JOURNAL OF VIROLOGY, Oct. 1982, p. 158–166
0022-538X/82/100158-09$02.00/0

Vol. 44, No. 1

Molecular Analysis of the c-*myc* Locus in Normal Tissue and in Avian Leukosis Virus-Induced Lymphomas

BENJAMIN G. NEEL,[1]* GREGORY P. GASIC,[1] CHARLES E. ROGLER,[2] ANN M. SKALKA,[3] GRACE JU,[3] FUMIO HISHINUMA,[3] TAKIS PAPAS,[4] SUSAN M. ASTRIN,[2] AND WILLIAM S. HAYWARD[1]†

The Rockefeller University, New York, New York 10021[1]; The Institute for Cancer Research, The Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111[2]; The Roche Institute of Molecular Biology, Nutley, New Jersey 07110[3]; and Laboratory of Tumor Virus Genetics, National Cancer Institute, Bethesda, Maryland 20205[4]

Received 19 March 1982/Accepted 10 June 1982

We isolated molecular clones of the provirus-host cell junctions (tumor junction fragments) from two avian leukosis virus-induced lymphomas and compared the structures of these clones with a clone of the normal c-*myc* gene. Restriction mapping and DNA sequencing demonstrated that normal proviral integration events occurred adjacent to c-*myc* in both tumors, without gross structural alteration of c-*myc*. The right long terminal repeat of an avian leukosis virus provirus is integrated upstream from the bulk of the c-*myc* coding sequences and oriented such that transcription can initiate within the long terminal repeat and proceed downstream into c-*myc*. A comparison of a tumor junction fragment with the v-*myc* gene showed that there are two regions of v-*myc*-related sequences (which are probably exons) separated by 1 kilobase of sequences unrelated to v-*myc* (probably an intron). A DNA sequence analysis of the tumor junction fragments suggested that integration had occurred in exons adjacent to splice donor sites. This suggests that there are additional exons and introns in c-*myc*. Based on these findings, a model is proposed for the genesis of the tumor-specific RNAs containing viral 5′ and c-*myc* information in avian leukosis virus-induced lymphomas.

The avian leukosis viruses (ALV) are a group of slowly transforming RNA tumor viruses that usually cause B-cell lymphomas in infected birds after a latent period of 4 to 12 months. Slowly transforming viruses lack specific transforming genes (*onc* genes), and until recently the mechanism by which these viruses induce neoplasia was obscure. These viruses differ from the rapidly transforming viruses, such as Rous sarcoma virus and the avian acute leukemia viruses, which contain transforming genes in their genomes (v-*onc* genes). All of these transforming genes are homologous to, and presumably derived from, normal cellular counterparts (c-*onc* genes) that are present in the genomes of all vertebrates. Rapidly transforming viruses usually cause neoplasia within a few weeks after infection and transform appropriate target cells in tissue culture. A wide variety of biochemical and genetic evidence suggests that rapidly transforming viruses induce neoplasia as a result of constitutive expression of v-*onc* genes or high expression of v-*onc* genes or both.

In the vast majority of ALV-induced lymphomas, oncogenesis results from the occasional chance integration of an ALV provirus adjacent to a specific cellular gene (12, 26, 28). In nearly all cases, this gene is c-*myc*, the cellular counterpart of v-*myc*, the transforming gene of MC29 virus (14). This finding has been confirmed by workers in other laboratories (7, 12, 27). The ALV provirus has its structural and replicative genes *gag*, *pol*, and *env* flanked by long terminal repeats (LTRs). The LTRs contain sequences that are implicated in the initiation of transcription by RNA polymerase II (8, 17, 41), and normal viral mRNA synthesis initiates in the left LTR. Integration of an ALV provirus upstream from c-*myc* permits an LTR to activate downstream transcription of c-*myc*, with the production of new mRNAs containing ALV 5′-terminal and c-*myc* information. Such transcripts are found in most ALV-induced lymphomas (14). However, recent evidence suggests that in some cases the LTR can activate transcription of c-*myc* by mechanisms that do not appear to involve transcription that initiates in the LTR (27). In any case, the constitutive expression of c-*myc* at high levels presumably results in neoplasia.

† Present address: Sloan-Kettering Institute for Cancer Research, New York, NY 10021.

To explore the mechanism of ALV oncogenesis in more detail, we compared the structures of the c-*myc* locus in normal tissue and in ALV-induced lymphomas. In this report we describe the molecular cloning of provirus-host cell tumor junction fragments from two independent lymphomas that each contained a portion of a single ALV provirus (26). Also, a clone of the c-*myc* gene of normal cells was selected from a chicken genomic library. The structures of these clones were compared with each other and with a clone containing the v-*myc* gene. Mapping and DNA sequence analysis showed that in the two tumors used a normal proviral integration event had occurred upstream from the bulk of the c-*myc* coding sequences and without gross structural rearrangement of c-*myc*. The ALV provirus in both tumors is oriented such that the viral LTR can promote downstream transcription of c-*myc*.

