Structural organization and nucleotide sequence of mouse c-myb oncogene: activation in ABPL tumors is due to viral integration in an intron which results in the deletion of the 5' coding sequences

Sukadev Lavu and E.Premkumar Reddy*

Department of Molecular Oncology, Roche Institute of Molecular Biology, Roche Research Center, Nutley, NJ 07110, USA

Received 27 February 1986; Revised and Accepted 11 June 1986

ABSTRACT

Bacteriophage libraries of mouse DNA were screened for sequences homologous to the v-myb oncogene and two overlapping clones containing the v-myb related region were isloated. Restriction enzyme mapping, heteroduplex analysis and nucleotide sequence analysis revealed the presence of nine exons. Six of these exons are homologous to the v-myb region while the other three exons are derived from the 5' region which is deleted in the viral oncogene. The sequences downstream to the sixth v-myb exon are not included in the 17 kbp of DNA sequences analyzed in this study. Comparison of the structure of the normal c-myb clone with its rearranged counterpart present in plasmacytoid lymphosarcomas revealed that the rearrangements occur in this locus as a result of viral integration. Present studies demonstrate that such a viral insertion interrupts the c-myb coding region at a region identical to that observed in the generation of the v-myb gene of avian myeloblastosis virus and results in the synthesis of mRNAs that lack the same 5' coding region.

INTRODUCTION

Avian myeloblastosis virus (AMV) induces myeloblastic leukemia in chickens and transforms a specific class of hematopoietic cells in vitro (1,2). However, it does not induce morphological transformation of cultured fibroblasts suggesting that only a restricted target cell population is responsive to its onc gene product (1,2). Like all other acute transforming viruses, AMV has arisen by recombination between helper viral and host cellular sequences. The latter sequences (termed v-myb) are responsible for the oncogenic properties of this virus. The cellular counterpart of this viral oncogene (c-myb) is highly conserved and is present in all vertebrate, and some invertebrate, species examined (3,4). During the past five years many studies have indicated that proto-oncogenes are associated with cell growth and differentiation and that structural or regulatory alterations of these genes lead to the development of neoplasia (5-11). In the case of c-myb, DNA rearrangements and altered expression have been reported in mouse and human lymphoid tumors (12-15). Characterization of the rearranged myb loci in four independently derived BALB/c plasmacytoid lymphosarcomas (ABPLs) revealed that the rearrangements are due to insertion of a defective Moloney murine leukemia virus (M-MuLV) at the 5' end of the v-myb-related sequences (13,14). This retroviral insertion

[©] IRL Press Limited, Oxford, England.

is associated with the abnormal transcription of <u>myb</u> sequences and probably represents a step in the neoplastic transformation of ABPL cells. This observation poses several important questions. 1) Does the viral integration occur within the coding region of the c-<u>myb</u> gene or is it immediately upstream of the c-<u>myb</u> coding domain? 2) Does the viral integration result in an alteration of the coding potential of the rearranged c-<u>myb</u> gene? 3) Does the viral integration result in the production of an aberrant gene product where specific coding sequences have been deleted? To answer these question, we undertook a comparative molecular characterization of the normal and aberrant c-<u>myb</u> loci and their transcriptional products.

MATERIALS AND METHODS

DNA probes.

The v-myb probe was derived from the AMV proviral DNA (16). The v-myb sequences located between the KpnI and XbaI sites were subcloned in pBR322 using BamHI linkers. This subclone (termed KX-1) contained the entire v-myb specific region and was used to detect the homologous sequences in mouse DNA. The molecular cloning of chicken c-myb sequences was achieved by reverse transcription of mRNA isolated from chicken thymus and molecular cloning of the cDNA in pBR322 or λ gt10 vectors (17). The subclone K555 contained a 500 bp insert from the 5' end of the c-myb gene that is deleted in the v-myb gene.

Molecular cloning of the mouse c-myb sequences.

