JOURNAL OF VIROLOGY, Apr. 1983, p. 212–220
0022-538X/83/040212-09$02.00/0

Vol. 46, No. 1

Structure and Transcription of the Cellular Homolog (c-*myb*) of the Avian Myeloblastosis Virus Transforming Gene (v-*myb*)

THOMAS J. GONDA†* AND J. MICHAEL BISHOP

Department of Microbiology and Immunology, University of California, San Francisco, California 94143

Received 4 November 1982/Accepted 17 January 1983

We isolated and characterized molecular clones containing the chicken cellular homolog (c-*myb*) of the avian myeloblastosis virus oncogene (v-*myb*). Mapping of the c-*myb* clones using restriction endonucleases and hybridization to radiolabeled v-*myb* probes revealed that the sequences homologous to v-*myb* are contained within four separate regions, which have since been shown by nucleotide sequencing (Klempnauer et al., Cell 31:453–463, 1982) to carry seven exons. Analysis of c-*myb* transcripts showed the presence of several large precursor RNAs in addition to the 4.0-kilobase cytoplasmic mRNA. We also determined the approximate positions of the c-*myb* sequences that are present in the 4.0-kilobase c-*myb* mRNA but not present in v-*myb*. Some of these sequences are found in a separate region 5′ to the v-*myb*-related sequences, whereas the remainder appear to be located immediately 3′ to the v-*myb*-related sequences. The data presented here, in conjunction with nucleotide sequence analysis (Klempnauer et al., Cell 31:453–463, 1982), indicate that the c-*myb* gene contains at least eight exons which span a total of about 16 kilobase pairs. The presence of exon sequences in c-*myb* outside the regions of homology with v-*myb* raises the possibility that the v-*myb* and c-*myb* gene products may differ significantly.

One of the major advances in the understanding of the molecular basis of malignant cellular transformation has been the discovery of a class of genes that are present in rapidly oncogenic retroviruses and also in normal cells (4). These genes, termed oncogenes, are responsible for the oncogenicity of the retroviruses which carry them, and their presence in normal cells raises two questions. (i) What is the role of the c-*onc* genes in normal cells? (ii) Are c-*onc* genes involved in tumors which are not caused by rapidly oncogenic retroviruses? (In this paper we will refer to viral oncogenes [in general] as v-*onc* genes and their cellular homologs as c-*onc* genes. Otherwise, the nomenclature of Coffin et al. [8] will be used.) Recent evidence suggests that the answer to the second question is positive (10, 18); however, the answer to the first is not at all clear at the moment. Studies of the transcriptional expression of c-*onc* genes (7, 13, 30) have suggested that one of these genes, c-*myb*, the cellular homolog of the avian myeloblastosis virus (AMV) oncogene (v-*myb*), may have a role in hemopoiesis. The c-*myb* gene is also transcribed in several human hemopoietic tumor cells (29), although there is no evidence to suggest that c-*myb* expression in these cells has any role in their transformation.

† Present address: Ludwig Institute for Cancer Research, PO Royal Melbourne Hospital, Victoria 3050, Australia.

Several studies have reported the presence of sequences homologous to v-*myb* in DNA from normal chickens (14, 22, 27) and from other vertebrates (2). In this report, we present an analysis of the overall structure of the chicken c-*myb* locus and a characterization of c-*myb* transcripts. The results show that as in most cellular genes, the coding sequences are interrupted by several large regions of DNA ("introns") which are transcribed but removed from the mature mRNA. Recently Perbal and Baluda have also reported the presence of at least one intron in the c-*myb* gene (20). We have also located sequences in the c-*myb* locus which are present in the mature c-*myb* message but not in the v-*myb* sequence.

The isolation of the c-*myb* clones described in this report has permitted a comparison of the nucleotide sequences of v-*myb* and the corresponding regions of c-*myb* (16). The observations presented here, taken in conjunction with this sequence analysis, suggest that there may be significant differences between the v-*myb* and c-*myb* gene products.

MATERIALS AND METHODS

Isolation of recombinant phage containing c-*myb* sequences. A library of chicken genomic DNA fragments in bacteriophage Charon 4A, prepared by partial digestion of the chicken DNA with *Alu*I and *Hae*III,

followed by attachment of *Eco*RI linkers, was kindly provided by J. Dodgson and J. Engel (11). The library was screened, and recombinant phage were isolated as described previously (17). The probe used was an AMV-specific cDNA ($cDNA_{myb}$; see below). Large-scale growth of purified phage was as follows. Bacteria (DP50 SupF) were grown in NZYDT medium (6) to an absorbance at 600 nm of 0.9, and 25 ml of the bacterial suspension was mixed with 2×10^9 PFU of phage in 25 ml of 0.1 M NaCl–0.01 M $MgCl_2$–0.01 M Tris-hydrochloride (pH 7.4). The mixture was added to 1 liter of NZYDT, and incubation with shaking was continued until lysis occurred (3 to 7 h). Phage were purified by polyethylene glycol precipitation, followed by two cycles of cesium chloride equilibrium centrifugation (as described by F. Blattner in the protocol which accompanies the Charon phages). DNA from the purified phage was isolated by proteinase K digestion in the presence of 0.5% sodium dodecyl sulfate, followed by phenol-chloroform extraction, and then was either dialyzed against 0.01 M Tris-hydrochloride (pH 7.4)–0.001 M EDTA or concentrated by ethanol precipitation.

All work with recombinant DNA was carried out according to National Institutes of Health guidelines.