MATERIALS AND METHODS

Cells and viruses. *Escherichia coli* ED8654 (25) was used for the isolation and propagation of recombinant bacteriophage. A lambda library of Rhode Island Red chicken genomic DNA (10) was obtained from R. Axel, Columbia University. The bacteriophage vector λgtWES · λB (19) was used to construct recombinant bacteriophage. This vector was digested with *Eco*RI and *Sac*I, and the *Eco*RI arms were purified from the dispensable λB fragments by sucrose density gradient centrifugation.

Restriction endonuclease digestions. All restriction enzymes except *Cla*I were obtained from New England Biolabs or Bethesda Research Laboratories; *Cla*I was obtained from Boehringer Mannheim Corp. The conditions used for digestion were those described by the manufacturers, except that excess enzyme (2 to 5 U/μg) was usually used.

Molecular cloning of tumor junction fragments. Tumor DNA was digested to completion with *Eco*RI. Approximately 600 μg of digested DNA from a splenic nodule in bird 7 (26) was fractionated by discontinuous electrophoresis in a "gene machine" (11). The gel contained 0.5% agarose (Seakem LE; Marine Colloids) in E buffer (40 mM Tris, 5 mM sodium acetate, 1 mM EDTA, pH 7.8). Electrophoresis was at 40 V for 15 min, and this was followed by a collection and filling cycle of 5 min. One fraction was collected every 20 min, and the procedure was carried out for 40 h. Three fractions in the expected molecular weight range (~3 kb kilobases [kb]) were precipitated and used separately in ligation reactions. *Eco*RI-digested tumor 10 DNA was size fractionated by electrophoresis in low-melting-point agarose (Sea Plaque; Marine Colloids). The techniques used for isolating DNA from low-melting-point agarose have been described previously (38). DNA from the appropriate size range was pooled and used for ligation. The insert DNAs were ligated to the *Eco*RI arms of λgtWES · λB (1 μg). The molar ratio of arms to insert was 1:0.2. Details of the procedures used for in vitro packaging (3, 15), screening (4), phage purification, amplification, and preparation of phage DNA have been described previously (16). The efficiencies of packaging of recombinant bacteriophage ranged from 10^5 to 10^6 infectious particles per μg of phage DNA. Recombinant bacteriophage (200,000 to 400,000 phage) were plated onto 150-mm petri dishes (20,000 phage per plate) and screened. Before hybridization, filters containing phage DNA imprints were prewashed at 37°C in 5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) containing 20 mM Tris-hydrochloride (pH 7.4) and 0.5% sodium dodecyl sulfate for at least 45 min. Hybridization was to a 5′ probe (see below) and was carried out at 37°C in a solution containing 50% formamide, 5× SSC, 20 mM Tris-hydrochloride (pH 7.4), 0.5% sodium dodecyl sulfate, and at least 200 μg of yeast RNA per ml for 18 to 24 h. A 10-ml sample of hybrid mixture containing 2 × 10^5 Cerenkov counts per min of 5′ probe was used for each filter. After hybridization, the filters were washed three times briefly in 0.2× SSC containing 0.2% sodium dodecyl sulfate and then incubated once or twice (45 min each time) at 37°C in the same solution. Washed filters were dried and exposed to Kodak XR-5 or XAR-5 film for 4 to 24 h with an intensifying screen. Positive plaques were purified three times before amplification.

Isolation of c-*myc* clones. Approximately 4 × 10^5 recombinant bacteriophage from the chicken genomic library were plated onto 150-mm petri dishes. Screening with the appropriate probe (see below) and phage isolation and purification were performed as described above.

Preparation of ^{32}P-labeled DNA probes. A cDNA probe homologous to the first 101 nucleotides of Rous-associated virus-2 (5′ probe) was prepared as described previously (13, 26). High-specific-activity nick-translated probes were prepared by the method of Maniatis et al. (21), using aqueous [α-^{32}P]dCTP (>2,000 Ci/mmol; Amersham Corp.). We routinely obtained specific activities of 1 × 10^8 to 2 × 10^8 Cerenkov counts per min per μg. Low-specific-activity radiolabeled restriction fragments were prepared by nick-translation for use in restriction mapping; the procedure was the same as that used for preparing high-specific-activity probes, except that the reaction buffer contained all four deoxynucleoside triphosphates at concentrations of 20 μM and one-half the amount of [α-^{32}P]dCTP was used per reaction. The specific activities obtained ranged from 5 × 10^5 to 5 × 10^6 Cerenkov counts per min per μg.