It was shown earlier (13,14) that hybridization of EcoRI digested mouse genomic DNA with v-myb (KX-1) probe revealed three restriction fragments of 4.2 kbp, 1.7 kbp and 1.4 kbp, respectively. To isolate molecular clones that contained these sequences, the subgenomic v-myb probe (KX-1) was used to screen a bacteriophage library of mouse DNA (partial EcoRI* digest) made using charon 4A vectors (kindly provided by B.K. Birshtein and L.A. Echardt, Albert Einstein College of Medicine). Six clones of similar organization were obtained and further characterization of one of them (MM1) revealed that it contained a 16 kbp insert. This clone encompassed the 4.2 kbp and 1.7 kbp v-myb related sequences but only 500 bp of the 1.4 kbp fragment. To obtain molecular clones of the latter region, normal mouse DNA was digested with EcoRI and fractionated on a preparative agarose gel. The DNA fraction containing the 1.4 kbp fragment was isolated by electroelution and cloned in a λ gt10 vector.

DNA sequence analysis.

The nucleotide sequence analysis was carried out by the procedure described by Maxam and Gilbert (18).

Heteroduplex analysis.

Equal amounts of DNAs (1 µg) were mixed, denatured with 30 mM NaOH,

neutralized with 0.1 M Tris- HCl (pH 8.0) and allowed to hybridize in the presence of 40% formamide at 25°C for 2 h. Samples were spread onto a distilled water hypophase and prepared for electronmicroscopy according to the method of Davis et al. (19). Uranyl acetate stained grids were rotary shadowed with platinum/palladium and examined in a Phillips 600 electron microscope.

Northern blot analysis.

RNA extractions and hybridization studies were as described by Mushinski et al. (12). To detect the presence of v-<u>myb</u> related sequences, a 1.1 kbp <u>XbaI</u> fragment containing exons 1 and 2 was nick-translated and used as a probe. To detect the presence of upstream sequences (termed UE2 and UE3) present towards the 5' end of the c-<u>myb</u> clone and absent in the v-<u>myb</u> gene, a 1.5 kbp <u>BstEII</u> fragment containing UE3 sequences and a 1.5 kbp <u>HindIII</u> fragment containing UE2 sequences were nick-translated, combined and used as a probe. The latter probe is referred to as the upstream exon probe.

RESULTS

Isolation of v-myb related clones from phage libraries of mouse DNA.

The subgenomic v-myb probe (KX1) was used to screen two bacteriophage libraries of mouse DNA. The first library was constructed using partial <u>Eco</u>RI* digested mouse DNA ligated into the charon 4A vector. Six identical clones were obtained from this library and further characterization of one of them (λ MM1) revealed that it contained a 16 kbp insert. A detailed restriction map of this clone was developed (Fig. 1). Digestion of the insert DNA with <u>Eco</u>RI produced six fragments which were 7.5 kbp, 4.2 kbp, 1.8 kbp, 1.7 kbp, 0.5 kbp and 0.4 kbp long. Hybridization with the v-myb specific probe (KX-1) revealed that the 4.2, 1.7 and 0.5 kbp fragments contained v-myb related sequences. Further analysis revealed that the λ MM1 clone lacked the 3' terminal sequences since it contained only a third of the 1.4 kbp EcoRI fragment detected by



Fig. 1. Restriction enzyme map of mouse c-<u>myb</u> genomic DNA. The extent of the two phage clones insolated from the mouse DNA libraries is indicated by the lines at the bottom. v-<u>myb</u> related regions are inidcated by solid boxes and upstream exons are indicated by striped boxes. The position of homologous regions was deduced by heteroduplex and nucleotide sequence analyses. E, <u>EcoRI</u>; H, <u>HindIII</u>, B, <u>BamHI</u>; Bg, <u>BgIII</u>, Bs, <u>BstEII</u>; S, <u>SstI</u>; X, <u>XbaI</u>.