Mapping of phage and genomic DNAs. Restriction endonucleases were purchased from New England Biolabs or Bethesda Research Laboratories and used as recommended by the suppliers. Restriction endonuclease digests were analyzed by agarose gel electrophoresis; after the gels were stained with ethidium bromide, DNA was transferred to nitrocellulose by the procedure of Southern (26) and analyzed by hybridization to $cDNA_{myb}$ (see below). Chromosomal DNA from a chicken lacking endogenous proviral sequences (15) was similarly analyzed.

Analysis of c-*myb* RNA. RNA was prepared from avian erythroblastosis virus (AEV)-transformed erythroblasts (kindly provided by T. Graf) and analyzed by formaldehyde-agarose gel electrophoresis, transfer to nitrocellulose, and hybridization to radiolabeled DNA probes, using procedures which have been described previously (13). Sizes of RNA species were estimated by comparison with rRNA markers, which were detected by ethidium bromide staining of the gels.

Nuclear and cytoplasmic fractions from AEV-transformed erythroblasts were prepared as follows. Cells were collected by centrifugation and resuspended at 10^7/ml in cold RSB (10 mM Tris-hydrochloride [pH 7.4], 10 mM NaCl, 3 mM $MgCl_2$, 0.5% Nonidet P-40). After thorough mixing, nuclei were pelleted by centrifugation (2,000 rpm, 5 min, 4°C). The supernatant (cytoplasmic fraction) was adjusted to 0.1 M NaCl, 10 mM EDTA, 0.5% sodium dodecyl sulfate, and 200 μg of proteinase K per ml and incubated at 37°C for 1 h. The nuclear pellet was washed in cold RSB and then resuspended in 0.1 M NaCl–1 mM EDTA–10 mM Tris (pH 7.4). Proteinase K was added to 200 μg/ml, followed by sodium dodecyl sulfate to 0.5%, and the sample was incubated at 37°C for 1 h. Polyadenylated RNA was then isolated as described previously (13).

Hybridization probes. (i) AMV-specific $cDNA_{myb}$ was prepared from AMV virion RNA as described previously (14). (ii) cDNA enriched for sequences from the 3′ portion of v-*myb* was prepared from AMV 70S virion RNA by procedures described previously (14) for the preparation of a 3′-specific probe from MAV 70S RNA, except that in the present work, the length distribution of the cDNA was 300 to 1,000 bases. (iii) The cloned v-*myb* probe was prepared using the 1.25-kilobase pair (kbp) *Kpn*I-*Xba*I fragment of cloned AMV DNA which contains mainly *myb* sequences (27). This fragment was prepared from a cloned *Bam*HI fragment containing the 3′ portion of an AMV provirus (which has the entire v-*myb* region [27]); the isolation of this clone is described elsewhere (16). (iv) Probes A, B, C, D, and E (see Fig. 3) were prepared with DNA isolated from the c-*myb* recombinant phage. The 1.3-kbp *Hin*dIII fragment used to prepare probe A and the 1.0-kbp *Eco*RI fragment used to prepare probe C were subcloned in plasmids. Probe B was prepared by *Eco*RI-*Bst*NI digestion of the probe A fragment. The DNA fragment used to prepare probe D was obtained by *Eco*RI-*Hin*dIII digestion of DNA from the recombinant phage λCM-1 (see below and Fig. 3), while that used to prepare probe E was obtained by *Sac*I digestion of λCM-2 DNA (see Fig. 3).

Preparative gel electrophoresis was performed with low-melting-point agarose (Seaplaque, FMC Inc.). Bands were visualized by ethidium bromide staining and excised from the gel. The agarose slices were melted at 68°C in the presence of 4 volumes of 0.3 M NaCl–0.02 M Tris-hydrochloride (pH 7.4)–0.001 M EDTA–0.1% sodium dodecyl sulfate and extracted twice with phenol. The aqueous phase was concentrated with butanol-isopropanol (7:3), and DNA was recovered by ethanol precipitation.

Synthesis of radiolabeled probes from DNA fragments using AMV polymerase and random primers has been described (14). Filter hybridization, washing, and autoradiography were carried out according to published procedures as previously described (14).

RESULTS

Detection of *myb* sequences in the chicken genome. DNA from a chicken which lacks endogenous proviral sequences (15) was digested with several restriction endonucleases, electrophoresed through an agarose gel, and transferred to a nitrocellulose filter by the Southern blotting procedure (26). The filter was then probed by hybridization to $cDNA_{myb}$ (see above). Several hybridizing fragments were detected with most of the endonucleases used (Fig. 1). In many cases, three or more fragments larger than 1.2 kbp (the size of v-*myb*) were detected; this suggests that the regions of homology to v-*myb* are separated by nonhomologous sequences, i.e., by introns. (Note that *Eco*RI digestion yielded four fragments, the band in Fig. 1, lane 2, between 2.0 and 2.3 kbp, being a doublet; this is apparent on the original autoradiogram and is shown by the analysis of the recombinant clones described below.)

Isolation of c-*myb* clones from a library of chicken DNA-bacteriophage recombinants. To further characterize c-*myb*, we isolated bacteriophage containing *myb* sequences from a library of chicken DNA fragments cloned in bacterio-

phage Charon 4A (11). Approximately 4×10^5 phage were screened by plaque hybridization techniques (1) with $cDNA_{myb}$ as a probe. Four positive plaques were detected, and the phage from each were purified; subsequent analysis showed two to be identical. DNA from the three nonidentical clones, λCM-1, λCM-2, and λCM-3, was digested with *Eco*RI and with *Eco*RI plus *Sac*I (Fig. 2A). This analysis, when followed by Southern blotting and hybridization to $cDNA_{myb}$ (Fig. 2B) showed the following. (i) λCM-1 plus λCM-3 together contain all four c-*myb* *Eco*RI fragments which are found in chicken chromosomal DNA (8.7, 5.5, 2.1, and 2.0 kbp; Fig. 1). (ii) DNA from λCM-3 contains all four c-*myb* *Eco*RI-*Sac*I fragments found in the genome (1.6, 2.1, 2.0, and 0.9 kbp; Fig. 1), implying that all of the sequences homologous to v-*myb* are present in this clone. This has been confirmed in experiments using a probe prepared from cloned v-*myb* sequences (data not shown). (iii) These data show that the three clones overlap, since all three contain the 2.1-kbp *Eco*RI fragment, λCM-1 and λCM-3 both contain the 2.0-kbp *Eco*RI and 1.6-kbp *Eco*RI-*Sac*I fragments, and λCM-2 and λCM-3 both contain the 8.7-kbp *Eco*RI fragment (Fig. 2B). The presence of a 1.9-kbp *Eco*RI fragment in λCM-2 (Fig. 2, lanes 3 and 9)