Nucleic acid blot hybridizations. DNA samples were electrophoresed in submerged horizontal agarose (Seakem) gels (0.8 to 2%, depending on the size range of the fragments to be analyzed) in E buffer. ^{32}P-labeled *Hin*dIII fragments or φX174 *Hae*III fragments were used as molecular weight markers. DNA was blotted onto nitrocellulose filters by the technique of Southern (36). The prewashing and hybridization conditions used were the same as those described above for filters containing phage DNA imprints.

Polyadenylated RNA was analyzed by electrophoresis followed by transfer to diazobenzyloxymethyl paper (2) or nitrocellulose paper (39) and by hybridization as described previously (26).

Restriction mapping. Appropriate inserts from molecular clones were purified by preparative electrophoresis in low-melting-point agarose. These fragments were nick-translated to low specific activity. Mapping was performed by standard procedures involving di-

gestions with one or a combination of restriction enzymes, followed by electrophoresis on agarose gels. The gels were dried onto DEAE paper and autoradiographed. The size values reported below are accurate to ±50 nucleotides.

DNA sequencing. All DNA sequence determinations were performed by the method of Maxam and Gilbert (22, 23). Both strands of DNA were sequenced. The λtjf-7 and λtjf-10 clones (containing tumor junction fragments from birds 7 and 10, respectively) were digested with *Eco*RI, and the inserts were purified on low-melting-point agarose gels. The inserts were labeled with ^{32}P at their 5′ ends by using [α-^{32}P]ATP (>3,000 Ci/mmol; Amersham Corp.) and T4 polynucleotide kinase or at their 3′ ends by using [α-^{32}P]dATP and the Klenow fragment of *E. coli* DNA polymerase I. After digestion with *Sac*I, the smaller fragments (containing LTR sequences) were isolated for sequencing.

RESULTS

Molecular cloning of tumor junction fragments. Tumor DNAs from birds 7 and 10 (26) were digested with restriction endonuclease *Eco*RI, blotted onto nitrocellulose filters (36), and hybridized to a probe specific for the first 101 nucleotides of ALV RNA (5′ probe). As described previously, *Eco*RI fragments that represent ALV provirus-host cell junctions (junction fragments) can be identified by virtue of their hybridization to 5′ probe but not to a probe representative of the total ALV genome. The tumor from bird 7 had a 3.1-kb junction fragment, and the tumor from bird 10 had a 3.6-kb junction fragment. Both tumors expressed 2.5-kb tumor-specific polyadenylated RNAs that contained ALV 5′ and *myc* information (14, 26).

Based on this size analysis, each *Eco*RI-digested tumor DNA was size fractionated, ligated to *Eco*RI-digested λgtWES · λB (19), packaged in vitro (3, 15), and screened by the procedure of Benton and Davis (4), using a 5′ probe. Two of the clones obtained, λtjf-7 and λtjf-10, were chosen for further analysis. λtjf-7 contains a 3.1-kb *Eco*RI fragment that hybridizes to the 5′ probe (Fig. 1A, arrow); this fragment is the tumor junction fragment. In addition, this clone contains a 4.8-kb *Eco*RI fragment that does not hybridize to the 5′ probe (Fig. 1A, asterisk); this fragment was not analyzed further. λtjf-10 contains the expected 3.6-kb tumor junction fragment, which hybridizes to the 5′ probe (data not shown).

The 3.1-kb *Eco*RI fragment of λtjf-7 hybridizes to the v-*myc* probe, as well as to the 5′ probe (Fig. 1A). A similar result was obtained for the 3.6-kb λtjf-10 insert (data not shown). This confirms unambiguously that viral 5′ and c-*myc* sequences are covalently linked in the DNAs of these tumors.

Isolation of a clone for the normal c-*myc* gene. A chicken genomic library (10) was kindly provided by Richard Axel. This library consists of 15- to 20-kb fragments from a partial *Hae*III-*Alu*I digestion of chicken DNA that have been tailed with *Eco*RI oligonucleotide linkers and inserted into bacteriophage Charon 4A (6). Recombinants were screened by the Benton-Davis method. The probe used was the 2.4-kb fragment of a *Sac*I digest of the λtjf-7 insert. This fragment contains c-*myc* sequences but not ALV sequences (Fig. 2 and 3). Several clones that hybridized to this probe were obtained, and one of these, λc-*myc*-2, was selected for further