Fig. 2. Heteroduplex analysis of mouse c-<u>myb</u> DNA. Heteroduplexes formed between a pBR322 subclone of 11.0 kbp <u>Bam</u>HI fragment isolated from λ MM1 and a 1.3 kbp <u>KpnI-Xbal</u> fragment of AMV (KX I) which contains the entire v-<u>myb</u> gene are shown. A diagramatic representation of the heteroduplex is shown below the electronmicrograph. Contour length (in kbp) of features are as follows: (1) 4.0+0.2, (2) 0.110+0.028, (3) 0.320+0.04, (4) 0.250+0.028, (5) 0.180+0.032, (6) 0.550+0.06, (7) 0.3+0.02, (8) 4.5+0.04, (9) 0.05+0.02, (10) 1.64+0.2, (11) 1.9+0.2, (12) 0.06+0.02, (13) 1.08+0.2. The 4.0 kbp and 0.3 kbp stretches of double stranded sequences at the two ends of the heteroduplex correspond to the pBR322 DNA.

hybridization of mouse DNA with the v-myb probe (apparantly due to cleavage at an <u>Eco</u>RI* site). This 1.4 kbp fragment was cloned separately into a λ gt10 vector and further characterized by restriction mapping. The restriction map of this clone is also shown in Figure 1. Thus, the two clones λ MM1 and λ gt34-1 contained the entire v-myb related sequences present in mouse DNA.

| ccaccggactttgtctagA A | l ACCGGACAGATGTGCAGTGCCAACA snArgThrAspValGlnCysGlnHi | 10 EXON 1 CCGGTGGCAGAAAGTGCTGAA sArgTrpGInLysValleuAs | 20 CCCTGAACTCATCAAAGGTCC nProGluLeuIleLysGlyPr | TTGGACCAAAGAA oTrpThrLysGlu |
|---|---|--|---|---|
| 30 GAAGATCAGAGAgtaagtt GluAspGlnArg | ctgtcc(0.05 kbp)t | tcacttgaagGTCATAGAGCT VallleGluLe | 40 TGTCCAGAAATATGGTCCGAA uValGInLysTyrGlyProLy | EXON II GCGTTGGTCTGTT sArgTrpSerVal |
| 50 ATTGCCAAGCACTTAAAAG IleAlalysHisleulysG | 60 GGAGAATTGGAAAGCAGTGTCGGGA ilyArgIleGlyLysGlnCysArgGl | GAGGTGGCACAACCATTTGAA uArgTrpHisAsnHisLeuAs | 70 TCCAGAAGTTAAGAAAACCTC nProGluValLysLysThrSe | CTGGACAGAAGAG rTrpThrGluGlu |
| 80 GAGGACAGAATCATTTACC GluAspArgileileTyrG | 90 AGGCACACAAGCGTCTGGGGAACAG SInAlaHisLysArgLeuGlyAsnAr | 100 ATGGGCAGAGATCGCAAAGCT gTrpAlaGluIleAlaLysLe | GCTGCCCGGACGgtaagactt uLeuProGlyAr | gg(1.6 kbp) |
