As shown in fig.3 several bands were obtained which were present regardless of whether the cells from which mRNA was isolated expressed the c-myb gene, and were therefore excluded as authentic cap sites. Only bands specifically appearing in the lanes containing RNA from WEHI3B(D+) or RB22.2 (lanes 3 and 4) but not in NIH3T3 mRNA, (lane 2) were considered as bona fide transcription initiation sites. Accordingly, on the basis of this criterion, sixteen potential c-myb cap sites were identified in WEHI3B(D+) cells.

FIG.1. THE 5' TERMINI OF THE MURINE C-MYB ONCOGENE

A: Schematic representation of the first 5 exons of the murine c-myb gene (compiled from this work and Castle and Sheenest, [49], Lavu and Reddy, [50], Gonda et al., [6]; locations of only the first 5 exons are marked). The restriction map of the 5.0 and 8.0 kbp EcoRI (RI) fragments was generated by single and multiple digestion of λmyb1 (which covers both EcoRI fragments), with HindIII (H), BamHI (B) and PstI (P) enzymes; large and small numbers indicate sizes of EcoRI and HindIII fragments (in kbp) respectively. B: expansion of the first and second exon (shaded boxes); primer extension oligonucleotides and RNase protection probes are indicated below. C. Independently defined 5' end sites of c-myb RNAs as detected by primer extension and RNase protection analysis in WEHI3B(D+), FDC-P1 and RB22.2 cells are shown. Symbols are as indicated in the legend.
A. DNA sequence of 2.4 kbp HindIII fragment containing the first exons of c-myb. The positions of oligonucleotides 'jul#1' and 'jul#2' are indicated by arrows; the first Met codon is circled; TATA sequence is boxed and differences with the published sequence [27, 33] are shown as nucleotide insertions or '?' indicating deletions. Potential Spl binding sites found on both strands of DNA are shown as stippled boxes. Only the sites with 80% (light shading) and 90% (dark shading) homology to the Spl consensus sequence are shown. Four major transcription initiation sites are shown. The sequence suggests a signal for poly(A) tail addition (Messing et al., 1976).
Nucleic Acids Research

(fig.3 and summarized in fig.1); however, only six of these, at positions 971, 1000, 1085, 1145, 1161 and 1181, were considered authentic cap sites based on the generation of primer extension products whose lengths reflected the distance separating the paired oligonucleotides used in this analysis.

Multiple 5' termini were also observed when RNA from a second cell line RB22.2, was used (lane 4, fig.3). Three cap sites in RB22.2 appear to also be utilized in WEHI3B(D+). In addition to these (at position 1063, 1071, 1085) a fourth cap site in RB22.2 (at 1101) differed by two nucleotides from the site found in WEHI3B(D+) (at position 1103). Two initiation sites present in WEHI3B(D+) (positions 971, 1085) were not detected in RB22.2. These results suggest that c-myb transcription is initiated in hemopoietic cells at multiple sites, some of them being common to different cell types, and some being cell line-specific.

Nuclease protection analysis. To confirm the primer extension data, nuclease protection analysis was carried out. Three radiolabelled antisense RNA riboprobes, 'P405', 'P570' and 'BANco' (as shown in fig.1) spanning the region containing the putative 5' end of transcript were annealed to mRNA followed by RNase T1 and RNase A digestion. Two of these, 'P405' and 'BANco', were used in riboprotection experiments with mRNA from WEHI3B(D+). Not only were the sizes of two riboprobes different ('BANco' of 1065 nucleotides and 'P405' of 360 nucleotides) but their 5' ends differed by 69 nucleotides. As shown in fig.4A several protected bands corresponding to 5' end sites at positions 1145, 1097-1101, 1069-1076, 1018-1021 and 991-997 were generated when riboprobe 'P405' was used, (lanes 3, 4 and 5). No protected fragments were generated in control experiments using tRNA (lane 6). Moreover the RNase resistant hybrids were not present prior to ribonuclease digestion (lane 7). When riboprobe 'BANco' was used protected bands corresponding to 5' end sites at positions 1122, 1098-1103, 1069-1077, 1017-1023 and 995-999 were generated (lanes 8, 9 and 10). A close correlation of the deduced positions (as indicated in fig.4A) of the 5' ends of c-myb mRNAs calculated from experiments using two independent probes (e.g. 5' boundaries at position 991-997 obtained from riboprobe 'P405' versus 995-999 from riboprobe 'BANco'), in four out of five cases suggests that the protected fragments correspond to genuine c-myb 5' ends, (see fig.1 for summary).