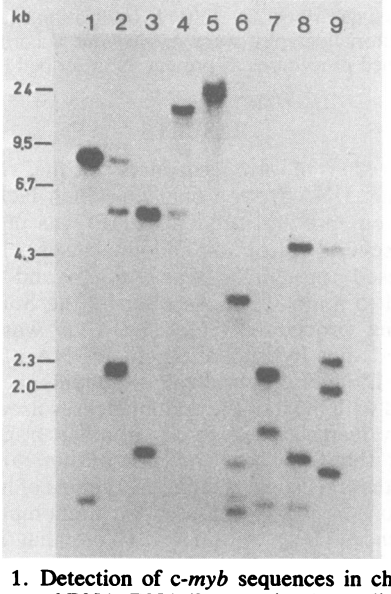

FIG. 1. Detection of c-*myb* sequences in chicken chromosomal DNA. DNA (8 μg per lane) was digested with various restriction endonucleases and analyzed by agarose gel electrophoresis, Southern blotting, and hybridization to $cDNA_{myb}$. The restriction endonucleases used were: *Sac*I (lane 1), *Eco*RI (lane 2), *Hin*dIII (lane 3), *Bam*HI (lane 4), *Xho*I (lane 5), *Pst*I (lane 6), *Sac*I plus *Eco*RI (lane 7), *Sac*I plus *Hin*dIII (lane 8), and *Hin*dIII plus *Eco*RI (lane 9). The positions of size markers are indicated at the left.

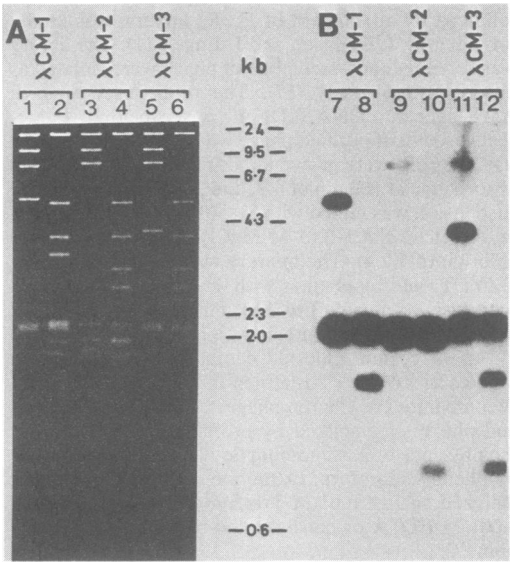

FIG. 2. Restriction analysis of DNA from c-*myb* recombinant phage. DNA from λCM-1, λCM-2, and λCM-3, as indicated, was digested with *Eco*RI (lanes 1, 3, 5, 7, 9, and 11) or *Eco*RI plus *Sac*I (lanes 2, 4, 6, 8, 10, and 12) and electrophoresed through an agarose gel. The DNA fragments were detected by staining with ethidium bromide (A) and then were transferred to nitrocellulose and hybridized to $cDNA_{myb}$. The resultant autoradiogram is shown in (B). The positions of size markers are indicated between the two panels.

suggests that the c-*myb* insert in this clone terminates within the 2.0-kbp *Eco*RI fragment found in chromosomal DNA and in λCM-1 and λCM-3. Similarly, the 4.0-kbp *Eco*RI fragment in λCM-3 presumably arises because the c-*myb* insert in this clone terminates within the 5.5-kbp *Eco*RI fragment found in chromosomal DNA and in λCM-1. Thus the data imply that the order of the *Eco*RI fragments containing sequences homologous to v-*myb* is 5.5 kbp-2.0 kbp-2.1 kbp-8.7 kbp; this is confirmed by more detailed restriction mapping (see below).

Detailed mapping of the c-*myb* locus. Having established the relative positions of the sequences contained in the recombinants, λCM-1, λCM-2, and λCM-3, we constructed a more detailed map of the c-*myb* locus. Data were obtained by single and double restriction endonuclease digestion, followed by Southern blotting and hybridization to $cDNA_{myb}$ or to probes prepared from cloned v-*myb* DNA. These data were compared with genomic blotting data (Fig. 1). The resultant map is shown in Fig. 3 along with the positions and orientations of the inserts of the three c-*myb* phage clones.