| ttttctacacagGACT6 gThrA | 110 EXON I GATAATGCTATCAAGAACCACTGGAA AspAsnAlaIleLysAsnHisTrpAs | II 120 TTCCACCATGCGTCGCAAGGT nSerThrMetArgArgLysVa | l GGAACAGGAAGGCTACCTGCA lGluGlnGluGlyTyrLeuGl | 130 NGGAGCCTTCCAAA InGluProSerLys |
| 140 GCCAGCCAGACGCCAGTGG AlaSerGInThrProVal |) GCCACGAGCTTCCAGAAGAACAATCA NlathrSerPheGlnLysAsnAsnHi | 150 ITTTGATGGGGTTTGGGCATGC sLeuMetGlyPheGlyHisAl | 160 CTCACCTCCATCTCAGCTCTC aSerProProSerGinLeuSc | CTCCAAGTGGCCAG erProSerGlyGln |
| 170 TCCTCCGTCAACAGCGAA1 SerSerValAsnSerGlul | 180 FATCCCTATTACCACATCGCCGAAGC FyrProTyrTyrHisIleAlaGluAl | ACAAAACgtaagccactcccg aGlnAsn | (1.9 kbp)ttttc1 | ttttcctggtcta |
| gATCTCCAGTCACGTTCCC lleSerSerHisValPrc | 190 EXON IV CTATCCTGTCGCATTGCATGTTAAT/ DTyrProValAlaLeuHisValAsnl | 200 ATAGTCAACGTCCCTCAGCCGG A lasnValProGlnProA | 210 CTGCGGCAGCCATCCAGgtga LaAlaAlaAlaIleGln | ngatcatctgacgg |
| g(0.1 kbp)0 | cttcacttgaaactctagAGACACT/ ArgHisTj | 22 NTAACGACGAAGACCCTGAGAA (rAsnAspG1uAspProG1uLy | O EXON V GGAAAAGCGAATAAAGGAGCI (sGlulysArgIlelysGlul) | 230 TGGAGTTGCTCCTG EuGluLeuleuleu |
| ATGTCAACAGAGAACGAG MetSerThrGluAsnGlul | 240 CTGAAGGGACAGCAGGCATTACCAg1 LeulysG1yG1nG1nA1aleuPro | tagctgtga(1.5 kbp) | caccccttccccagA(T | 250 CACAGAACCACACT hrGlnAsnHisThr |
| TGCAGCTACCCCGGGTGG CysSerTyrProGlyTrp | 260 EXON VI CACAGCACCTCCATTGTGGACCAGA HisSerThrSerIleValAspGInT | 270 CCAGACCTCATGGGGATAGTGC hrArgProHisGlyAspSerAl | ACCTGTTTCCTGTTTGGGAG aproValSerCysLeuGlyG | 280 AACACCATGCCACC luHisHisAlaThr |
| CCATCTCTGCCTGCAGAT ProSerLeuProAlaAsp | 290 CCCGGCTCCCTACCTGAAGAAAGTG ProG1ySerLeuProG1uG1uSerA' | 300 CCTCACCAGCAAGGTGCATGA1 I a SerProA1 a ArgCysMETI 1 | 310 [CGTCCACCAG66CACCATTC leValHisGlnGlyThrIleL | TGGACAATGTTAAG euAspAsnValLys |
| AACCTCTTAGAATTTGCA AsnleuleuG1uPheA1a | 320 330 GAAACACTCCAGTTTATAGATTCTg GluThrLeuGluPhelleAspSer | taagtagaattgctaaggca | | |