Additional experiments in which RNA from FDC-P1 cells (which also express high levels of c-myb mRNA) was used, gave a slightly different pattern of 5' end sites. Multiple sites were seen at positions 976 to 1137 with riboprobe 'P405', (lane 1 fig.4B), and four more discrete sites, which were further 5' at positions 729 to 785 were generated with riboprobe 'P570', lane 4. The intensities of the bands corresponding to the furthest 5' ends were lower than those of the sites corresponding to the shorter mRNA species, suggesting that the longer species are less abundant than the shorter ones. Their authenticity is supported by the failure of mRNA from P3K cells (that do not express c-myb) to generate fragments of a similar size to those obtained with FDC-P1 RNA (lanes 2 and 5). The slightly different pattern of 5' end boundaries utilized in FDC-P1 and WEHI3B(D+) cells probably represents cell-specific variation in the selection of cap sites.

sites are indicated by stars. There is approximately 3.5 kbp HindIII to HindIII segment of unsequenced DNA covering first intron. The arbitrary numbering of the 5' segment of the 1.4 kbp HindIII to EcoRI genomic fragment containing the second exon of c-myb starts at the first nucleotide of the HindIII recognition site as nucleotide 2001. The position of oligonucleotides 'myb #2' and 'myb #3' are indicated by arrows.
Functional Activity of the 5' End of c-myb Gene

The data presented above indicate that a 2.4 kbp HindIII fragment of λmyb1 includes a number of sites where c-myb transcription is initiated in vivo. Since it was considered possible that the expression of c-myb might be regulated at the level of transcriptional initiation and that the sequences responsible for regulation are present in the 5' region of the gene, as exemplified in many other eukaryotic transcription units, a promoter assay that employs a readily assayable enzymatic function has been used in a transient expression system in vitro.

Two groups of vectors and derivatives were used in this study which utilize bacterial CAT as a reporter gene; the original CAT vector, pSV2CAT [30] includes a bacterial origin of DNA replication and the bacterial CAT gene under the control of the SV40 early promoter. pSV0CAT is derived from pSV2CAT and lacks the SV40 72 and 21 bp repeats of the SV40 promoter/enhancer as well as the TATA box. Thus pSV0CAT serves as a vector for assaying both promoter and promoter/enhancer activity. The second vector pUCAT2 [31] containing the SV40 21 bp repeats and the TATA box but lacking the 72 bp enhancer repeats serves in assays designed to detect sequences with enhancer activity capable of complementing the SV40 promoter.

In order to test whether the region of the gene that included the major cap sites contained a functional promoter, the 2.4 kbp HindIII fragment of λmyb1 (containing the 1st exon and 1.5 kbp upstream of translational initiation codon) was inserted into pSV0CAT in the correct (pSVm1CAT) and inverse (pSVm2CAT) transcriptional orientations (fig.5). These constructs, as well as pSV0CAT alone, were transfected into NIH3T3 cells and the level of CAT activity determined after 48 - 60h. As shown in fig.6A only the pSVm1CAT (lane 3) displayed a slight, but statistically significant (P = 0.009; statistical analysis using one-tailed, two sample, T-test), elevation of activity compared to the negative control, pSV0CAT vectors (lane 2). By contrast the pSVm2CAT plasmid (lane 4) which represents the same 2.4 kbp fragment cloned in the opposite orientation shows the same level of activity as pSV0CAT.

The 2.4 kbp c-myb HindIII fragment of λmyb1 contains two genetic elements present in the first exon which may interfere with efficient expression of the CAT gene resulting in artificially low levels of CAT activity. The splice donor site could interfere with processing of the CAT mRNA, while the protein synthesis initiation codon (ATG) could interfere at the translational level. In order to remove both elements from pSVm1CAT, two constructs were generated by deleting the 3' end of HindIII fragment. Deletion between the NcoI site at position 1205 (coinciding with the ATG start codon) and the 3' HindHI site led to removal of the splice donor site, generating the plasmid designated pSVm1CATΔNco (fig.5). An extended deletion from the BstEII site, at position 1047 to the 3' HindIII site removed the translational initiation codon as well as the splice donor

FIG.3. PRIMER EXTENSION ANALYSIS

Oligonucleotide primers "jul #1", "jul #2" and "myb #2" complementary to sections of the First and second exons (figs 1 and 2) were hybridized to cytoplasmic mRNAs from NIH3T3, WEHI3B(D+) and RB22.2 cell lines and extended with reverse transcriptase as described in materials and methods.

