MOLECULAR AND CELLULAR BIOLOGY, Aug. 1986, p. 2800–2806
0270-7306/86/082800-07$02.00/0

Vol. 6, No. 8

Features of the Chicken c-*myc* Gene That Influence the Structure of c-*myc* RNA in Normal Cells and Bursal Lymphomas

CAROL NOTTENBURG* AND HAROLD E. VARMUS

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

Received 2 January 1986/Accepted 6 May 1986

The chicken c-*myc* gene is the target for proviral insertion mutations in bursal lymphomas and has been transduced to generate several viral oncogenes, but the boundaries of its exons have not been securely established. To define the landmarks of the chicken c-*myc* gene necessary to produce its mRNA, we used an RNase protection assay and a cDNA clone to analyze the c-*myc* mRNAs from normal chicken embryos and from two bursal lymphomas: LL6, which contains an avian leukosis virus provirus downstream of the c-*myc* coding region, and LL7, which contains an avian leukosis virus provirus upstream of the c-*myc* coding region. Two initiation sites for normal c-*myc* mRNA are less than 7 bases apart, downstream of a GC-rich region lacking canonical TATA and CAAT sequences. The first exon has two open reading frames for the entire length but no initiator methionine codons. The splice donor and acceptor sites at the boundary of the first intron were assigned by comparing a sequence of an LL6 c-*myc* cDNA clone with a genomic DNA sequence and confirmed by RNase protection of labeled RNA probes by normal and LL6-derived mRNAs. Two potential polyadenylation signals are located approximately 250 and 400 bases downstream of the c-*myc* coding region in the third exon, but only the more distal signal is utilized in both normal cells and the LL7 tumor. The proviral integration in the LL6 tumor occurred upstream of the authentic c-*myc* polyadenylation signal accounting for polyadenylation of this transcript in the proviral long terminal repeat.

The importance of the c-*myc* gene in tumorigenesis is manifest by its frequent association with chromosomal alterations and augmented levels of c-*myc* RNA in a variety of lymphoid tumors (46). In chicken bursal lymphomas caused by avian leukosis virus and chicken syncticial virus, the c-*myc* locus is almost always interrupted by a provirus (19, 21, 31, 32). The disrupted chicken c-*myc* locus is often transcribed from a viral promoter present in the long terminal repeat (18, 21, 32) and sometimes from cryptic promoters (25, 32). Murine thymomas caused by murine leukemia viruses as well as feline T-cell tumors caused by feline leukemia virus also frequently contain a disrupted c-*myc* locus (10, 23, 24, 29, 39). Chicken and feline c-*myc* genes are both progenitors of v-*myc* oncogenes contained in several replication-defective viruses and causing rapid onset of myelomonocytic tumors and carcinomas in chickens (1, 46) and lymphoid tumors in cats (23, 29). Although the function of the *myc* protein is speculative, the protein binds DNA in vitro (13, 33), is associated with the nucleus in vivo (2, 13, 16, 33), and may play a central role in the control of cell growth (8, 22).

Incomplete mapping of the normal chicken c-*myc* locus hampers efforts to interpret the contribution of gene alterations to bursal lymphomagenesis, especially when both the gene structure and the resulting c-*myc* RNA are suspected to contain physical alterations (21, 25, 32, 40). The entire coding region of the chicken c-*myc* gene appears to be contained within two exons (1, 40, 47), but the length of the normal c-*myc* RNA is substantially greater than the combined length of the coding regions. Evidence for an additional, noncoding exon in the chicken c-*myc* gene (25, 40) suggests that chicken c-*myc* is structurally analogous to the better defined mammalian c-*myc* genes which are composed of three exons, a noncoding exon followed by two coding exons (5, 41–43, 48).

* Corresponding author.

Using a nearly full-length cDNA clone from a chicken bursal lymphoma, LL6, with a provirus integrated 3′ of the c-*myc* coding sequences (32), and RNAs from normal chicken embryos, LL6, and a second bursal lymphoma, LL7 (which has a provirus 5′ of the coding regions), we rigorously determined the landmarks in the chicken c-*myc* gene important for the synthesis and processing of its RNA: the transcriptional initiation site, the splice sites for the first and second intron, and the single polyadenylation site. In conjunction with previous analyses of v-*myc* and c-*myc*, these findings provide a more complete picture of the structure of chicken c-*myc* and reveal several striking dissimilarities in gene organization to the mammalian c-*myc* genes.

MATERIALS AND METHODS

Isolation of a c-*myc* cDNA clone. mRNA from the bursal lymphoma LL6 (32) was the source of material for the construction of a cDNA library, performed by the method of T. St. John, J. Rosen, and H. Gershenfeld (personal communication). Approximately 1.5×10^6 independent clones longer than 500 base pairs (bp) were constructed and amplified by one round of replication in C600(*Hfl*A). Clones were plated and screened by hybridization with a ^{32}P-labeled plasmid containing 2.8 kilobases (kb) of chicken c-*myc* coding and intervening sequences, including sequences from the *Sst*I site in exon 1 to the *Eco*RI site in exon 3 (see Fig. 1). Purified clones hybridizing to this probe were further analyzed for length and appropriate restriction sites. One clone, λ13, contains a 3.1-kb insert and was utilized for subsequent experiments.

DNA sequencing. Appropriate restriction fragments were purified by agarose gel electrophoresis and inserted into either M13mp10 or M13mp11 vectors. Sequencing was performed by the dideoxy chain termination method (38).