Fig. 3. Nucleotide sequence of the mouse c-myb locus. The nucleotide sequence of the coding region corresponding to the v-myb homologous region is shown. The sequence of exons is given in capital letters and of introns in lower case letters. The amino acid sequence deduced from the open reading frame is given below the nucleotide sequence.

Organization of myb-related sequences within the mouse c-myb domain.

To define the organization of <u>myb</u>-related sequences in mouse DNA, heteroduplex analysis was carried out. For these experiments, a 11.0 kbp <u>Bam</u>HI fragment of λ MM1 was subcloned in pBR322. The resulting plasmid was then linearized within the vector region by digestion with <u>Cla</u>I, annealed to the 1.3 kbp v-<u>myb</u> clone (KX1) derived from AMV and examined under an electronmicroscope. Such analysis (Fig. 2) revealed that the two DNA molecules shared five regions of homology of approximately 100, 200, 250, 80 and 100 bp interrupted by approximately 200, 1600, 1900 and 300 bp long intron sequences. The region of homology commenced approximately 100 bp to the right of <u>KpnI</u> site of AMV. Thus, approximately 700 bps of sequence related to v-<u>myb</u> are present within the

NRTDVQCQHR WQKVLNPELI KGPWTKEEDQ RVIELVQKYG PKRWSVIAKH c-<u>myb</u> (mouse) NRTDVQCQHR WQKVLNPELN KGPWTKEEDQ RVIEHVQKYG PKRWSDIAKH v-<u>myb</u> C-MYD (Chicken) NRTDVQCQHR WQKVLNPELI KGPWTKEEDQ RVIELVQKYG PKRWSVIAKH c-<u>myb</u> (mouse) LKGRIGKQCR ERWHNHLNPE VKKTSWTEEE DRIIYQAHKR LGNRWAEIAK LKGRIGKQCR ERWHNHLNPE VKKTSWTEEE DRIIYQAHKR LGNRWAEIAK v-myb c-<u>myb</u> (chicken) LKGRIGKQCR ERWHNHLNPE VKKTSWTEEE DRIIYQAHKR LGNRWAEIAK C-<u>myb</u> (mouse) LLPGRTDNAI KNHWNSTMRR KVEQEGYLQE PSKASQTPVA TSFQKNNHLM LLPGRTDNAV KNHWNSTMRR KVEQEGYPQE SSKAGPPSAT TGFQKSSHLM v - <u>m.yb</u> c-myb (chicken) LLPGRTDNAI KNHWNSTMRR KVEQEGYLQE SSKAGLPSAT TGFQKSSHLM c-<u>myb</u> (mouse) GFGHASPPSQ LSPSGQSSVN SEYPYYHIAE AQNISSHVPY PVALHVNIVN AFAHNPPAGP LPGAGQAPLG SDYPYYHIAE PQNVPGQIPY PVALHINIIN v-myb C-MYD (chicken) AFAHNPPAGP LPGAGQAPLG SDYPYYHIAE PQNVPGQIPY PVALHVNIVN C-<u>myb</u> (mouse) VPQPAAAAIQ RHYNDEDPEK EKRIKELELL LMSTENELKG QQALPTQNHT VPQPAAAAIQ RHYTDEDPEK EKRIKELELL LMSTENELKG QQALPTQNHT v-myb c-myb (chicken) VPQPAAAAIQ RHYNDEDPEK EKRIKELELL LMSTENELKG QQALPTQNHT c-myb (mouse) CSYPGWHSTS IVDQTRPHGD SAPVSCLGE- HHATPSLPAD PGSLPEESAS ANYPGWHSTT VADNTRTSGD NAPVSCLGEH HHCTPSPPVD HGCLPEESAS v-mvb c-myb (chicken) ANYPGWHSTT VADNTRTSGD NAPVSCLGEH HHCTPSPPVD HGCLPEESAS c-myb (mouse) PARCMIVHQG TILDNVKNLL EFAETLQFIDS PARCMIVHQS NILDNVKNLL EFAETLQLIDS v-ayb c-myb (chicken) PARCMIVHQS NILDNVKNLL EFAETLQLIDS

Fig. 4. Homology between the mouse $c-\underline{myb}$ and the chicken $v-\underline{myb}$ and $c-\underline{myb}$ gene products. The symbol * indicates non-homology between the two adjacent sequences. The amino acid sequence of the chicken $v-\underline{myb}$ and $c-\underline{myb}$ sequences is derived from Rushlow et al. (20) and Klempnauer et al. (21).

11.0 kbp <u>Bam</u>HI subclone. Nucleotide sequence analysis revealed that most of the remainder of the v-<u>myb</u> related sequences lie to the right of <u>Bam</u>HI site within the 1.4 kbp <u>Eco</u>RI fragment of λ gt34-1 clone.

Nucleotide sequence of the mouse c-myb gene.

We identified and sequenced all the domains within the mouse $c-\underline{myb}$ locus that displayed homology with $v-\underline{myb}$. Figure 3 shows the nucleotide sequence of these domains. There are six domains of homology with $v-\underline{myb}$ which are similar to those identified in chicken $c-\underline{myb}$ sequences (20,21). The DNA clones described here do not contain sequences homologous to the 7th exon of $v-\underline{myb}$ (AMV) and the other downstream exons. The intron-exon boundaries could be readily identified by the consensus splice donor and acceptor sequences that are usually observed in these positions in eukaryotic DNA (as indicated in Figure 3). The predicted amino acid sequence is given below the



Fig. 5. Mechanism of activation of mouse c-<u>myb</u> locus by M-MuLV insertion in ABPL tumors and its structural similarity with the activation of the chicken <u>myb</u> locus as seen in avian myeloblastosis virus. Homologous regions between mouse and chicken cDNA sequences are indicated.

nucleotide sequence. The deduced amino acid sequence of the c-myb (mouse) gene product is compared with the c-myb (chicken) gene product in Figure 4. This comparison reveals certain areas of high conservation and other areas of considerable divergence. Thus, 100% conservation of the amino acid sequence was seen between chicken and mouse c-myb encoded proteins in exons 1, 2, and 5 while there was a divergence of 33%, 19% and 21% in exons 3, 4 and 6 respectively. It is possible that the conserved regions encode functional domains such as those involved in DNA binding or nuclear localization. Conversely, the divergent regions may confer a degree of species specificity for this protein that may be closely associated with gene regulation.