Lane 1. M.W. markers (pBR322/HinfI) in nucleotides; lane 2. NIH3T3 poly(A)+ RNA/"jul #1" primer; lane 3. WEHI3B(D+) poly(A)+ RNA/"jul #1"; lane 4. RB22.2 poly(A)+ RNA/"jul #1"; lane 5. NIH3T3 poly(A)+ RNA/"jul #2"; lane 6. WEHI3B(D+) poly(A)+ RNA/"jul #2"; lane 7. RB22.2 poly(A)+ RNA/"jul #2"; lane 8. NIH3T3/"myb #2"; lane 9. WEHI3B(D+) poly(A)+ RNA/"myb #2"; lanes 10. to 12. Sequencing ladder (A and G mix, C and T) of an unrelated fragment of DNA as a single nucleotide size marker.
FIG. 4 RNase PROTECTION ANALYSIS

A. Polyadenylated RNA from WEHI3B(D+) was hybridized with a molar excess of complementary radiolabeled RNA, followed by digestion with RNase A and RNase T, fractionated on an 8% acrylamide sequencing gel and established as an autoradiograph. Lanes: 1. and 2. different amounts of M. W. markers; lane 3. 10 μg of WEHI3B(D+) mRNA/'P405' probe; lane 4. 5 μg of WEHI3B(D+) mRNA/'P405'; lane 5. 1 μg of WEHI3B(D+) mRNA/'P405'; lane 6. 10 μg of tRNA/'P570'; lane 7. as in lane 3 but no RNases added; lane 8. 10 μg of WEHI3B(D+) mRNA/'ΔNco'; lane 9. 5 μg of WEHI3B(D+) mRNA/'ΔNco'; lane 10. 1 μg of WEHI3B(D+) mRNA/'ΔNco'.

B. Polyadenylated RNAs from FDC-P1 and P3K cells were hybridized with an excess of complementary radiolabeled RNA, followed by digestion with RNase A and RNase T, resolved on an 8% acrylamide sequencing gel and established as an autoradiograph.

Lane: 1. 10 μg FDC-P1 mRNA/'P405' probe; lane 2. 10 μg P3K mRNA/'P405'; lane 3. 10 μg tRNA/'P405'; lane 4. 10 μg FDC-P1 mRNA/'P570'; lane 5. 10 μg P3K mRNA/'P570'; lane 6. 10 μg tRNA/'P570';
site (pSVm1CATΔBst). Both constructs, along with a number of control plasmids, were transfected in triplicate into NIH3T3 cells and the level of CAT activity determined 48-60 h post-transfection. As revealed in fig.6B, both pSVm1CATΔNco and pSVm1CATΔBst (lanes 5 and 6), showed an approximately 3-fold increase in CAT activity compared with the promoter-less control pUCAT0 (P=0.016 and P=0.014 respectively), which was comparable to the activity observed in cells transfected with the vector pUCAT2 that contains the SV40 promoter region without the enhancer element. The elevated level of CAT activity is however, considerably less than that achieved using the same reporter construct containing the polyoma enhancer element (lane 2, fig.6B). These results suggest that the 1.25 kbp c-myb fragment that lies between the 5' HindIII and BstEII sites contains a functional promoter region permitting relatively low levels of expression in NIH3T3 cells.

In order to directly address the question of whether a transcriptional regulatory element of the enhancer type exists in the 5' end of c-myb, plasmid pUCAT2 was employed. Transfection of a number of constructs (myb/pUCAT2; see fig.5) into NIH3T3 cells and quantitation of CAT activity in the cell extracts, failed however to produce a significant elevation of enzyme activity (data not shown). Therefore, it was concluded that there was no transcriptional stimulatory element capable of complementing the enhancer-dependent SV40 promoter in the 5' fragment (up to approximately 3.5 kbp upstream of the first ATG) of c-myb that functioned in NIH3T3 cells.

Taken together, these results suggest that the 5' region of c-myb contains promoter elements that have a low level of transcriptional activity in NIH3T3 cells. There was no evidence for any enhancer activity in this region that functions in NIH3T3 cells, although it should be stressed that endogenous c-myb mRNAs have not been detected in NIH3T3 cells. Accordingly, these experiments might not register a promoter/enhancer that might only be active in the particular cell types that express c-myb. In an attempt to approach this question similar experiments were undertaken in hemopoietic cells that express the c-myb gene.