The transcriptional orientation of the gene relative to the restriction map was determined with a cDNA probe enriched in sequences from

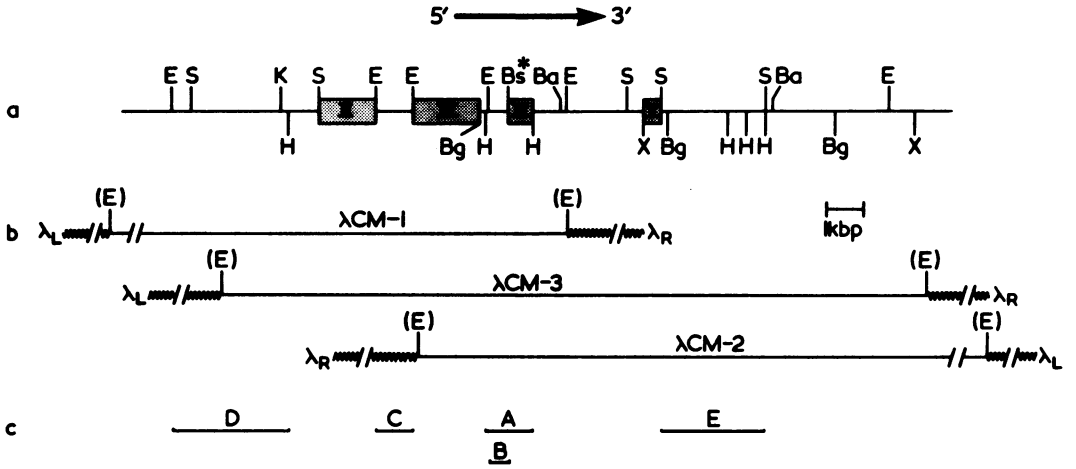

FIG. 3. (a) Restriction map of the c-*myb* locus. The sites of cleavage of the following restriction endonucleases are shown: *Eco*RI (E), *Sac*I (S), *Hin*dIII (H), *Bam*HI (Ba), *Bgl*II (Bg), *Kpn*I (K), and *Xba*I (X). In addition, the position of one *Bst*NI site (Bs*) is shown, as it defines one boundary of probe B; the locations of other *Bst*NI sites have not been determined. Also shown are the positions of the four regions of homology with v-*myb* (stippled boxes), numbered II, III, IV, and V; the precise positions of the v-*myb*-related sequences within these four regions were not determined in this study. The arrow indicates the direction of transcription (see text). (b) Position and orientation (see text) of the three c-*myb* recombinant phage, λCM-1, λCM-2, and λCM-3, with respect to the map of the c-*myb* locus. (Note that the *Eco*RI sites at the boundaries of the inserts in these phage were created by addition of linkers used in constructing the library.) Wavy lines indicate the left (λ_L) and right (λ_R) arms of the Charon 4A vector (not to scale). (c) Positions of the DNA fragments used to prepare probes A, B, C, D, and E, in relation to the restriction map shown in (a) above.

the 3′ portion of the v-*myb* gene. This probe was prepared from AMV virion RNA by reverse transcription with oligodeoxythymidylate as a primer (see above). The final oligodeoxythymidylate-containing cDNA had a length distribution of 300 to 1,000 bases, so that a proportion of the DNA molecules extended into the v-*myb* gene (about 600 bases from the 3′ [polyadenylate] terminus of the RNA), but none of it extended to the 5′ end of the v-*myb* gene (about 1.8 kb from the 3′ terminus of the RNA [12, 26]). Hybridization of this probe to *Hin*dIII-*Eco*RI digests of λCM-1, λCM-2, and λCM-3 (Fig. 4) showed that 3′ sequences were present in λCM-2 and λCM-3 but not in λCM-1. Furthermore, only the 4.6-kbp *Eco*RI-*Hin*dIII fragment hybridized to the 3′-enriched probe in λCM-2 and λCM-3; location of this fragment on the map shown in Fig. 3 yields the orientation of the gene. This is consistent with the relative orientation of the three clones (Fig. 3). Note that the preferential reaction of the 3′-enriched probe with one fragment rules out the possibility that the four hybridizing *Eco*RI fragments detected with $cDNA_{myb}$ or cloned v-*myb* probes represent four separate copies of the complete *myb* sequence. Southern blotting experiments with chromosomal DNA and probes derived from the recombinant molecules reinforce this conclusion (data not shown). In the light of the restriction map of c-*myb* shown in Fig. 3, the data of Souza et al. (27) and Perbal and Baluda (20) are also consistent with the orientation deduced here.

The data summarized in Fig. 3 reveal that the c-*myb* sequences homologous to v-*myb*, which have a total length of about 1.2 kbp, are distributed between at least four separate regions (numbered II to V, Fig. 3) of the c-*myb* locus which span a minimum of 8 kbp. Thus, these data suggest that the c-*myb* gene contains at least four exons; although the positions of the exons could not be further localized within each region by mapping the phage clones, nucleotide sequencing has subsequently revealed that regions II to V contain a total of seven exons (16).

Transcription of the c-*myb* locus. RNA from several chicken tissues and cell lines has been shown to contain *myb* sequences (7, 13). Previous work from this laboratory (13) has shown the presence of a major polyadenylated transcript of 4.0 kb in these cells and tissues; in addition, several larger RNA species were detected. Further analysis of c-*myb* transcription was performed with polyadenylated RNA from AEV-transformed erythroid cells (erythroblasts), which was previously shown to contain a relatively high level of c-*myb* sequences (13). The RNA was fractionated on formaldehyde-agarose gels, transferred to nitrocellulose, and hybridized to probes prepared from cloned v-*myb* DNA and from λCM-3 DNA.

Figure 5 shows that several RNA species

hybridized to both probes (lanes 1 and 2). Earlier studies, in which the single-stranded $cDNA_{myb}$ was used as a hybridization probe, also demonstrated the presence of the two smaller RNA species in polyadenylated RNA from uninfected chicken tissues (data not shown). These observations show that the major transcripts from the c-*myb* locus contain the sequences homologous to v-*myb* and are of the same polarity as the v-*myb* RNA species.