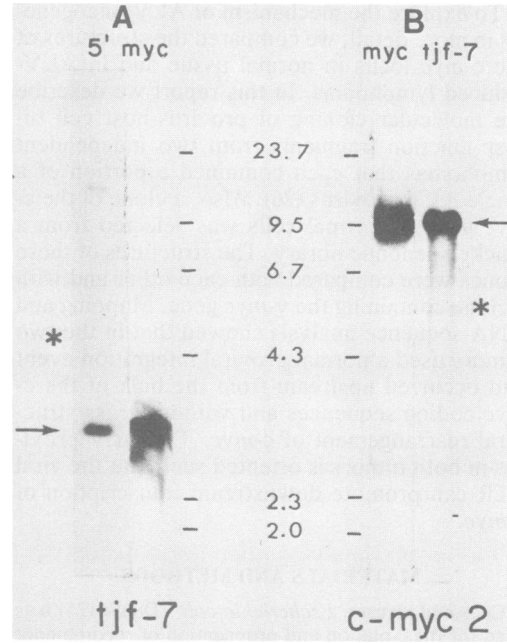

FIG. 1. Molecular cloning of a tumor junction fragment and the normal c-*myc* gene. (A) The *Eco*RI-generated tumor junction fragment from a lymphoma in bird 7 was molecularly cloned into λgtWES (19) to give clone λtjf-7. Bacteriophage DNA was isolated, digested with *Eco*RI, transferred to a nitrocellulose filter (36), and hybridized to either a 5′ probe or a v-*myc* probe (see text). The arrow indicates the 3.1-kb tumor junction fragment. The asterisk indicates the position of a 4.8-kb band, identified by ethidium bromide staining, which was present in the λtjf-7 clone but did not hybridize to the 5′ or *myc* probe. The *Hin*dIII fragments of lambda DNA were used as size markers. Numbers indicate length in kilobases. (B) A molecular clone containing the c-*myc* gene from normal chickens was selected from a chicken genomic library (see text). DNA from this clone (λc-*myc*-2) was digested with *Eco*RI, transferred to nitrocellulose, and hybridized to either a v-*myc* probe or to the 2.4-kb *Sac*I fragment of the λtjf-7 clone, which lacks ALV sequences (see text and Fig. 2). The arrow indicates the *myc*-containing *Eco*RI fragment. The position of a 5.8-kb fragment of λc-*myc*-2 that did not contain *myc* information is indicated by the asterisk.

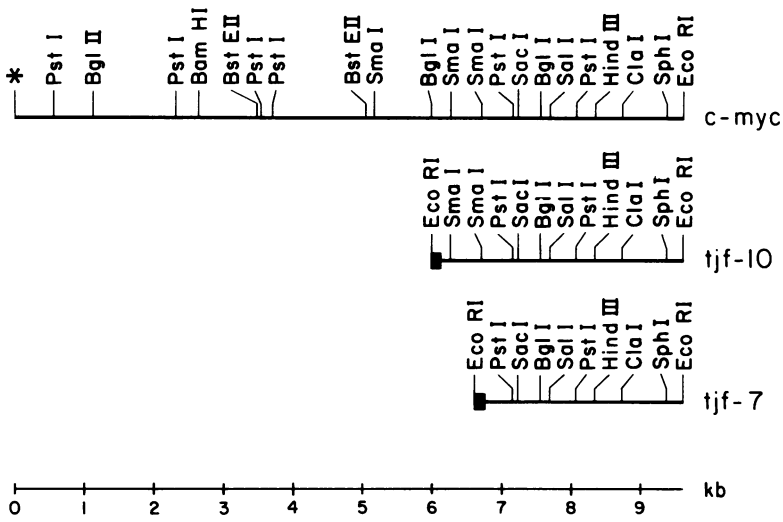

FIG. 2. Restriction maps of tumor junction fragments: comparison with the c-*myc* gene. The 3.1-kb *Eco*RI insert of λtjf-7 and the 3.6-kb insert of λtjf-10 were isolated by preparative electrophoresis, as was the 9.6-kb *myc*-specific fragment from λc-*myc*-2. After low-specific-activity labeling of the fragments with [α-^{32}P]dCTP, the restriction endonuclease cleavage maps of these inserts were determined and compared. The LTR sequences in the tumor junction fragments are represented by solid boxes. The left end of the c-*myc* clone (asterisk) is defined by either a *Hae*III site or an *Alu*I site. The cleavage positions are accurate to within 50 nucleotides. There are no *Kpn*I, *Pvu*I, *Xba*I, or *Xho*I sites in any of the clones.

analysis because it contained the largest *myc*-specific fragment. An *Eco*RI digest of λc-*myc*-2 produced two restriction fragments in addition to the Charon 4A arms, a 9.6-kb fragment that contains all of the c-*myc* sequences homologous to v-*myc* (Fig. 1B, arrow) and a 5.5-kb fragment that does not contain *myc*-related sequences (Fig. 1B, asterisk). The restriction map of the 9.6-kb insert (see below) corresponds to earlier restriction data on the c-*myc* locus (33; Astrin, unpublished data).