Construction of probes and radioactive labeling. Four different probes were constructed for the RNA protection experiments. A plasmid containing an 8-kb *Bam*HI fragment

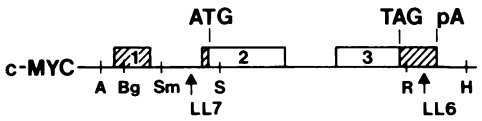

FIG. 1. Organization of chicken c-*myc* gene. A simplified restriction map of chicken c-*myc* and its major features is presented. The open boxes represent the coding region of c-*myc*. The shaded boxes represent noncoding sequences present in the mature c-*myc* mRNA. The translational initiator codon (ATG), the translational stop codon (TAG), and the polyadenylation site (pA) are indicated. Approximate proviral integration points in two different bursal lymphomas, LL6 and LL7, are indicated by arrows below the map. Relevant restriction sites are dipslayed. Not all restriction sites for each enzyme are shown. A, *Apa*I; Bg, *Bgl*I; Sm, *Sma*I; S, *Sst*I; R, *Eco*RI; H, *Hin*dIII.

encompassing the entire c-*myc* gene was doubly digested with *Bgl*I and *Sma*I, or *Apa*I and *Sma*I, or *Eco*RI and *Hin*dIII. The ends of *Bgl*I or *Apa*I digestion products were made blunt by the 3'-to-5' exonuclease action of the large fragment of *Escherichia coli* DNA polymerase I in 50 mM Tris (pH 8.0)–10 mM $MgCl_2$–50 μM deoxynucleotides for 10 min at 37°C. Two *Bgl*I-*Sma*I fragments (300 and 700 bp), an *Apa*I-*Sma*I fragment (450 bp; see Fig. 3A), and an *Eco*RI-*Hin*dIII fragment (400 bp; see Fig. 4A) were inserted into either *Sma*I-digested or *Eco*RI-*Hin*dIII-digested pT7-1 and pT7-2 vectors (US Biochemicals). Resulting clones were screened for proper orientation, and plasmid DNAs were prepared. A plasmid containing the *Bgl*I-*Sma*I 700-bp fragment was digested with *Apa*I and *Sma*I, treated with *E. coli* DNA polymerase I, and ligated to obtain a plasmid containing the 150-bp *Apa*I-*Bgl*I fragment (see Fig. 3A). Plasmid DNAs were linearized by digestion with *Eco*RI (for the pT7-1 vector) or *Hin*dIII (for the pT7-2 vector), extracted once with phenol, and ethanol precipitated. ^{32}P-labeled RNA complementary to the coding strand was synthesized from templates of plasmid DNAs by transcription with T7 RNA polymerase at 37°C for 30 min (9). Intact plasmid DNA was subsequently digested by the addition of pancreatic DNase I to 10 μg/ml and 2 U of RNAsin. Labeled RNA was extracted with phenol-chloroform (1:1 [vol/vol]) and ethanol precipitated twice.

RNA protection analysis. Approximately 100,000 dpm of ^{32}P-labeled RNA was mixed with 1 to 5 μg of LL6 mRNA, LL7 mRNA, chicken embryo mRNA, or yeast RNA in 50% recrystallized formamide–0.4 M NaCl–10 mM MOPS (morpholinepropanesulfonic acid) (pH 7.0)–1 mM trisodium EDTA in a total volume of 25 μl. Samples were heated briefly to 75°C and incubated overnight at 45 or 37°C. Hybridizations were diluted 10-fold with RNase A (30 μg/ml)–RNase T_1 (3 μg/ml) in 0.3 M NaCl–5 mM trisodium EDTA and incubated at 30°C for 30 min. RNases were inactivated by incubation with proteinase K (150 μg/ml)–0.7% sodium dodecyl sulfate at 37°C for 15 min. Samples were extracted once with an equal volume of phenol-chloroform (1:1) and precipitated with ethanol. After suspension in water and the addition of 90% formamide-dye solution, samples were heat denatured at 80°C for 10 min and applied to a 6 or 12% polyacrylamide–8 M urea gel. The resulting gel was dried and exposed to Cronex-4 X-ray film.

RESULTS

General structure of normal chicken and tumor c-*myc* genes and mRNAs. The c-*myc* gene is composed of three exons (Fig. 1). The following description of c-*myc* and Fig. 1 are compiled from published data (1, 40, 47) and data presented in this paper. Exon 1, which is entirely noncoding (40), and the first intervening sequence were mapped by RNase protection experiments in conjunction with analysis of a cDNA clone (see below). Placement of the intron between exons 2 and 3, which contain the entire coding region, was determined by comparison of the length of a restriction fragment from an LL6 cDNA clone containing the putative spliced junction with the corresponding v-*myc* restriction fragment (1, 47). The normal c-*myc* mRNA contains the information present in all three exons and is polyadenylated at a site approximately 575 bases downstream of the termination codon.

The integration sites of avian leukosis virus proviruses in two primary bursal tumors, LL6 and LL7, are shown with arrows in Fig. 1. LL6 DNA contains a deleted provirus in exon 3 located approximately 475 bp 3' of the coding sequence and in the same transcriptional orientation as c-*myc* (32; unpublished observations). LL6 contains a 3.0-kb c-*myc* mRNA, approximately 20-fold more abundant than c-*myc* RNA in normal cells and with viral sequences at its 3' terminus. More detailed structural analysis of the c-*myc* gene in LL6 will be presented elsewhere (C. Nottenberg and H. E. Varmus, manuscript in preparation). The LL7 bursal tumor allele contains a provirus integrated in the first intervening sequence, in a transcriptional orientation opposite to that of c-*myc*, although the precise location has not been determined. The c-*myc* mRNA in LL7 is similar in length to the normal mRNA but lacks sequences from exon 1 and viral sequences (32).