Transcriptional Activity of the 5' End of c-myb in Hemopoietic Cells

In order to test the possibility that expression of c-myb in hemopoietic cells is regulated at the level of transcriptional initiation, transfection of various myb-reporter gene constructs into cells expressing high levels of c-myb mRNA was undertaken. In preliminary experiments severe difficulties were encountered in attempts to transfect hemopoietic cell lines by calcium phosphate precipitation. Accordingly, two cell lines (RB22.2 and FDC-P1 [32]) were chosen on the basis of their ability to take up DNA by electroporation. When pSVm1CAT was electroporated into RB22.2 cells (lane 4, fig.7) a small (about 2 fold) elevation of CAT activity was observed compared to that obtained with the pUCAT0 vector alone (lane 2) and which was marginally greater than the activity of pUCAT2, (lane 3). However, no increase in the level of CAT expression was observed when pSVm2CAT (lane 5) containing the c-myb fragment in the opposite transcriptional polarity was used. These results, like those obtained in NIH3T3 cells, suggested that the 2.4 kbp HindIII fragment contains a weak promoter element leading to a low level of expression. Similarly, the derivatives of pSVm1CAT containing deletions of the splice donor and translational start codon, pSVm1ΔNco and pSVm1ΔBst (lanes 6 and 7, fig.7), also showed only a modest elevation of activity, again comparable to that of pUCAT2.

As before, a second vector was utilized in order to directly address the question of potential enhancer activity in this region. Inserting the 1.1 kbp c-myb BamHI and 3.5 kbp NcoI fragments into the vector pUCAT2 generated plasmids pUCAT2m21, pUCAT2m22 and
pUCAT2mN3.5,5', (fig.5). When these constructs were electroporated into RB22.2 cells, a small elevation of CAT activity was observed in the case of pUCAT2m21 compared to that obtained with pUCAT2 (fig.7 lanes 8 and 3). By contrast, pUCAT2m22 (lane 9) containing the 1.1 kbp BamHI fragment in the inverted orientation showed a level of activity lower than that of pUCAT2. Similarly the pUCAT2m21 derivatives lacking the splice donor site and the translation initiation codon, pUCATm21ΔNco and pUCATm21ΔBst (lanes 11 and 12, respectively, fig.7), as well as a plasmid containing a longer, 3.5 kbp NcoI myb fragment (lane 10) showed levels of CAT activity above the background represented by pUCATO (lane 2). In each case, more activity was obtained with vectors containing fragments in the correct, as compared to the inverse transcriptional orientation. These values were again comparable to that achieved with the pUCAT2 construct (lane 3, fig.7), suggesting that the c-myb fragments tested do not contain an enhancer element that can complement the SV40 promoter in RB22.2 cells. Similar results were obtained in experiments with another hemopoietic line FDC-P1 (data not shown).

In summary, the results presented here suggest that the 5' region of the murine c-myb gene contains some promoter activity detectable in hemopoietic cells (RB22.2 and FDC-P1) as well as in fibroblastic cells (NIH3T3), indicating that this promoter region may function in both hemopoietic and non-hemopoietic cells. There was no evidence for any enhancer-like activity within first 3.5 kbp 5' from the region of multiple initiation sites of the c-myb gene in the three cell lines tested.

DISCUSSION

The nucleotide sequence and structure of the 5' region of the murine c-myb gene determined in the present study, while substantially in agreement with that reported by Bender and Kuehl [27] and Watson et al. [33], differs at 6 nucleotide positions as shown in fig.2. The published nucleotide sequences correspond to positions 140–1279 and 321–1231, respectively of the sequence generated in the present study. While the differences may represent errors in sequencing due to special difficulties in sequencing GC rich DNA it is more likely that they correspond to polymorphic differences (somatic or non-somatic) between the different strains of mice used to prepare the genomic libraries.

Two well defined structural elements, 'TATA' and 'CCAAT', are features of many eukaryotic promoters. These DNA elements are usually located 25–30 and 40–110 residues respectively upstream of the transcriptional start site. A search of the 5' region in the c-myb DNA sequence revealed the presence of one such element. The sequence at position 994—TAATAT— displays a high level of homology to the consensus sequence