Restriction mapping of tumor junction fragment and c-*myc* clones. To compare the structures of the c-*myc* locus in normal tissue and ALV-induced lymphomas, restriction maps of the c-*myc* clone and two tumor junction fragment clones were prepared. A comparison of these maps (Fig. 2) revealed that (i) the tumor junction fragment clones were colinear with the right end of the c-*myc* clone (thus, proviral integration in birds 7 and 10 occurred without gross structural alteration of the c-*myc* locus) and (ii) the orientation of the integrated provirus in the tumor junction fragment clones was such that the LTR (Fig. 2, solid boxes) could be used to initiate transcription at the viral initiation site that would proceed downstream into c-*myc*. Orientation of the provirus was confirmed by digestion of the tumor junction fragment clones with *Sac*I, followed by Southern blotting and hybridization to 5′ or *myc* probes. As expected (Fig. 2), the left *Sac*I fragment (0.7 kb) hybridized both to the 5′ probe and to the v-*myc* probe, whereas the right *Sac*I fragment (2.4 kb) hybridized only to the *myc* probe (data not shown).

Localization of putative coding domains in c-*myc*. The c-*myc* and v-*myc* genes share extensive nucleotide sequence homology (31, 32). It is generally thought that MC29, which contains v-*myc*, arose as a result of recombination between a slowly transforming virus and c-*myc*. The v-*myc* gene is translated as part of a *gag-myc* fusion protein in MC29-infected cells; thus, the entire v-*myc* gene may represent protein-encoding sequences (5). We reasoned that sequences in common between v-*myc* and c-*myc* might represent coding sequences in the c-*myc* locus. Therefore, we prepared a detailed restriction map of a clone containing the v-*myc* gene and a more detailed restriction map of the λtjf-10 insert than is shown in Fig. 2.

The clone containing the v-*myc* gene has a 2.9-kb *Bam*HI fragment of an integrated MC29 provirus (18) inserted into pBR322. The map of this insert is shown in Fig. 3. This insert contains some viral *gag* sequences at the left end, followed by the v-*myc* gene, and a small amount of sequence to the right of v-*myc*. A comparison of this insert with the λtjf-10 insert shows the following. (i) There are regions on the left (~0.7 kb) and right (~0.85 kb) of the λtjf-10 insert in which the restriction map corresponds exactly to the

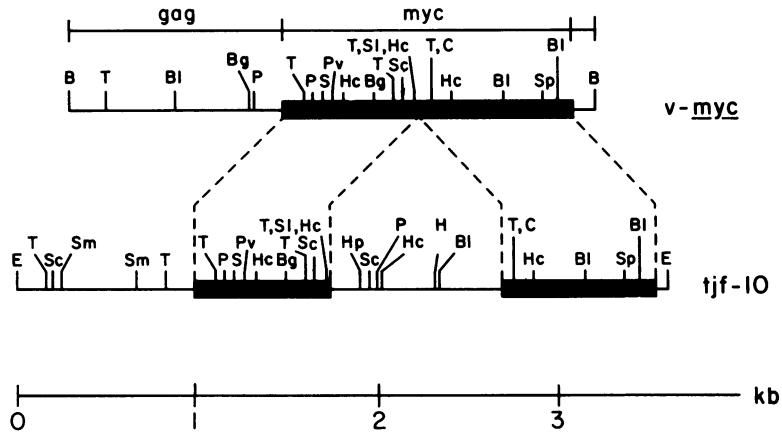

FIG. 3. Identification of putative coding and noncoding domains in c-*myc*. A more detailed restriction map of the 3.6-kb λtjf-10 insert was determined and was compared with the map of part of an integrated MC29 provirus containing the entire v-*myc* gene (18). The positions of the MC29 *gag* and *myc* genes are indicated above the clone. The regions of the two clones having identical restriction maps are shown as thick bars. Restriction sites were determined by using low-specific-activity-labeled insert DNAs, as described in the legend to Fig. 2. Abbreviations: B, *Bam*HI; Bg, *Bgl*I; Bl, *Bal*I; C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; Hc, *Hin*cII; Hp, *Hpa*I; P, *Pst*I; Pv, *Pvu*II; S, *Sac*I; Sc, *Sac*II; Sl, *Sal*I; Sm, *Sma*I; Sp, *Sph*I; T, *Taq*I.

left and right of v-*myc*. (ii) These two regions in λtjf-10 are separated by a region of about 1 kb that appears to be absent from the v-*myc* clone. The simplest interpretation of these results is that there is a 1-kb noncoding domain separating two coding domains. S1 deletion mapping experiments using the technique of Shenk et al. (34) with the v-*myc* and λtjf-10 clones indicated that the putative coding domains are 0.72 and 0.85 kb long, respectively, whereas the noncoding domain is 0.96 kb long (data not shown). (iii) The v-*myc* and λtjf-10 clones also diverge at their extreme left ends. This suggests that proviral integration in tumor 10 had occurred upstream from the bulk of the c-*myc* coding sequences. However, there may be coding sequences or leader sequences or both in c-*myc* that are not represented in v-*myc*.