Identification of exons upstream of v-myb related regions.

During the process of recombination, both AMV and E26 viruses have has lost a substantial portion of coding regions of the c-myb gene both at the 5' and 3' ends. This is demonstrated schematically in Figure 5. Comparison of the v-myb sequences with those of a chicken c-myb cDNA revealed that approximaely 0.8 kb of the cDNA sequences are deleted from the 5' end, while a stretch of 2.0 kbp is deleted from the 3' end (17). As can be predicted, these changes affect the size of the proteins encoded by the v-myb and

Fig. 6. Nucleotide sequence of the mouse $c-\underline{myb}$ locus upstream to the v- \underline{myb} related region. The sequence of exons is given in capital letters and that of introns in lower case letters. Predicted amino acid sequence is presented below the nucleotide sequence. The ATG codon that could constitute the potential initiator codon is indicated in capital letters.

c-<u>myb</u> genes. Thus, the v-<u>myb</u> gene encodes a protein of 45 kd while the c-<u>myb</u> gene encodes a protein of 75 kd (22,23).

Based on supposed homology between chicken and mouse c-<u>myb</u> coding sequences, we used our cloned chicken c-<u>myb</u> sequences as a hybridization probe to identify mouse c-<u>myb</u> exons not included in v-<u>myb</u>. Further localization of these sequences was carried out by nucleotide sequence analysis which allowed us to position the upstream exons precisely within the λ MMI clone. The position of these exons is shown in Figure 5. An examination of the nucleotide sequence of these exons (Fig. 6) revealed the presence of an AUG codon in the first exon (UE3) which could function as an initiator methionine. However, in the absence of a complete cDNA clone derived from mouse c-<u>myb</u> mRNA, we cannot rule out the possibility that the reading frame extends even further upstream into as yet unidentified exons.

Deletion of upstream coding sequences (UE1-3) in aberrant c-myb mRNAs present in ABPL tumors.

It was demonstrated earlier that ABPL tumors exhibit rearrangements in the c-myb locus as a result of M-MuLV provirus insertion (12-14). Molecular cloning and characterization of these fragments revealed that in all these tumors the viral insertion occurred towards the 5' end of the v-myb related sequences (13,14). Sequence analysis reported here allows us to position the integration site between exons UE1 and 1, both of which constitute the coding domains of the c-myb gene (Fig. 5). The rearranged myb locus in ABPL tumors has been shown to code for aberrant mRNAs that are larger than the normal c-myb encoded mRNAS (12). It was of interest to see if these mRNAs contain upstream myb sequences that are deleted in v-myb. For this purpose two mouse c-myb probes were prepared; one containing exon 1 and exon 2 sequences (that are homologous to v-myb) and a second probe containing upstream exons 2 and 3 (UE2 and 3)



Fig. 7. Comparison of c-<u>myb</u> transcripts from normal mouse thymus and ABPL-2 tumor cell line. Lanes 1 and 3, RNA from ABPL tumors; Lanes 2 and 4, RNA isolated from normal mouse thymus. 10 μ g of RNA was used in all cases. The blots in panel A were hybridized with a nick translated probe derived from DNA fragments (see Material and Methods) containing sequences from UE2 and UE3 (Fig. 6). The blots in panel B were hybridized with a nick translated probe derived from a 1.1 kbp <u>Xbal</u> fragment containing sequences from 2 (Fig. 5). The blot in panel I was autoradiographed for 48 h while the blot in panels 2, 3 and 4 were autoradiographed for 16 h. The estimated molecular weight of the three major RNA species are indicated.