The relationship between the various species was first investigated by examining RNA isolated from nuclear and cytoplasmic fractions of the AEV-transformed erythroblasts. In this experiment (Fig. 6), a 1.3-kbp *Hin*dIII subclone from the c-*myb* region (Fig. 3, probe A) was used; this probe detects the same RNA species as v-*myb* and λCM-3 probes (cf. Fig. 5, lanes 1 and 2, and Fig. 6, lane 1) (the 7.9-kb RNA species can be seen on longer exposures of the autoradiogram shown in Fig. 6). Despite some degradation of the RNA which occurred during the fractionation, it is apparent from Fig. 6 that only 4.0-kb species is present in the cytoplasm, whereas the 5.4-kb species and, presumably, the other, larger RNAs are confined to the nucleus. This suggests that only the 4.0-kb species is a mature mRNA and that the larger species may be precursors.

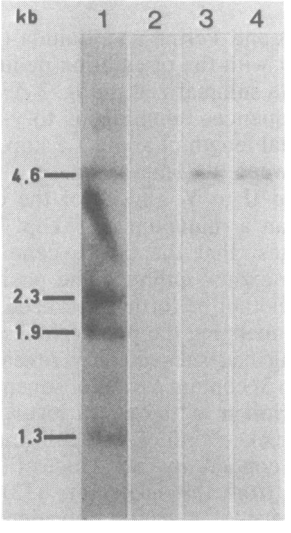

FIG. 4. Determination of the transcriptional orientation of c-*myb*. DNA from λCM-1 (lane 2), λCM-2 (lane 3), and λCM-3 (lane 4) was digested with *Eco*RI plus *Hin*dIII and analyzed by agarose gel electrophoresis and Southern blotting and hybridized to cDNA enriched for 3′ *myb* sequences (see text). For comparison, DNA from a parallel *Eco*RI-plus-*Hin*dIII digestion of λCM-3 was similarly analyzed by hybridization to the cloned v-*myb* probe (lane 1). Note that the fragments detected by this latter probe are the same as those detected in an *Eco*RI-plus-*Hin*dIII digest of chromosomal DNA (cf. with Fig. 1, lane 9). The sizes of these fragments are indicated at the left.

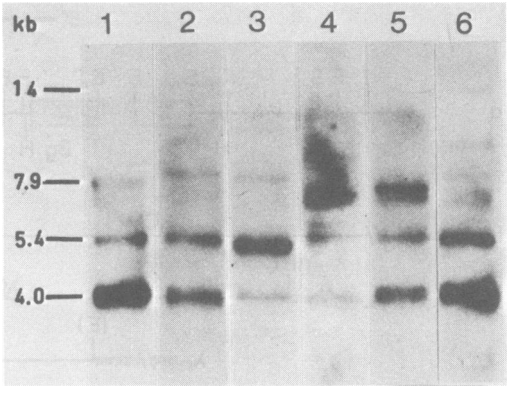

FIG. 5. Transcription from c-*myb*. Polyadenylated RNA (4 μg per lane) from AEV-transformed erythroblasts was analyzed by gel electrophoresis and filter hybridization as described in the text. The probes used were cloned v-*myb* DNA (lane 1), λCM-3 DNA (lane 2), probe B (lane 3), probe C (lane 4), probe D (lane 5), and probe E (lane 6). The sizes of several of the RNA species detected are indicated at the left.

The nature of the c-*myb* transcripts was further investigated by hybridization to several probes prepared from the c-*myb* clones (see Fig. 3). As mentioned above, probe A detected many of the same RNA species as did the v-*myb* and λCM-3 probes; however, comparison of lane 1 of Fig. 5 with lane 1 of Fig. 6 shows that the reaction of the 5.4-kb species was greater with probe A than with the v-*myb* probe (relative to the reactions of each of these probes with the 4.0-kb species). This suggests that the 5.4-kb species contains sequences which are between the v-*myb*-related (exon) sequences, i.e., intervening or intron sequences. In support of this, probe B reacted predominantly with the 5.4-kb species (Fig. 5, lane 3); the reaction of the 4.0-kb species with this probe was probably due to low-level contamination of the probe with v-*myb*-related sequences, which in turn probably originated with the preparative electrophoresis step used in the isolation of the template for probe B. (DNA sequencing has subsequently shown that the exon in region IV [see Fig. 3] is outside the region covered by probe B [16]; furthermore, a probe prepared with cloned DNA from the region contained within probe B does not react with the 4.0-kb RNA [personal communication, K.-H. Klempnauer]). Probe C, prepared from a fragment which does not react detectably with v-*myb* probes (see Fig. 2), reacts most extensively with two of the larger c-*myb* RNAs (of 7.2 and 7.4 kb) and several other species (Fig. 5, lane 4). A very faint reaction was also detected with the

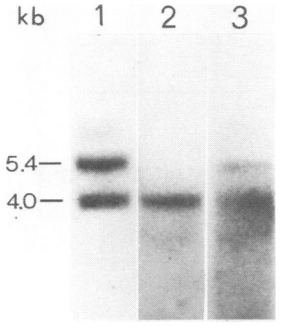

FIG. 6. Subcellular distribution of c-*myb* transcripts. Polyadenylated RNA was prepared from nuclear and cytoplasmic fractions of AEV-transformed erythroblasts as described in the text. The RNA was analyzed by gel electrophoresis and filter hybridization with probe A (see Fig. 3). Lane: 1, total polyadenylated RNA; 2, RNA from the cytoplasmic fraction; 3, RNA from the nuclear fraction. The sizes of the two major transcripts detected are indicated at the left.

4.0-kb species; this is explained by the results of DNA sequencing work (16) which has shown that the *Eco*RI site at the 3′ boundary of probe C (see Fig. 3) actually lies 26 bases within one exon in region III. (The failure of v-*myb* probes to react with the 1.0-kb *Eco*RI fragment used to prepare probe C, despite the 26-base region of homology, may be due to [i] the lower stability of DNA-DNA duplexes compared with RNA-DNA hybrids or [ii] the 1-base mismatch between v-*myb* and c-*myb* in the center of this 26-base sequence or both [16].)