Nucleotide sequencing of the left ends of the tumor junction fragment clones. The DNA sequences of the left ends of both the λtjf-7 and the λtjf-10 inserts were determined by the technique of Maxam and Gilbert (22, 23). These sequences have several interesting characteristics (Fig. 4). The first 152 nucleotides of both clones are identical to each other and to the known sequence of the right half of the viral LTR (17, 37). The last two nucleotides of the viral LTR (thymine-thymine at positions 100 and 101) are absent from the tumor junction fragment clones (cytosine-cytosine in λtjf-7 and guanine-guanine in λtjf-10). Deletion of the terminal two nucleotides of each LTR has been reported previously in sequence analyses of integrated proviruses (9, 17, 35, 40) and probably reflects a normal feature of the integration event. The deletion of LTR nucleotides in the tumor junction fragment clones suggests that a normal integration event had occurred. The sequence following the LTR is different in λtjf-10 and λtjf-7, since the proviruses in tumors 7 and 10 are integrated at different sites, which are about 500 nucleotides apart. As expected for a right LTR, the sequence following the LTR bears no resemblance to the viral sequence (37). Thus, no viral sequences appear to be interposed between the LTR and c-*myc*, so no other viral information besides the LTR-derived sequence should be present in tumor-specific transcripts, in agreement with previous results (14, 26). Interestingly, in both tumor junction fragment clones (Fig. 5) the LTR is located upstream from cellular sequence similar to the consensus splice donor sequence of eucaryotes (20, 30) and the adenovirus type 2 major late leader splice donor sequence (1, 42). If these are functional splice donor sites in vivo, then they may serve as splice sites for processing the primary transcript of the c-*myc* gene of normal cells. In this case, integration in both tumor 7 and tumor 10 has occurred in exons. In the λtjf-7 insert, there is an adenine-thymine-rich stretch beyond the putative splice donor site consisting of TTA and ATT repeats, which continues for about 35 nucleotides.

This finding suggests a possible explanation for why tumors with proviruses integrated at different sites upstream from c-*myc* produce

```
          -60          -50          -40          -30          -20
RSV    :  *******AAT TCCGCATTGC AGAGATATTG TATTTAAGTG CCTAGCTCGA
tjf-7  :        AAT TCCGCATTGC AGAGATATTG TATTTAAGTG CCTAGCTCGA
tjf-10 :        AAT TCCGCATTGC AGAGATATTG TATTTAAGTG CCTAGCTCGA

                   cap
          -10       1        10          20          30          40
RSV    :  TACAATAAAC GCCATTTGAC CATTCACCAC ATTGGTGTGC ACCTGGGTTG
tjf-7  :  TACAATAAAC GCCATTTGAC CATTCACCAC ATTGGTGTGC ACCTGGGTTG
tjf-10 :  TACAATAAAC GCCATTTGAC CATTCACCAC ATTGGTGTGC ACCTGGGTTG

                  50          60          70          80          90
RSV    :  ATGGCCGGAC CGTTGATTCC CTGACGACTA CGAGCACCTG CATGAAGCAG
tjf-7  :  ATGGCCGGAC CGTTGATTCC CTGACGACTA CGAGCACCTG CATGAAGCAG
tjf-10 :  ATGGCCGGAC CGTTGATTCC CTGACGACTA CGAGCACCTG CATGAAGCAG

                 100         110         120         130         140
RSV    :  AAGGCTTCAT TTGGTGACCC CGACGTGATA GTTAGGGAAT AGTGGTCGGC
tjf-7  :  AAGGCTTCAC CCCGCTGAGC TGGGGAGGGG GTGAGGCGGG GGGCTCACGA
tjf-10 :  AAGGCTTCAG GTGAGTGGCG GGGCCGG

                 150         160         170         180         190
RSV    :  CACAGACGGC GTGGCGATCC TGTCTCCATC CGTCTCGTCT ATCGGGAGGC
tjf-7  :  GGGGGTCGTG CTCTTTTATT ATTATTATTA TTTTATTATT ATTTAACCAA
```

FIG. 4. Comparison of viral and tumor junction fragment DNA sequences. The DNA sequences of the two tumor junction fragments (λtjf-7 and λtjf-10) were determined by the technique of Maxam and Gilbert (22, 23). These sequences were aligned with the sequence of Schmidt-Ruppin A Rous sarcoma virus (RSV) (37) that begins to the right of the *Eco*RI site in the LTR. This viral strain was chosen for comparison because the tumors in birds 7 and 10 were induced by td103, a *src* deletion mutant of the Schmidt-Ruppin A strain. The sequences are numbered such that the cap site (initiation site for viral RNA synthesis) within the LTR is at position 1. The sequence of the positive strand is shown for each DNA. The sequences upstream from the cap site are given negative numbers, and those downstream are given positive numbers. The boldface letters indicate the positions of (i) the TATA box (positions −30 to −24), (ii) the LTR-cell junction (positions 100 and 101), showing deletion of the two terminal LTR nucleotides in the tumor junction fragment clones, and (iii) putative splice donor sites in λtjf-7 (positions 119 to 125) and λtjf-10 (positions 99 to 107).