FIG.5. Structure of plasmids carrying the 5' region of the murine c-myb gene coupled to the SV40/chloramphenicol acetyltransferase gene. Top lines represent a 5' fragment of the gene with the first five exons marked. Bars above and below indicate the 2.4 kbp HindIII (H), 1.1 kbp BamHI (B), 3.5 kbp NcoI (N), 1.8 kbp HindIII to NcoI (H3ΔNco), 1.6 kbp HindIII to BstEII (H3ΔBst), 1.0 kbp BamHI to NcoI (BΔNco) and 0.9 kbp BamHI to BstEII (BΔBst) restriction fragments spanning the first exon and upstream sequences; the 1.7 BamHI fragment comes from the first c-myb intron. Fragments cloned into the HindIII site of pSVOCAT or pUCAT2: m1 = pSVm1CAT; m2 = pSVm2CAT; m1ΔNco = pSVm1CATΔNco; m1ΔBst = pSVm1CATΔBst; m21 = pUCAT2m21, m22 = pUCAT2m22, m23 = pUCAT2m23, m24 = pUCAT2m24, mN3.5,5' = pUCAT2mN3.5,5'. Py = represents a 240 bp fragments that includes the polyoma viral enhancer element inserted in the correct (pUCAT2Py1) and inverse (pUCAT2Py2) transcriptional orientation. The body of the CAT gene is indicated by an open box and 5' untranslated region by an open dashed box. The NcoI restriction site in c-myb coincides with translation initiation codon (ATG).
However, as discussed below, this motif does not correlate with any of the cap sites detected in the primer extension analysis.

Inspection of the 2.4 kbp HindIII fragment of cmyb1 has revealed an unusually high content of G and C nucleotides. A moving average value for percentage G+C content is plotted in fig.8A. Two regions at position 580—905 and position 980—1300 show a high level of G+C content of 76.7% and 69.2%, respectively. Furthermore the ratio of the observed to the expected occurrence of the dinucleotide CpG, (as shown in fig.8B) in those two regions is characteristic of ‘CpG islands’ [35]. Almost all CpG islands identified to date have been associated with the 5’ ends of genes [36]. Some of the G-C rich regions associated with several eukaryotic genes include binding sites for the transcription factor Sp1 [37]. A number of potential binding sites within the 5’ flanking sequences of the c-myb gene that have 80% or greater homology to the consensus binding sequence of Sp1 (GCGGCCTT) are located within two regions (nucleotides 715—950 and 1060—1230) that also correlate with majority of the 5’ ends of c-myb mRNA, as shown in fig.2.

An interesting 120 bp region of the c-myb gene lying 5’ of the translational initiation site (nucleotides 901—1022, fig. 2) is conserved (92% identity) between chicken and mouse [38]. This homology, that includes the most 5’ cap sites detected in this study is therefore likely to be of some importance to c-myb regulation.

Riboprotection and primer extension analysis leads to the conclusion that the c-myb gene contains multiple transcriptional initiation sites. Four cap sites at position 1000, 1069, 1103 and 1145 which were identified using both assays (in the majority of cases using two different priming oligonucleotides or two independent riboprobes), are likely to represent genuine transcriptional initiation sites for the c-myb gene in WEHI3B(D+) cells (see fig.1 for summary). Moreover, multiple cap sites were also identified in two other cell lines, (FDC-P1 and RB22.2); in some cases the major cap sites were the same as these detected in WEHI3B(D+) while other cap sites were unique to each cell line.

In addition to the four major transcriptional initiation sites detected in this study, several minor cap sites have been detected using the RNase protection assay. Utilization of some of the minor cap sites might account for the larger, less abundant 4.2 kbp c-myb mRNA observed by some investigators [39—41], since these sites are approximately 280—350 bp 5’ of the major cap sites whose utilization presumably leads to the synthesis of the major 3.9 kb mRNA. This interpretation is, however, complicated by the observation that at least some 4.2 kb c-myb mRNA molecules include an additional exon as a result of alternative splicing within the coding region of the gene [42]. Alternatively, the minor ‘cap’ sites detected in the present study may correspond to sites located within a putative upstream exon(s) that is subject to differential splicing. Indeed, a comparison of some of the ‘cap’ sites detected only by the riboprotection procedure with potential splice acceptor sites [43], reveals that in two cases, apparent ‘cap’ sites at positions 1043—51 and 1087—91 (detected in FDC-P1 mRNA) do in fact match the positions of potential 5’ splice acceptors.
FIG. 7. Transient expression of chloramphenicol acetyl transferase in RB22.2 cells electroporated with plasmid DNA: pUCAT2Py (lane 1); pUCAT0 (lane 2); pUCAT2 (lane 3); pSVm1CAT (lane 4); pSVm2CAT (lane 5); pSVm1CATΔNco (lane 6); pSVm1CATΔBst (lane 7); pUCAT2m21 (lane 8); pUCAT2m22 (lane 9); pUCAT2mN3.5,5' (lane 10); pUCAT2m21ΔNco (lane 11); pUCAT2m21ΔBst (lane 12) and mock DNA (lane 13). CAT assay positive control, 1.0 u of CAT enzyme (lane 14). Electroporation and CAT assays were carried out as described in materials and methods. The bars represent standard deviations calculated from the triplicate samples.