Hybridization probes were also prepared from fragments which map both 5′ (probe D) and 3′ (probe E) to the regions homologous to v-*myb* (see Fig. 3). Both of these probes reacted with the 4.0-kb c-*myb* transcript and also with several larger species; this demonstrates that the c-*myb* transcriptional unit extends beyond the v-*myb*-related sequences. The relative amounts of hybridization of the 4.0-kb RNA compared with the larger RNAs were similar for probe E and for the v-*myb* probe (cf. Fig. 5, lanes 6 and 1). This suggests that the sequences in probe E are not removed by processing of the original transcript. Probe D, however, hybridized disproportionately (compared with the v-*myb* probe) to two large RNAs of 7.6 and 7.2 kb (Fig. 5, lane 5), suggesting that some (but not all) of the sequences in probe D which are present in the 7.6- and 7.2-kb species are removed during processing of the original transcript.

In addition, an experiment complementary to those described above was performed, in which a cDNA probe was prepared from polyadenylated RNA (from AEV-transformed erythroblasts) which had been enriched (by sucrose gradient centrifugation) for the 4.0-kb c-*myb* transcript. This probe hybridized to sequences within the *Eco*RI-*Hin*dIII fragment used to prepare probe D and to sequences immediately 5′ to this fragment, but not to sequences in λCM-3 which are 5′ to region II (unpublished observations) (Fig. 3 and 7). This result limits the extent of region I (Fig. 7) on the 3′ side. Furthermore, the enriched cDNA did not hybridize to any sequences 3′ to the *Sac*I fragment used to prepare probe E (unpublished observations); thus, the c-*myb* transcript does not extend beyond the 3′ boundary of this fragment (Fig. 3 and 7).

DISCUSSION

In this report, we have described the isolation and characterization of recombinant bacteriophage clones containing chicken DNA homologous to the presumptive oncogene of AMV. Since the sequences detected and cloned here represent all of the *myb* sequences in the chicken genome, it is likely that this gene (c-*myb*) represents the progenitor of v-*myb*. This conclusion is strongly supported by nucleotide sequence analysis (16).

Our analysis of the c-*myb* locus has shown the presence of a minimum of four separate regions which contain sequences homologous to v-*myb*; nucleotide sequencing has shown that there are, in fact, seven exons within these four regions (16). Since most of these sequences are present in mature messenger-like RNA and probably encode a protein (16), we have referred to them as exons and to the sequences in between them as introns. Thus, our analysis of the c-*myb* gene has shown that it is organized and transcribed in a manner similar to that of most eucaryotic cellular genes, in that the sequences present in the mature mRNA, including the presumptive coding sequences, are interrupted by several introns. In this respect, it resembles several other c-*onc* genes, e.g., c-*src* (19, 25), c-*myc* (21, 27), c-Ha-*ras*1 (9), and the two c-*erb* loci, c-*erb*A and c-*erb*B (24, 28), although each of the c-*erb* loci gives rise to more than one cytoplasmic transcript (28). We have also presented evidence for an additional exon located 5′ to the v-*myb*-related sequences and for the presence of c-*myb* exon sequences 3′ to the v-*myb*-related sequences. These 3′ c-*myb* sequences may be contiguous with the 3′-most v-*myb*-related exon, or alternatively, they may comprise a separate exon.

The 4.0-kb c-*myb* transcript is almost certainly the mRNA encoding the c-*myb* gene product, since it is polyadenylated and is the only c-*myb* transcript present in the cytoplasm. The other c-*myb* transcripts are thus likely to be precursors of the 4.0-kb species, since they are larger, contain intron sequences, and are associated

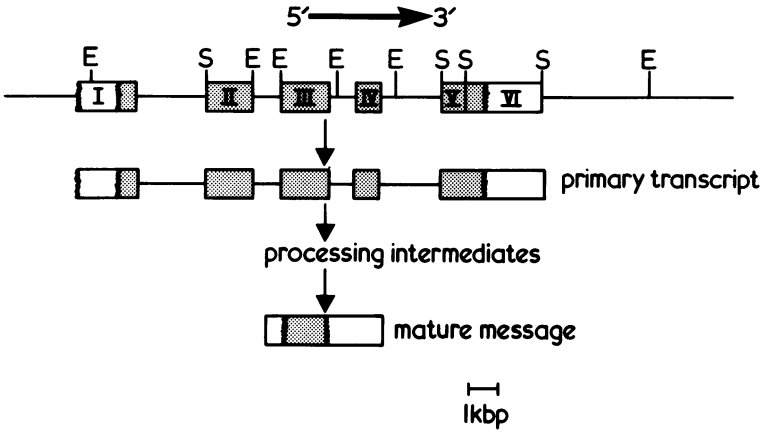

FIG. 7. Possible locations of sequences found in mature c-*myb* RNA on the map of the c-*myb* locus. Only the *Eco*RI (E) and *Sac*I (S) sites are shown here (see Fig. 3). The locations of sequences present in the mature c-*myb* mRNA are indicated by the boxed regions, numbered I to VI; regions II to IV contain sequences homologous to v-*myb* (see text). The putative untranslated regions of the message are indicated by the open boxes; the putative coding regions (see text) are within the regions shown in stippled boxes. Jagged boundaries are used where the limits of a particular region have not been defined.

with the nucleus. The c-*myb* transcripts also contain sequences located 5′ and 3′ to the v-*myb*-related exons; these sequences are likely to include the 5′ and 3′ untranslated regions of the c-*myb* mRNA (see below).