Characterization of 5' end of c-*myc* mRNA. Mapping of the 5' end of the c-*myc* locus was initiated by attempting to isolate full-length c-*myc* cDNA clones. A cDNA library was constructed from LL6 mRNA and screened with a c-*myc* containing probe. A clone, λ13, with a 3.1-kb insert extending from exon 1 into the virus-specific 3' end (confirmed by restriction mapping), was selected. The splice donor and acceptor sites that define the intron between exons 1 and 2 were deduced from this cDNA clone by comparing the sequence of the 5' part to the published genomic sequence located upstream of the c-*myc* coding region (Fig. 2) (40). The sequence is in good correspondence with the compiled consensus splice donor site sequence (28). However, the deduced splice donor site is in conflict with the site mapped by Linial and Groudine (25) 78 bases further downstream. Further mapping revealed that the exon 1-exon 2 junction in the cDNA clone represents the normal c-*myc* RNA species (see below).

The 5' ends of normal c-*myc* and LL6 c-*myc* RNA were

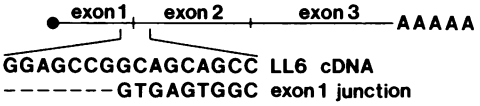

FIG. 2. Sequence of the first splice junction of c-*myc*. A diagram of a normal chicken c-*myc* mRNA is displayed, and the boundaries of the three exons are marked. DNA sequences surrounding the first splice junction are presented below the map. The uppermost sequence was determined for λ13, a cDNA clone of c-*myc* RNA derived from the bursal lymphoma LL6 (see text). The second and third sequences are genomic DNA sequences (40) from the regions of the first exon splice donor (exon 1 junction) and the second exon splice acceptor (exon 2 junction). Dashes indicate bases identical to the LL6 DNA sequence.

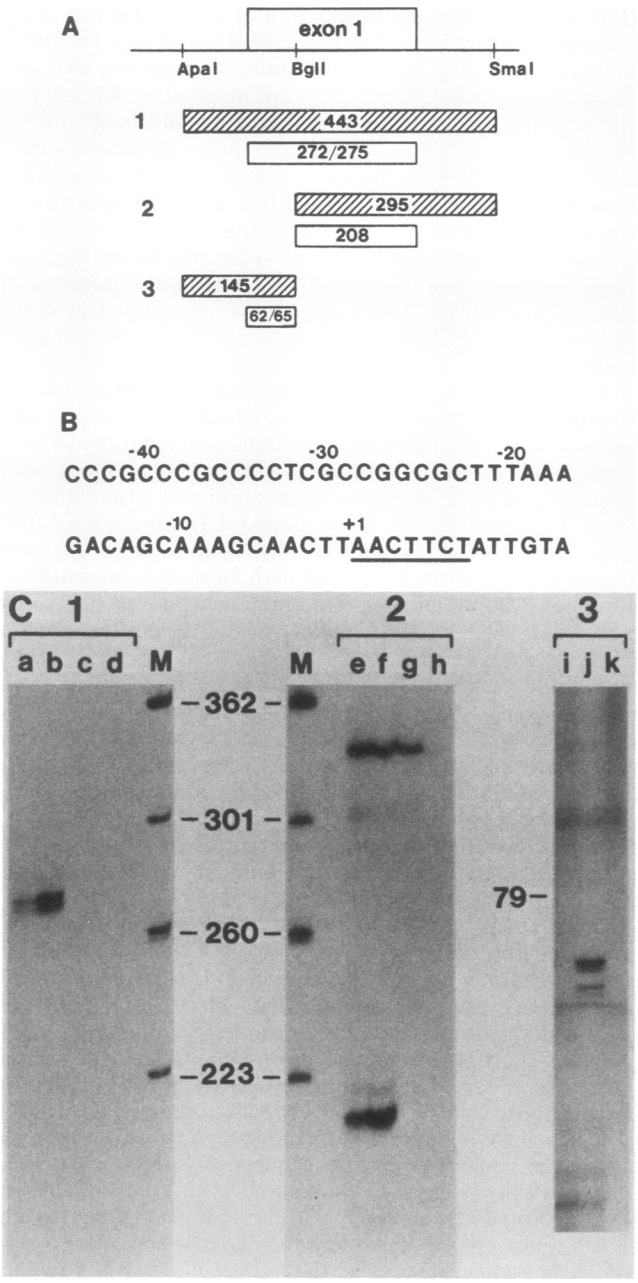

FIG. 3. RNA probe protection of the 5′ end of chicken c-*myc* mRNA. (A) The upper drawing is a schematic of exon 1 showing the restriction sites used for constructing plasmids containing probe sequences. The three numbered hatched bars underneath represent the RNA probes generated and are positioned with respect to the exon 1 map. The open bars display the probe fragments protected by the target mRNAs. They are placed relative to the exon 1 map according to our interpretation of the data. The lengths of the probes and the protected fragments (in bases) are shown within the bars. (B) The DNA sequence (40) is the genomic sequence surrounding the mapped RNA cap sites. The bar underneath indicates the region within which all RNA start sites occur. Numbering of the DNA sequence is based upon the assignment of +1 as the most 5′ of the possible RNA transcriptional start sites. (C) The three autoradiograms show the results of hybridization of labeled probe RNA to target RNAs. The numbers above the gels indicate which of the probes shown in panel A was used. Either 1 μg of RNA (LL6 mRNA, LL7 mRNA) or 3 μg of RNA (chicken embryo RNA, yeast RNA) was annealed to 10^5 dpm of ^{32}P-labeled probe RNA. After RNase digestion, the strands were denatured and run on either a 6% (probes 1 and 2) or an 8% (probe 3) polyacrylamide–urea gel. The resulting gel was dried and exposed to Cronex-4 film. Lanes: a and e, chicken embryo mRNA; b, f, and j, LL6 mRNA; c, g, and i, LL7 mRNA; d, h, and k, yeast RNA; M, ^{32}P-labeled *Hin*fI-*Bam*HI-digested pSJ118 DNA. The lengths of the marker fragments (in bases) are shown.