(Fig. 1) which constitute the 5' end of c-myb encoded mRNA but are absent in v-myb. Two separate RNA blots containing RNA from normal mouse thymus or from ABPL-2 tumor line were hybridized with these two probes. The results (Fig. 7) show that the RNA from normal mouse thymus hybridized well with both probes and two closely migrating bands in the region of 3.8 and 4.2 kb were seen with both these probes. On the other hand ABPL-2 RNA hybridized with the 3' probe (comprising exon 1 and exon 2) to reveal the presence of 3.8, 4.2 and 5.0 kb RNA species that were observed earlier in these cells using v-<u>myb</u> specific probe. However, the 5' probe (containing the upstream exons 2 and 3) failed to hybridize to the 5.0 kb mRNA present in these cells. Even the 3.8 and 4.2 kb RNA species fialed to hybridize with the same intensity as the normal mouse thymus RNA suggesting that only minor amounts of normal <u>myb</u> message is produced in these cells. These results demonstrate that the rearranged c-<u>myb</u> locus in ABPL tumors codes for mRNAs that lack the 5' exon sequences - a situation identical to that observed with AMV.

DISCUSSION

The present studies describe the detailed organization and structure of the mouse genomic DNA which include sequences related to the v-myb oncogene. The six exons that comprise the v-myb coding region have been identified. Several findings indicate that these exons do not account for the entire mouse c-myb proto-oncogene. The first v-myb-related exon lacks an initiator methionine codon. Moreover, the mouse c-myb transcript is known to be 4.0 Kb in length (12) while the v-myb related sequences are only 1.1 Kb (20,21). We have recently isolated a chicken c-myb c-DNA clone which includes the entire coding region (17). This clone contains sequences at both the 5' and 3' ends of the molecule that are not present in the v-myb gene. Using this clone as a probe we have isolated mouse genomic sequences corresponding to the 5' coding regions present in the chicken c-myb gene. While this work was in progress, Gonda et al. (24) have isolated a 2.8 Kb clone of mouse c-myb cDNA which also identifies the coding sequences present in the c-myb transcript. The ATG codon identified here corrsesponds to the proposed initiator codon in their cDNA sequence. In addition to these coding sequences, the c-myb encoded mRNA appears to contain at least 800 bases of non-coding sequences. In the absence of the complete cDNA sequence, it has not been possible to identify the upstream non-coding exons of the c-myb locus. The 17 kb of DNA sequences present in λ MM1 and λ gt 34-1 do not contain the 3' coding regions of the c-myb gene which are also absent in the viral oncogene.

One of the interesting observations that emerges from these studies pertains to the similarities in the mechanism of activation of c-<u>myb</u> proto-oncogene in avian and murine systems. In the avian system, this has been achieved by transduction of the <u>myb</u> oncogene into a retrovirus. Such a transduction resulted in the deletion of coding sequences from both the 5' and 3' ends of the gene. Initiation and terminator codons in helper viral sequences have been substituted for the analogous sequence in the proto-oncogene (Fig. 5). The important biological question that needs to be addressed is whether these deletions in coding sequences are responsible for the conversion of the

normal c-myb protein to its transforming counterpart. The availability of the murine model system allowed us to examine this question further. In the murine system, represented by the ABPL tumors, the activation of the myb locus occurred due to the integration of Moloney MuLV in a region which is immediately upstream to the v-myb related sequences (12-14). Sequence analysis presented here demonstrates that this integration occurred in the middle of the c-myb coding region at least three exons downstream of the normal translation-initiation site. Such an integration results in the synthesis of aberrant myb mRNAs that lack the 5' coding sequences present upstream of the viral integration site. These aberrant mRNAs are similar to that of the v-myb encoded mRNA in that they have suffered an identical deletion of the upstream protein coding sequences. Recently, Shen-ong et al. (25) have reached similar conclusions following the sequence analysis of one cDNA clone derived from the ABPL-2 tumor line.

This suggests that activation of the myb proto-oncogene is accomplished by exon deletion and/or replacement. It is possible that such deletions activate the oncogene by removing a regulatory domain. A similar activation may take place in the oncogene abl (26-28). The availability of c-myb proto-oncogene clones will allow direct analysis of the effects of deletions on the oncogenic activation of this gene.

ACKNOWLEDGEMENTS

We gratefully acknowledge A.M. Skalka and T. Curran for critical review of the manuscript.