at position 1046 and 1091, respectively. The possibility of differential splicing is further supported by the observation that two chicken cDNA clones [44,45] have different 5' ends, although it has not been possible to identify genomic sequences corresponding to the 5' end of the cDNA sequence of Rosson and Reddy [44] within 8.0 kbp upstream of the
FIG. 8. G+C CONTENT IN 5' REGION OF MURINE c-myb GENE.
A: a moving average value for the percentage of G+C calculated using a 51 nucleotides window (N=51) moving across the sequence at 1 nucleotide intervals.
B: analysis of the distribution of CpG dinucleotides in the 5' c-myb region; note, that the first and last 51 nucleotides in this and top panels are blank, due to the size of the window used in the calculations.
The ratio observed/expected CpG was calculated according to the formula:

$$\frac{\text{Obs}}{\text{Exp}} \times \frac{\text{number of CpG}}{\text{number of C} \times \text{number of G}} \times N$$

where N is the total number of nucleotides in the analysed sequence [35].

ATG [46]. The possible existence of additional upstream exons and differential splicing may be further explored by attempting to isolate murine cDNA clones that encompass sequences in the 5' region of the c-myb gene.

While the numerous 'cap' sites utilized in c-myb transcription in a murine pre-B cell leukemia [27], WEHI3B, murine thymocytes and myeloma Ag-8 [33] lie in the same upstream region (between nucleotides 861 and 1108) none of 13 putative cap sites defined
by Bender and Kuehl match precisely the 15 putative cap sites established by Watson et al., nor do they match many of the cap sites presented here. The cap sites identified in this study at positions 1069 and 1103 probably correspond to those found by Watson et al., at positions 1067 and 1108, respectively; the few nucleotides difference may represent inaccuracy in determining the size of DNA or RNA fragments on acrylamide gels. Intriguingly, these sites were determined using RNA from the same myelomonocytic cell line, WEHI3B(D+), used in one of the published studies [33] which also utilized S-1 or riboprotection assays to determine the 5' ends of c-myb mRNAs. It is unclear what the basis of the apparently different initiation sites may be but it seems likely that discrepancies in c-myb 'cap' sites within the same cell line simply reflect differences between populations of cells that have been passaged in different laboratories over long periods of time.

Evidence has been presented here that a c-myb DNA fragment containing the first exon and at least 800 bp further upstream from the splice donor site contains a biologically active promoter(s). The activity of this upstream region appears to be similar in hemopoietic cells (RB22.2 and FDC-P1) and murine fibroblasts (NIH3T3). The similar levels of CAT activity that we have observed in cells of hemopoietic and fibroblastic origin is consistent with nuclear run-on assays in which the transcription rate of c-myb exon 1 has been shown to be equal in a number of different hemopoietic cells as well as murine fibroblasts [47], indicating that an important mode of regulation of transcription of the c-myb gene is at the level of elongation. Apparently, sequences located within the body of the c-myb gene serve to attenuate transcription in certain cells, while in other cells they appear inoperative [47,48]. Since the transfected promoter/reporter gene constructs used in the present study lack those sequences that are responsible for the block in elongation it seems likely that this level of regulation is lost leading to the accumulation of mRNAs and associated basal levels of CAT activity.

Finally, it should be noted that other post-transcriptional initiation events such as capping, splicing, transport of myb RNAs to the cytoplasm, and degradation might also serve to regulate the levels of steady-state c-myb mRNAs and protein. A detailed understanding of the contribution that each of these processes make will be necessary to provide a full understanding of the regulation of expression of the c-myb proto-oncogene in normal cellular growth and oncogenic transformation.

ACKNOWLEDGEMENTS
We thank Julie King for technical assistance, Greg Thege and Marina Spaulding for running many computer programs for DNA analysis and Pierre Smith and Alan May for their assistance with artwork.

*To whom correspondence should be addressed at Centre for Early Human Development, Monash Medical Centre, 246 Clayton Rd., Clayton, Victoria 3168, Australia

REFERENCES