Although the experiments described here do not allow the precise mapping of the c-*myb* transcripts onto the gene, they do suggest that the c-*myb* transcriptional unit is arranged in a manner similar to that illustrated in Fig. 7. The main features of this scheme, which are discussed below, are the following. (i) The sequences detected by probe D include an exon in region I (Fig. 7) which presumably includes the 5′ untranslated region present in the c-*myb* mRNA. (ii) There is a region of up to 2.6 kbp (region VI, detected by probe E) which is adjacent to the 3′-most v-*myb*-related exon (in region V) and presumably includes the 3′ untranslated region of the c-*myb* mRNA. (iii) The c-*myb* protein coding sequences include those sequences in regions II to V which are homologous to the coding sequences of v-*myb*. This last feature is supported by comparison of the nucleotide sequences of v-*myb* and the homologous exons of c-*myb*, in that these sequences contain an open reading frame which comprises all but the first 84 nucleotides of their length (of 1.2 kbp) (16). The scheme shown in Fig. 7 predicts a primary c-*myb* transcript of about 16 kb; a transcript of approximately this size (14 kb) has been detected (a faint band of this size is visible in Fig. 5, lane 6; the 14-kb species is also seen on longer exposures [not shown] of other lanes shown in Fig. 5). Note that the difference between the length of the v-*myb* sequence (1.2 kb) and the size of the c-*myb* mRNA (4.0 kb) implies that the exons in regions I and VI contain a total of about 2.8 kb; the distribution of these 2.8 kb between the 5′ and 3′ regions of the c-*myb* mRNA is as yet unknown.

To date, the gene products of v-*myb* and c-*myb* have not been identified, and the nucleotide sequences of v-*myb* (16, 23) and the corresponding regions of c-*myb* (16) do not allow conclusive identification of the initiation codons for either protein or of the termination codon for c-*myb*. Since, as we have shown, there are exon sequences in c-*myb* both 5′ (region I) and 3′ (region VI) to the sequences homologous to v-*myb*, it is possible that protein coding sequences are present in one or both of these regions. Thus, the c-*myb* gene product may be significantly larger than the v-*myb* product, since (as mentioned above) regions I plus VI contain a total of about 2.8 kb of exon sequences.

In view of the possibility that the c-*myb* gene product may be larger than the v-*myb* gene product, it is of interest to consider the genome structure and protein coding capacity of E26, an avian acute leukemia virus which carries v-*myb*-related sequences (5, 12, 22). It appears likely that more genetic information is required to encode the E26-specific fusion protein P135 than can be accounted for by *gag*- and v-*myb*-related sequences in the E26 genome (5). E26 (like AMV) probably arose by capture of part of c-*myb* by a nondefective retrovirus; thus, we might reasonably predict that the additional P135 coding sequences in the E26 genome would be derived from c-*myb* sequences other than those homologous to v-*myb*, i.e., from sequences in region I or VI or both. (In fact, region VI is a better candidate, since the v-*myb*-related

sequences appear to be located in the 5′ portion of the E26 genome [5].) If this reasoning is correct, it follows that there is a significant amount of coding sequence in c-*myb* not present in v-*myb*.

In summary, the data presented in this report show that the c-*myb* gene spans about 16 kb and comprises at least four regions of sequences which are present in v-*myb* plus (at least) two additional regions found in the c-*myb* mRNA. Taken in conjunction with the results of nucleotide sequence analysis (16), the data raise the possibility that the c-*myb* gene product is significantly larger than, and thus may differ functionally from, the v-*myb* gene product. These results suggest that further analysis of the c-*myb* gene should focus on the nucleotide sequences in regions I and VI; for example, the 5′ and 3′ termini of the c-*myb* mRNA could be located using restriction fragments from these regions in S1 mapping (3) experiments.

ACKNOWLEDGMENTS

We thank K.-H. Klempnauer for providing two of the DNA fragments used in this study. We also thank Susan Blackford for typing the manuscript and Pierre Smith and Sonia Belan for assistance with photography.

This work was supported by grants from the American Cancer Society and the National Cancer Institute; T.J.G. was the recipient of a postdoctoral fellowship from the American Cancer Society (California division).