mapped by RNase protection experiments with multiple single-stranded RNA probes. Uniformly labeled RNA probes complementary to sequences encompassing the entire exon 1, as well as probes complementary to either the 5′ or 3′ part of the exon 1 (Fig. 3A) were annealed to normal chicken embryonic RNA, LL7 RNA, and LL6 RNA. Hybridization of the whole first exon probe (probe 1) to these RNAs confirmed that the 5′ untranslated sequences are derived from a single exon; hybridization of two nonoverlapping probes (probes 2 and 3), each containing part of the first exon, assigned the cap site and the splice donor site.

Hybridization of a ^{32}P-labeled RNA probe containing sequences from the exon 1 region (the 443-bp *Apa*I-*Sma*I fragment [Fig. 3A]) to embryonic mRNA (Fig. 3C, lane a) and LL6 mRNA (lane b) yielded two protected fragments of 272 and 275 bases. No hybridization of this probe occurred to either LL7 mRNA (lane c) or yeast RNA (lane d). The LL7-derived c-*myc* RNA did not anneal to any exon 1 probe (Fig. 3C, lanes c, g, and k) implying that the transcriptional start site lies outside this region in this tumor. Further mapping of the 5′ end of the LL7 c-*myc* transcript was not pursued.

To determine the splice donor site at the end of exon 1, a probe extending from the *Bgl*I site to the *Sma*I site (Fig. 3A) was annealed to chicken embryo RNA and LL6 mRNAs. Only a single species of identical length, 208 bases, was protected by both mRNAs (Fig. 3B, lanes e and f). The simplest interpretation is that the splice donor site is 208 bp from the *Bgl*I site, at position 1462 in reference 40, confirming the conclusion drawn from analysis of cDNA λ13 from the bursal tumor LL6 (Fig. 2). Again, no fragments were protected by RNA from the bursal tumor LL7 (lane g) or yeast cells (lane h).

The splice donor and acceptor sites that define the boundaries of the intron between exons 2 and 3 were initially placed by analogy to the v-*myc* sequence (1, 47). Comparison of the lengths of the *Sal*I-*Cla*I restriction fragment containing the putative spliced junction in the LL6 cDNA clone, λc13, and in v-*myc* revealed identical-length fragments, thus confirming usage of the same splice sites (data not shown).

The RNA cap sites of chicken embryo and LL6 c-*myc* RNAs were mapped more accurately by the following strategy. A 145-base *Apa*I-*Bgl*I RNA probe (Fig. 3A) was protected for lengths of approximately 62 and 65 bases by LL6 RNA (Fig. 3B, lane j) and chicken embryo RNA (data not shown). This places the 5′-most initiation site at position 1188 in reference 40. The two protected doublets may represent imprecise RNase digestion; alternatively, there could be four start sites with differential utilization. The 7-base sequence surrounding all the possible initiation sites is presented in Fig. 3B. Further definition was precluded because the RNases we used do not cleave 3′ linkages of rA. Previous mapping of the cap site by primer extension placed the RNA start sites approximately 50 bases (rather than our 60 to 67 bases) upstream of the *Bgl*I site (25) and with a pattern of RNA start sites very similar to that determined here by RNA probe protection, suggesting that both meth-