new, tumor-specific RNAs of approximately the same size (2.5 to 2.9 kb) (14, 26). It may be that c-*myc* can be activated more efficiently when the provirus integrates into a c-*myc* exon. The exon could then provide a splice donor site to enable transcripts initiated in the LTR (which lacks a splice donor site) to be spliced to the bulk of c-*myc* coding information located further downstream. If splicing eliminated the majority of the sequence between the integration site and the 0.72-kb coding domain, the mRNAs from tumors with proviruses integrated at different sites could be about the same size. Overexposures of Northern blots of tumor polyadenylated RNA hybridized to *myc* or 5′ probes revealed larger RNAs, which appeared to be precursors of the 2.5- and 2.9-kb *myc*-containing mRNAs. We have performed preliminary analyses of the genetic content of these RNAs using restriction fragments of the c-*myc* locus that map upstream from the regions homologous to v-*myc*. The results of these studies suggest that there are additional exons and introns in the c-*myc* primary transcript (B. G. Neel, C.-K. Shih, and W. S. Hayward, unpublished data).

DISCUSSION

We isolated molecular clones containing provirus-host cell junctions from the DNAs of two independent ALV-induced lymphomas and compared the structures of these clones with the structure of a clone of the normal chicken c-*myc* gene. The results of restriction mapping (Fig. 2), Southern blot hybridizations (data not shown),

and DNA sequencing (Fig. 4) show that in the two tumors used, the right LTR of an ALV provirus has integrated adjacent to and upstream from the bulk of the c-*myc* coding sequences. Integration has occurred without gross structural rearrangement of the c-*myc* locus. The proviruses in these clones are oriented such that the LTR could activate downstream transcription of c-*myc*, as suggested previously (14, 26).

To help localize potential coding sequences in the c-*myc* locus, the λtjf-10 clone was compared with a clone containing the v-*myc* gene by restriction mapping. There are two regions of v-*myc*-related sequences (we feel that these are probably exons), separated by about 1 kb of sequences unrelated to v-*myc*, which is probably an intron. The exon sizes are 0.72 kb (left) and 0.85 kb (right). The total amount of coding information (~1.6 kb) shared between v-*myc* and c-*myc* is essentially identical to previously reported estimates of the size of the v-*myc* gene (24). Also, Robins et al. (29) have recently reported similar values for these putative c-*myc* exons based upon heteroduplex mapping of c-*myc*/v-*myc* hybrids. In the v-*myc*-related regions the restriction maps of the c-*myc* clone (derived from a Rhode Island Red chicken) and the tumor junction fragment clones (derived from White Leghorn chickens) are identical to each other and to the map of the v-*myc* clone. This suggests that c-*myc* information is highly conserved in chickens, which is in agreement with previous data (32).

It is formally possible that the 1 kb of sequence between the two coding regions in c-*myc* that is not present in v-*myc*, which we have suggested is an intron, is actually present in the tumor-specific mRNAs. However, this is highly unlikely. The size of the tumor-specific mRNA produced in tumors 7 and 10 is 2.5 kb (26). This RNA is not big enough to encode all of the information present in the 0.72-, 0.85-, and 0.96-kb regions in addition to 100 to 200 nucleotides of polyadenylic acid and at least 99 nucleotides from the ALV LTR. It seems highly unlikely that sequences in c-*myc* that are also present in v-*myc* are absent from c-*myc* mRNA. Current evidence (S. Braverman and W. S. Hayward, manuscript in preparation) suggests that recombination between RNA tumor viruses occurs after packaging of the heterologous RNAs into virus particles. It is likely that recombination between slowly transforming viruses and c-*onc* genes involves packaging of the c-*onc* mRNA. If recombination involves a processed RNA intermediate, introns could not be regenerated, and v-*myc* should only contain sequences present in c-*myc* mRNA.