LITERATURE CITED

1. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science **196**:180–182.
2. **Bergmann, D. G., L. M. Souza, and M. A. Baluda.** 1981. Vertebrate DNAs contain nucleotide sequences related to the transforming gene of avian myeloblastosis virus. J. Virol. **40**:450–455.
3. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S_1 endonuclease-digested hybrids. Cell **12**:721–732.
4. **Bishop, J. M.** 1981. Enemies within: the genesis of retrovirus oncogenes. Cell **23**:5–6.
5. **Bister, K., M. Nunn, C. Moscovici, B. Perbal, M. A. Baluda, and P. H. Duesberg.** 1982. Acute leukemia viruses E26 and avian myeloblastosis virus have related transformation-specific RNA sequences but different genetic structures, gene products, and oncogenic properties. Proc. Natl. Acad. Sci. U.S.A. **79**:3677–3681.
6. **Blattner, F. R., B. G. Williams, A. E. Blechl, K. Denniston-Thompson, H. E. Faber, L.-A. Furlong, D. J. Grunwald, D. O. Kiefer, D. O. Moore, J. W. Schumm, E. L. Sheldon, and O. Smithies.** 1977. Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. Science **19**:161–169.
7. **Chen, J. H.** 1980. Expression of endogenous avian myeloblastosis virus information in different chicken cells. J. Virol. **36**:162–170.
8. **Coffin, J. M., H. E. Varmus, J. M. Bishop, M. Essex, W. D. Hardy, Jr., G. S. Martin, N. E. Rosenberg, E. M. Scolnick, R. A. Weinberg, and P. K. Vogt.** 1981. Proposal for naming host cell-derived inserts in retrovirus genomes. J. Virol. **40**:953–957.
9. **DeFeo, D., M. A. Gonda, H. A. Young, E. H. Chang, D. R. Lowy, E. M. Scolnick, and R. W. Ellis.** 1981. Analysis of two divergent rat genomic clones homologous to the transforming gene of Harvey murine sarcoma virus. Proc. Natl. Acad. Sci. U.S.A. **78**:3328–3332.
10. **Der, C. J., T. G. Krontiris, and G. M. Cooper.** 1982. Transforming genes of human bladder and lung carcinoma cell lines are homologous to the *ras* genes of Harvey and Kirsten sarcoma viruses. Proc. Natl. Acad. Sci. U.S.A. **79**:3637–3640.
11. **Dodgson, J. B., J. Strommer, and J. D. Engel.** 1979. Isolation of the chicken β-globin gene and a linked embryonic β-like globin gene from a chicken DNA recombinant library. Cell **17**:879–887.
12. **Duesberg, P. H., K. Bister, and C. Moscovici.** 1980. Genetic structure of avian myeloblastosis virus, released from transformed myeloblasts as a defective virus particle. Proc. Natl. Acad. Sci. U.S.A. **77**:5120–5124.
13. **Gonda, T. J., D. K. Sheiness, and J. M. Bishop.** 1982. Transcripts from the cellular homologs of retroviral oncogenes: distribution among chicken tissues. Mol. Cell. Biol. **2**:617–624.
14. **Gonda, T. J., D. K. Sheiness, L. Fanshier, J. M. Bishop, C. Moscovici, and M. G. Moscovici.** 1981. The genome and the intracellular RNAs of avian myeloblastosis virus. Cell **23**:279–290.
15. **Hughes, S. H., F. Payvar, D. Spector, R. T. Schimke, H. L. Robinson, G. S. Payne, J. M. Bishop, and H. E. Varmus.** 1979. Heterogeneity of genetic loci in chickens: analysis of endogenous viral and nonviral genes by cleavage of DNA with restriction endonucleases. Cell **18**:347–359.
16. **Klempnauer, K.-H., T. J. Gonda, and J. M. Bishop.** 1982. Nucleotide sequence of the retroviral leukemia gene *v-myb* and its cellular progenitor *c-myb*: the architecture of a transduced oncogene. Cell **31**:453–463.
17. **Maniatis, T., R. C. Hardison, E. Lacy, J. Laver, C. O'Connell, D. Quon, G. K. Sim, and A. Efstratiadis.** 1978. The isolation of structural genes from libraries of eukaryotic DNA. Cell **15**:687–701.
18. **Parada, L. F., C. J. Tabin, C. Shih, and R. A. Weinberg.** 1982. Human EJ bladder carcinoma oncogene is homologue of Harvey sarcoma virus *ras* gene. Nature (London) **297**:474–478.
19. **Parker, R. C., H. E. Varmus, and J. M. Bishop.** 1981. Cellular homologue (c-*src*) of the transforming gene of Rous sarcoma virus: isolation, mapping, and transcriptional analysis of *c-src* and flanking regions. Proc. Natl. Acad. Sci. U.S.A. **78**:5842–5846.
20. **Perbal, B., and M. A. Baluda.** 1982. Avian myeloblastosis virus transforming gene is related to unique chicken DNA regions separated by at least one intervening sequence. J. Virol. **41**:250–257.
21. **Robins, T., K. Bister, C. Garon, T. Papas, and P. Duesberg.** 1982. Structural relationship between a normal chicken DNA locus and the transforming gene of the avian acute leukemia virus MC29. J. Virol. **41**:635–642.
22. **Roussel, M., S. Saule, C. Lagrou, C. Rommens, H. Beug, T. Graf, and D. Stehelin.** 1979. Three new types of viral oncogene of cellular origin specific for haematopoietic cell transformation. Nature (London) **281**:452–455.
23. **Rushlow, K. E., J. A. Lautenberger, T. S. Papas, M. A. Baluda, B. Perbal, J. G. Chirikjian, and E. P. Reddy.** 1982. Nucleotide sequence of the transforming gene of avian myeloblastosis virus. Science **216**:1421–1423.
24. **Saule, S., M. Roussel, C. Lagrou, and D. Stehelin.** 1981. Characterization of the oncogene (*erb*) of avian erythroblastosis virus and its cellular progenitor. J. Virol. **38**:409–419.
25. **Shalloway, D., A. D. Zelenetz, and G. M. Cooper.** 1981. Molecular cloning and characterization of the chicken gene homologous to the transforming gene of Rous sarcoma virus. Cell **24**:531–541.
26. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
27. **Souza, L. M., J. N. Strommer, R. L. Hillyard, M. C. Komaromy, and M. A. Baluda.** 1980. Cellular sequences are present in the presumptive avian myeloblastosis virus

genome. Proc. Natl. Acad. Sci. U.S.A. **77**:5177–5181.
28. **Vennstrom, B., and J. M. Bishop.** 1982. Isolation and characterization of chicken DNA homologous to the two putative oncogenes of avian erythroblastosis virus. Cell **28**:135–143.
29. **Vennstrom, B., D. Sheiness, J. Zabielski, and J. M. Bishop.** 1982. Isolation and characterization of c-*myc*, a cellular homolog of the oncogene (v-*myc*) of avian myelocytomatosis virus strain 29. J. Virol. **42**:773–779.
30. **Westin, E. H., R. C. Gallo, S. K. Arya, A. Eva, L. M. Souza, M. A. Baluda, S. A. Aaronson, and F. Wong-Staal.** 1982. Differential expression of the *amv* gene in human hematopoietic cells. Proc. Natl. Acad. Sci. U.S.A. **79**:2194–2198.