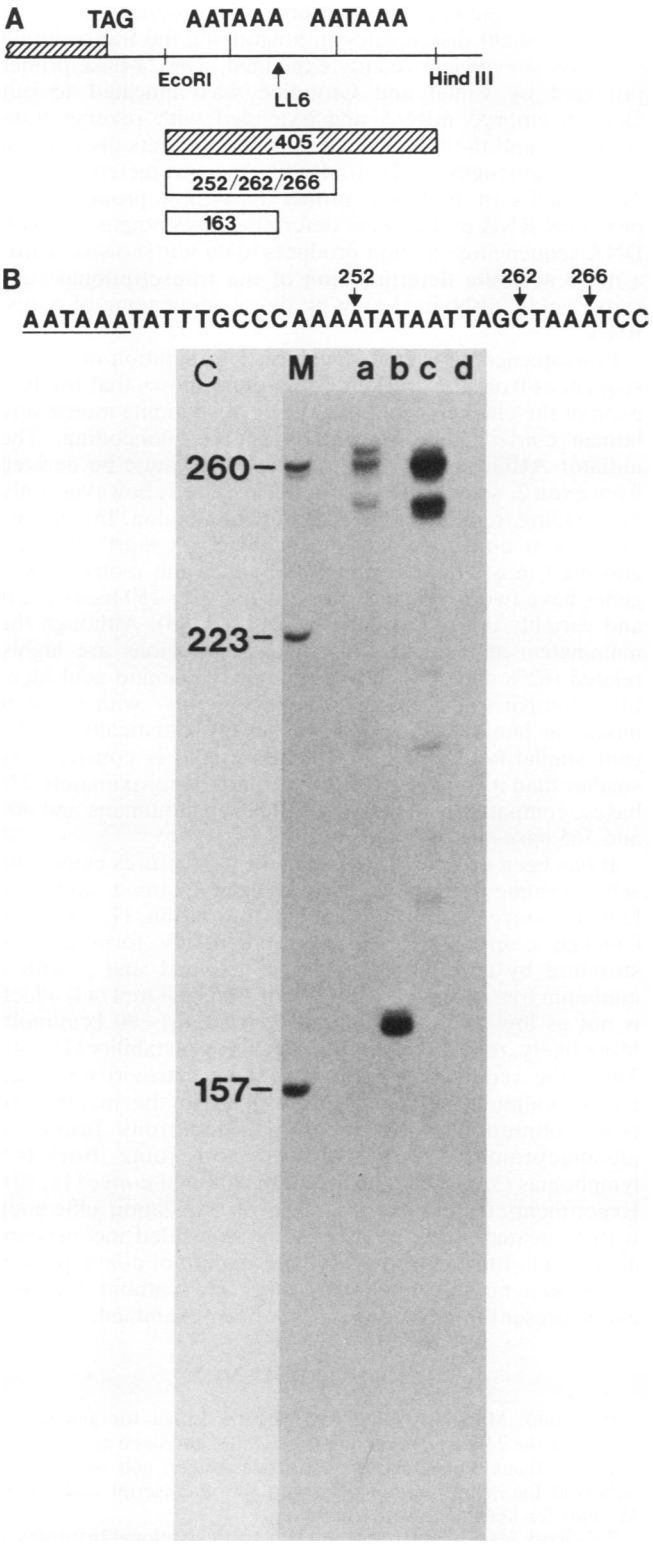

FIG. 4. RNA probe protection analysis of the 3′ end of c-*myc* mRNA. (A) The diagram shows the genomic region of the 3′ end of the c-*myc* gene. The hatched bar represents the 3′-most coding region with the TAG termination codon shown. The two potential polyadenylation signals are shown above the line, and the arrow below the line indicates the proviral integration point in the LL6 bursal lymphoma. The restriction sites shown mark the fragment used to generate the RNA probe. The hatched bar illustrates the probe with the length written in the bar. The open bars show the lengths of the protected fragments with the sizes indicated and are positioned relative to the 3′ region according to our interpretation of the data. (B) The DNA sequence shown is the genomic sequence at the 3′-most potential polyadenylation signal. The signal sequence AATAAA is underlined, and the numbered arrows indicate the 3′ penultimate probe-protected base of each protected fragment. Numbering of the sequence is based upon the assignment of the *Eco*RI cleavage site as number 1. (C) Autoradiogram shows the data resulting from hybridization of the ^{32}P-labeled probe to 3 μg of chicken embryo mRNA (lane a), 1 μg of LL6 mRNA (lane b), 1 μg of LL7 mRNA (lane c), and 3 μg of yeast RNA (lane d). Lane M contains ^{32}P-labeled *Hin*fI-*Bam*HI fragments of pSJ118. The lengths of the marker fragments (in bases) are shown.

ods detect the same 5′ ends. When we annealed the identical 24-base primer (kindly provided by M. Groudine and M. Linial) to our chicken embryo mRNA and extended it with reverse transcriptase, we obtained four products with a pattern and length consistant with the RNA probe protection experiments (data not shown).

Others (25, 40) have examined normal mRNA with probes encompassing sequences 4 to 5 kb upstream of the region detected by our probes for exon 1 and have not found any mRNAs. Taken together with other results (Fig. 3 and reference 25), it appears that we are detecting and mapping the 5′-most region of the normal c-*myc* transcript.

Characterization of polyadenylation site. Two consensus polyadenylation sequences, AATAAA, occur near the 3′ end of the c-*myc* coding region, 70 bases and 230 bases, respectively, downstream of the *Eco*RI site (Fig. 4A). The proviral integration in LL6 occurred between these two sequences, eliminating the distal site and producing a c-*myc*-virus fusion RNA (32). Therefore, either the first potential polyadenylation signal is not recognized normally or it is specifically bypassed or suppressed or both in the tumor. To clarify the status of the two polyadenylation sequences, RNAs from LL6, LL7, and chicken embryo were annealed to ^{32}P-labeled RNA corresponding to the noncoding strand of the *Eco*RI-*Hin*dIII fragment at the 3′ end of the normal chicken c-*myc* gene (Fig. 4A). Three protected fragments approximately 250 to 260 bases in length were obtained by hybridization to chicken embryo RNA (lane a) and LL7 RNA (lane c). This suggests that exon 3 ends slightly downstream of the second polyadenylation signal. No protected fragments that would indicate use of the proximal polyadenylation site were detected, even in long exposures of the gel (data not shown). Thus, the proximal site is used rarely, or not at all, in the tissues tested. LL6 RNA protects a fragment of 163 bases (lane b), consistent with the distance from the *Eco*RI site to the proviral integration point, which was independently determined by sequencing (unpublished observations).

More detailed analysis of the chicken embryo and LL7 RNA with a probe extending from 50 bases upstream of the polyadenylation signal to the *Hin*dIII site (data not shown) indicated that the three protected fragments correspond to discontinuities in the RNA at positions indicated by arrows above the sequence shown in Fig. 4B. In the absence of evidence for downstream exons or splice donor consensus sequences, we presume these sites are polyadenylation sites. Adenylations at 252 and 266 bases from the *Eco*RI site correspond well to proposed addition sites, rNrA, where rN is preferentially rC (4, 17, 26, 36, 49), whereas putative poly(A) addition at 262 bases from the *Eco*RI site would not occur at a consensus polyadenylation sequence and may represent an unusual adenylation site.

DISCUSSION

Characterization of c-*myc* transcriptional elements. By using the general strategy of RNase protection of uniformly labeled probes combined with analysis of a LL6 c-*myc* cDNA clone, we mapped the elements necessary to form the chicken c-*myc* transcript. Several unusual features of chicken c-*myc* emerged from our work and that of others: uncommonly long 5′ and 3′ untranslated regions, including a noncoding first exon; the absence of consensus promoter sequences in the region 5′ of exon 1; and two potential polyadenylation signals in exon 3, one of which is used exclusively.