*To whom correspondence should be addressed

REFERENCES

- Baluda, M.A. and Goetz, I.E. (1961) J. Virol. 15, 185-199. 1.
- 2. Moscovici, C. (1975) Curr. Top. Microbiol. Immunol. 71, 79-96.
- 3. Franchini, G., Wong-Staal, F., Baluda, M.A., Lengel, C. and Tronick, S.A. (1983) Proc. Natl. Acad. Sci. USA 80, 7385-7389.
- 4.
- Katzen, A.L., Kornberg, T.B. and Bishop, J.M. (1985) Cell. 41, 449-456. Doolittle, R.F., Hunkapiller, M.W., Hood, L.E., Devare, S.G., Robbins, K.C., 5. Aaronson, S.A. and Antoniades, H.N. (1983) Science 221, 275-277.
- 6. Waterfield, M.D., Scrace, G.T., Whittle, N., Stroobant, P., Johnson, A., Wasteson, A., Westermark, B., Heldin, C.H., Juang, J.S., and Deuel, T. (1983) Nature 304, 35-39.
- 7. Downward, J., Yarden, Y., Mayes, E., Scrace, G., Totty, N., Stockwell, P., Ullrich, A., Schlessinger, J. and Waterfield, M.D. (1984) Nature 307, 521-527.
- 8. Kelly, K., Cochran, B.H., Stiles, C.D. and Leder, P. (1983) Cell 35, 603-610.
- 9. Greenberg, M.E. and Ziff, E.B. (1984) Nature 311, 433-438.
- 10. Kruijer, W., Cooper, J.S., Hunter, T. and Verma, I.M. (1985) Nature 312, 711-716.
- Muller, R., Bravo, R., Burckhardt, J. and Curran, T. (1984) Nature 312, 716-720. 11.
- 12. Mushinski, J.F., Potter, M., Bauer, S. and Reddy, E.P. (1983) Science 220, 795-798.
- Shen-Ong, G.L.C., Potter, M., Mushinski, J.F., Lavu, S. and Reddy, E.P. (1984) Science 226, 1077-1080. 13.

- 14. Lavu, S, Mushinski, J.F., Shen-Ong, G.L.C., Potter, M. and Reddy, E.P. (1985) Cancer Cells 3, 301-306.
- 15. Pelicci, P-G., Lanfrancone, L., Brathwaite, M.D., Wolman, S.R. and Dalla-Favara, R. (1984) Science 224, 1117-1121.
- Souza. L.M., Strommer, J.N., Hillyard, R.L., Komaromy, M.C. and Baluda, M.A. (1980) Proc. Natl. Acad. Sci. USA 77, 5177-5181.
- 17. Rosson, D. and Reddy, E.P. (1986) Nature 319, 604-606.
- 18. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymology 65, 499-560.
- 19. Davis, R.W., Simon, M. and Davidson, N. (1971) Methods in Enzymology 21, 414-428.
- Rushlow, K.E., Lautenberger, J.A., Papas, T.S., Baluda, M.A., Perbal, B., Chirikjian, J.G. and Reddy, E.P. (1982) Science 216, 1421-1423.
- 21. Klempnauer, K.-H., Gonda, T.J. and Bishop, J.M. (1982) Cell 31, 453-463.
- Klempnauer, K.H., Ramsay, G., Bishop, J.M., Moscovici, M.G., Moscovici, C., McGrath, J.P. and Levinson, A.D. (1983) Cell 33, 345-355.
- 23. Boyle, W.J., Lipsik, J.S., Reddy, E.P. and Baluda, M.A. (1983) Proc. Natl. Acad. Sci. USA 80, 2834-2839.
- 24. Gonda, T.J., Grough, N.M., Dunn, A.R. and de Blaquieve, J. (1985) The EMBO J. 4, 2003-2008.
- 25. Shen-ong, G.L.C., Morse, H.C., Potter, M. and Mushinski, J.F. (1986) Mole. and Cell. Biol. 6, 380-392.
- Reddy, E.P., Smith, M.J. and Srinivasan, A. (1983) Proc. Natl. Acad. Sci. USA 80, 3623-3627.
- 27. Groffen, J., Stephenson, J.R., Heisterkamp, N., deKlein A., Bartram, C.R. and Grosfeld, G. (1984) Cell 36, 93-99.
- Shitvelman, E., Lifshitz, B., Gale, R.P. and Canaani, E. (1985) Nature 315, 550-554.