We mapped at least two transcriptional initiation sites of c-*myc* to within a 7-bp region. The region upstream of the cap sites lacks consensus sequences for the common promoter-associated elements, the TATA box (6; D. Goldberg, Ph.D. thesis, Stanford University, Stanford, Calif., 1979) and the CAAT box (4, 12), although the sequence TTTAAA found approximately 20 bases upstream of the start site may substitute for TATA sequences. Instead, the region 5′ of the RNA start site, which is likely to be part of the promoter, is found to be extremely GC rich, averaging nearly 80% GC over the 250 bases upstream of the start site. Included are several CCGCCC repeats which bear resemblance to part of the 21-bp repeat of the simian virus 40 promoter region that serves as a part of the binding site for the Sp1 transcription factor in vitro (14). Other genes, including those for adenosine deaminase (45), hypoxanthine phosphoribosyltransferase (27), and 3-hydroxy-3-methyl glutaryl coenzyme A reductase (35), have similar promoter regions.

Polyadenylation of chicken c-*myc* occurs exclusively in response to one of two potential AATAAA signals in the noncoding portion of exon 3. Factors governing the polyadenylation following AATAAA sequences are the subject of intensive work (17, 26, 36, 49), but it is not apparent why only one signal is utilized in this case. Both the normal and LL7 c-*myc* RNAs are polyadenylated in response to the same signal. The LL6 provirus is integrated 5′ to the authentic polyadenylation sequence, and the polyadenylation signal in the viral sequences is utilized.

Three closely spaced potential sites of polyadenylation were found downstream of the AATAAA signal. Two of these occur at consensus adenylation sequences (the more usual one being rCrA), whereas the third site conflicts with established nucleotide pattern. We are uncertain whether this site represents a novel polyadenylation site or is an artifact of nuclease digestion. Isolation of cDNAs that include the poly(A) tail could confirm that the protected fragments denote polyadenylation sites.

Our proposal for the structure of exon 1 is in disagreement with a report from other investigators (25) who describe two overlapping exons based upon S1 mapping of c-*myc* RNA with uniformly labeled DNA. They place the RNA initiation sites approximately 15 bases downstream and the splice donor of the major first exon approximately 80 bases downstream of those described in this report. The minor first exon they describe most likely corresponds to the first exon of 272 to 275 bases presented in Fig. 3. We are reasonably confident of our placement of the splice donor site (Fig. 3) because it was based on identical RNA probe protection of samples from both normal and tumor cells, as well as the nucleotide sequence analysis of a cDNA clone from LL6. Several reasons could account for the discrepancy of the splice donor site, including the presence of substantial quantities of unprocessed RNA in their samples, incomplete S1 nuclease digestion, or perhaps polymorphic splice donor sites. The relatively slight discrepancy in positioning the transcription initiation sites is less readily explained. The 24-base primer provided by Linial and Groudine was annealed to our chicken embryo mRNA and extended with reverse transcriptase, and the visualized extension products displayed a pattern and length similar to the RNAase-protected species. The lengths of both the primer extension products and protected RNA probes were determined by comparison with DNA sequencing reaction products (data not shown). Thus, a more accurate determination of the transcriptional start sites could be obtained only by direct sequencing of c-*myc* RNA.

Consequences of exon 1 structure. Examination of exon 1 sequences from the chicken c-*myc* gene shows that the first exon of the chicken gene, like the corresponding mouse and human c-*myc* (3, 5, 41, 42), is entirely noncoding. The initiator AUG for $p58^{c\text{-}myc}$ in c-*myc* RNA must be derived from exon 2, since there is no ATG in exon 1; however, only one reading frame contains a termination codon. In contrast to chicken c-*myc*, in which several RNA start sites are clustered in a 7-base region, the human and mouse c-*myc* genes have two RNA start sites located over 150 bases apart and variably utilized in different cells (3, 43). Although the mammalian and chicken c-*myc* coding regions are highly related (62% nucleotide identity and 73% amino acid identity), comparison of exon 1 of chicken c-*myc* with those of mouse or human c-*myc* failed to reveal statistically significant similarity. Exon 1 of chicken c-*myc* is considerably smaller than its mammalian counterparts, approximately 275 bases, compared with 390 and 550 bases in humans and 400 and 565 bases in the mouse.

It has been proposed that secondary structures caused by self-annealing sequences derived from exons 1 and 2 of human c-*myc* RNA can inhibit translation (7, 30, 37). Chicken c-*myc* RNA can also potentially form a stem structure by annealing portions of exons 1 and 2 with a minimum free energy of ΔG^0 (25°) of −45 kcal/mol (44) which is not as low as in the human RNA stem (−90 kcal/mol). More likely, removal of exon 1 affects RNA stability (15, 34). Thus, the removal of exon 1 in most retrovirus-induced bursal lymphomas (11, 37), as well as in the majority of c-*myc*-immunoglobulin locus translocations found in plasmacytomas (3, 20, 41, 43) and some Burkitt's lymphomas (3, 5), may enhance translation of c-*myc* (11, 37). Experiments to address the issue of translation efficiency with truncated c-*myc* transcripts have yielded inconsistant data (7, 11, 30). Unfortunately, the amount of c-*myc* protein in chicken bursal lymphomas with and without the first exons present in RNA has not yet been examined.

ACKNOWLEDGMENTS

We thank Mark Groudine and Maxine Linial for generously providing the 24-base primer and for friendly and open communication. We thank Jackie Papkoff, Titia de Lange, and Mary Anne Schofield for helpful comments about the manuscript and Janine Marinos for assistance with the typing.

This work was supported by grants from the National Institutes of Health and the American Cancer Society. C.N. is a postdoctoral fellow of the American Cancer Society, California Division. H.E.V. is an American Cancer Society Research Professor.

LITERATURE CITED

1. **Alitalo, K., J. M. Bishop, D. H. Smith, E. Y. Chen, W. W. Colby, and A. D. Levinson.** 1983. Nucleotide sequence of the v-*myc* oncogene of avian retrovirus MC29. Proc. Natl. Acad.

Sci. USA **80**:100–104.
2. **Alitalo, K., G. Ramsay, J. M. Bishop, S. Ohlsson, W. W. Colby, J. P. McGrath, and A. D. Levinson.** 1983. Identification of nuclear proteins encoded by viral and cellular *myc* oncogenes. Nature (London) **306**:274–277.
3. **Battey, J., C. Moulding, R. Taub, W. Murphy, T. Stewart, H. Potter, G. Lenoir, and P. Leder.** 1983. The human c-*myc* oncogene: structural consequences of translocation into the IgH locus in Burkitt lymphoma. Cell **34**:779–787.
4. **Benoist, C., K. O'Hare, R. Breathnach, and C. Chambon.** 1980. The ovalbumin gene-sequence of putative control regions. Nucleic Acids Res. **8**:127–142.
5. **Bernard, O., S. Cory, S. Gerondakis, E. Webb, and J. Adams.** 1983. Sequence of the murine and human cellular *myc* oncogenes and two modes of *myc* transcription resulting from chromosome translocation in B lymphoid tumors. EMBO J. **2**:2375–2383.
6. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
7. **Butnick, N., C. Miyamoto, R. Chizzonite, B. R. Cullen, G. Ju, and A. M. Skalka.** 1985. Regulation of the human c-*myc* gene: 5′ noncoding sequences do not affect translation. Mol. Cell. Biol. **5**:3009–2016.
8. **Campisi, J., H. E. Gray, A. B. Pardee, M. Dean, and G. E. Sonenshein.** 1984. Cell-cycle control of c-*myc* but not c-*ras* expression is lost following chemical transformation. Cell **36**:241–247.
9. **Chamberlin, M., and J. Ring.** 1973. Characterization of T7-specific ribonucleic acid polymerase. I. General properties of the enzymatic reaction and the template specificity of the enzyme. J. Biol. Chem. **6**:2235–2244.
10. **Corcoran, L. M., J. M. Adams, A. R. Dunn, and S. Cory.** 1984. Murine T lymphomas in which the cellular *myc* oncogene has been activated by retroviral insertion. Cell **137**:113–122.
11. **Darveau, A., J. Pelletier, and N. Sonenberg.** 1985. Differential efficiencies of *in vitro* translation of mouse c-*myc* transcripts differing in the 5′ untranslated region. Proc. Natl. Acad. Sci. USA **82**:2315–2319.
12. **Dierks, P., A. van Ooyen, M. D. Cochran, C. Oobkin, J. Reiser, and C. Weissman.** 1983. Three regions upstream from the capsite are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell **32**:695–706.
13. **Donner, P., I. Greiser-Wilke, and K. Moelling.** 1982. Nuclear localization and DNA binding of the transforming gene product of avian myelocytomatosis virus. Nature (London) **296**:262–265.
14. **Dynan, W. S., and R. Tjian.** 1983. The promoter-specific transcription factor SP1 binds to upstream sequences in the SV40 early promoter. Cell **35**:79–87.
15. **Eick, D., M. Piechaczyk, B. Henglein, J.-M. Blanchard, B. Traub, E. Kofler, S. Wiest, G. Lenoir, and W. Bornkamm.** 1985. Aberrant c-*myc* RNAs of Burkitt's lymphoma cells have longer half-lives. EMBO J. **4**:3717–3725.
16. **Eisenman, R. N., C. Y. Tachibana, H. D. Abrams, and S. R. Hann.** 1985. v-*myc*- and c-*myc*-encoded proteins are associated with the nuclear matrix. Mol. Cell. Biol. **5**:114–126.
17. **Fitzgerald, M., and T. Shenk.** 1981. The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell **24**:251–260.
18. **Fung, Y. K. T., L. Crittenden, and H. J. Kung.** 1981. Orientation and position of avian leukosis virus DNA relative to the cellular oncogene c-*myc* in B-lymphocyte tumors of highly susceptible $15I_5X7_2$ chickens. J. Virol. **44**:742–746.
19. **Fung, Y. K. T., A. Fadly, L. Crittenden, and H. J. Kung.** 1981. On the mechanism of retrovirus-induced avian lymphoid leukosis: deletion and integration of the proviruses. Proc. Natl. Acad. Sci. USA **78**:3418–3422.
20. **Gerondakis, S., S. Cory, and J. M. Adams.** 1984. Translocation of the *myc* cellular oncogene to the immunoglobulin heavy chain locus in murine plasmacytomas is an imprecise reciprocal exchange. Cell **36**:973–982.
21. **Hayward, W. S., B. G. Neel, and S. M. Astrin.** 1981. Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis. Nature (London) **209**:475–479.
22. **Kelly, K., B. H. Cochran, C. D. Stiles, and P. Leder.** 1983. Cell-specific regulation of the c-*myc* gene by lymphocyte mitogens and platelet-derived growth factor. Cell **35**:603–610.
23. **Levy, L. S., M. B. Gardner, and J. W. Casey.** 1984. Isolation of a feline provirus containing the oncogene *myc* from a feline lymphosarcoma. Nature (London) **308**:853–856.
24. **Li, Y., C. A. Holland, J. W. Hartley, and N. Hopkins** 1984. Viral integrations near c-*myc* in 10–20% of MCF 247-induced AKR lymphomas. Proc. Natl. Acad. Sci. USA **81**:6808–6811.
25. **Linial, M., and M. Groudine.** 1985. Transcription of three c-*myc* exons is enhanced in chicken bursal lymphoma cell lines. Proc. Natl. Acad. Sci. USA **82**:53–57.
26. **McDevitt, M. A., M. J. Imperiale, H. Ali, and J. R. Nevins.** 1984. Requirement of a downstream sequence for generation of a poly(A) addition site. Cell **37**:993–999.
27. **Melton, D. W., D. S. Konecki, J. Brennard, and C. T. Caskey.** 1984. Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. Proc. Natl. Acad. Sci. USA **81**:2147–2151.
28. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
29. **Neil, J. C., D. Hughes, R. McFarlane, N. M. Wilkie, D. E. Onions, G. Lees, and O. Jarrett.** 1984. Transduction and rearrangement of the *myc* gene by feline leukemia virus in naturally occurring T-cell leukemias. Nature (London) **308**:814–820.
30. **Nilsen, T. W., and P. A. Maroney.** 1984. Translational efficiency of c-*myc* mRNA in Burkitt lymphoma cells. Mol. Cell. Biol. **4**:2235–2238.
31. **Noori-Daloii, M. R., R. A. Swift, H. J. Kung, L. B. Crittenden, and R. L. Witter.** 1981. Specific integration of REV proviruses in avian bursal lymphomas. Nature (London) **294**:544–546.
32. **Payne, G. S., J. M. Bishop, and H. E. Varmus.** 1982. Multiple arrangements of viral DNA and an activated host oncogene c-*myc* in bursal lymphomas. Nature (London) **295**:209–217.
33. **Persson, H., and P. Leder.** 1985. Nuclear localization and DNA binding properties of a protein expressed by human c-*myc* oncogene. Science **225**:718–721.
34. **Rabbitts, P. J., A. Forster, M. A. Stinson, and T. H. Rabbitts.** 1985. Truncation of exon 1 from the c-*myc* gene results in prolonged c-*myc* mRNA stability. EMBO J. **4**:3727–3733.
35. **Reynolds, G. A., S. K. Basu, T. F. Osborne, D. J. Chin, G. Gil, M. S. Brown, J. L. Goldstein, and K. L. Luskey.** 1984. HMG CoA reductase: a negatively regulated gene with unusual promoter and 5′ untranslated regions. Cell **38**:275–286.
36. **Sadofsky, M., and J. C. Alwine.** 1984. Sequences on the 3′ side of hexanucleotide AAUAAA affect efficiency of cleavage at the polyadenylation site. Mol. Cell. Biol. **4**:1460–1468.
37. **Saito, H., A. D. Hayday, K. Wiman, W. S. Hayward, and S. Tonegawa.** 1983. Activation of the c-*myc* gene by translocation, a model for translational control. Proc. Natl. Acad. Sci. USA **80**:7476–7480.
38. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
39. **Selten, G., H. T. Cuypers, M. Ziljlstra, C. Melief, and A. Berns.** 1984. Involvement of c-*myc* in MuLV-induced T cell lymphomas in mice: frequency and mechanisms of activation. EMBO J. **3**:3215–3222.
40. **Shih, C.-K. M. Linial, M. M. Goodenow, and W. S. Hayward.** 1984. Nucleotide sequence 5′ of the chicken c-*myc* coding region: localization of a noncoding exon that is absent from *myc* transcripts in most avian leukosis virus-induced glycoproteins. Proc. Natl. Acad. Sci. USA **81**:4697–4701.
41. **Stanton, L. W., P. D. Fahrlander, P. M. Tesser, and K. Marcu.** 1984. Nucleotide sequence comparison of normal and translocated murine c-*myc* genes. Nature (London) **310**: 423–425.
42. **Stanton, L. W., R. Watt, and K. B. Marcu.** 1983. Translocation, breakage and truncated transcripts of c-*myc* oncogene in murine plasmacytomas. Nature (London) **303**:401–405.
43. **Taub, R., C. Moulding, J. Battey, W. Murphy, T. Vasicek, G. M. Lenoir, and P. Leder.** 1984. Activation and somatic

mutation of the translocated c-*myc* gene in Burkitt lymphoma cells. Cell **36**:339–348.

44. **Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–41.
45. **Valerio, D., M. G. C. Duyvesteyn, B. M. M. Dekker, G. Weeda, T. M. Berkvens, L. van der Voorn, H. van Ormondt, and A. J. van der Eb.** 1985. Adenosine deaminase: characterization and expression of a gene with a remarkable promoter. EMBO J. **4**:437–443.
46. **Varmus, H. E.** 1984. The molecular genetics of cellular oncogenes. Annu. Rev. Genet. **18**:553–612.
47. **Watson, D. K., E. P. Reddy, P. H. Duesberg, and T. S. Papas.** 1983. Nucleotide sequence analysis of the chicken c-*myc* gene reveals homologous and unique coding regions by comparison with the transforming gene of avian myelocytomatosis virus MC29, delta-gag-*myc*. Proc. Natl. Acad. Sci. USA **80**: 2146–2149.
48. **Watt, R., L. W. Stanton, K. B. Marcu, R. C. Gallo, C. M. Croce, and G. Rovera.** 1983. Nucleotide sequence of cloned cDNA of human c-*myc* oncogene. Nature (London) **303**: 725–728.
49. **Wickens, M., and P. Stephenson.** 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent mRNA 3′ end formation. Science **226**:1045–1051.