Based on these studies, we have constructed a model of what is now known about the c-*myc* locus and its putative transcript after integration of an ALV provirus. The 0.85- and 0.72-kb exons account for about 1.6 kb of information in the 2.5-kb mRNA produced in tumors 7 and 10. Whether the mRNA proceeds past the right end of the 0.85-kb exon has not been established yet. However, Payne et al. (27) have presented some evidence that tumor-specific mRNAs may extend beyond the *Eco*RI site in c-*myc*. Also, the tumor-specific RNA contains at least 99 nucleotides from the ALV LTR and 100 to 200 nucleotides of polyadenylic acid. Therefore, most of the region from the LTR to the 0.72-kb exon probably is not present in the 2.5-kb RNA. Instead, there is probably at least one more exon and one more intron, both of unknown length, within this region of c-*myc*. The finding of putative splice donor sites in both the λtjf-7 and the λtjf-10 inserts suggests that there may be at least four exons and three introns in c-*myc*. The additional c-*myc* exons are probably small (not more than a few hundred nucleotides long) and may form part of a leader sequence in c-*myc* mRNA. The results of preliminary analyses of putative precursors to the 2.5- and 2.9-kb tumor-specific mRNAs are consistent with this interpretation.

Our data suggest that proviral integration in tumors 7 and 10 occurred upstream from all of the c-*myc* information that is found in v-*myc*, which presumably represents the bulk of the c-*myc* coding sequences as well. All other tumors which we studied had *Eco*RI tumor junction fragments at least 2.7 kb long (14, 26; Astrin, unpublished data). Assuming that integration in these tumors also occurred without gross alteration of c-*myc*, then integration in these tumors occurred upstream from these putative coding sequences. However, our data also suggest that there may be additional exons in c-*myc*. It is unclear whether these exons contain coding or

	splice ↓
Consensus:	A/C AGGTAAGT
Ad-2 Major Late leader:	TT*G*G**GT***C*AGTAC
tjf-7:	AG*C*G**GT***C*AGGCG
tjf-10:	TCAGG**T***C*AGTG

FIG. 5. Putative splice donor sites in the tumor junction fragments. In λtjf-7, a sequence strongly suggestive of a splice donor site is found approximately 20 nucleotides downstream from the LTR, whereas a similar sequence lies immediately adjacent to the LTR in λtjf-10 (see Fig. 4). The adenovirus type 2 (Ad-2) major late leader splice donor sequence (1, 42) and the consensus splice donor sequence (20) are shown for comparison. Agreement between the sequence of the tumor junction fragments and the consensus sequence is shown in boldface letters, whereas agreement with the adenovirus type 2 splice donor is shown in italics.

leader sequences. This issue cannot be resolved conclusively at present because the protein encoded by the c-*myc* gene has not been identified yet and the DNA sequence of the entire c-*myc* locus (and thus the start codon for c-*myc* protein) has not been determined. We cannot unambiguously state that there is no alteration in c-*myc* coding information in tumors 7 and 10. However, there are a few tumors that have considerably larger tumor junction fragments, on the order of 5 kb (Astrin, unpublished data). Recently, we have located putative RNA polymerase II initiation sites for c-*myc* RNA synthesis by means of in vitro transcription experiments, using the c-*myc* clone as the template. The furthest initiation site is 4.4 kb upstream from the *Eco*RI site at the right of c-*myc* (Neel and Hayward, manuscript in preparation). Thus, it is likely that, at least in these few tumors, proviral integration has occurred without any alteration of c-*myc*.

We initially proposed that the ALV provirus activates c-*myc* by inserting the strong viral promoter present in the LTR upstream from c-*myc*. The two tumors studied in detail here are clearly consistent with this mechanism. Payne et al. (27) have reported that other orientations of the LTR are also associated with activation of c-*myc*. In some tumors, the LTR is upstream from c-*myc* but in the opposite orientation, and in one tumor it is downstream from c-*myc*. We checked all of our previous tumors for which samples were still available. In about 85% of those tumors where the provirus was integrated adjacent to c-*myc*, proviral orientation was such that the LTR could initiate downstream transcription. In about 15% of the tumors it is still possible that alternative orientations exist. At present, we cannot explain why Payne et al. found such a high percentage (~40%) of alternative orientations. The only differences between the protocols of these authors and ours are in time of injection and virus strain used. Nevertheless, the fundamental lymphomagenic mechanism of ALV remains the activation of c-*myc* by insertion of the viral LTR.

ACKNOWLEDGMENTS

We thank Stuart Kahn and Richard Malavarca (Roche Institute of Molecular Biology) for help with the cloning of the tumor junction fragments, Richard Axel (Columbia University) for supplying the chicken genomic library, and Hidesaburo Hanafusa and Bernard Mathey-Prevot (Rockefeller University) for critical reading of the manuscript. Excellent technical assistance was provided by Nancy Goldberg.

This investigation was supported by Public Health Service grants from the National Cancer Institute to W.S.H. (CA16668 and CA18213) and S.M.A. (CA06927) and by a grant from the Flora E. Griffin Memorial Fund to W.S.H. B.G.N. is a biomedical fellow in the Medical Scientist Training Program of the National Institutes of Health.

LITERATURE CITED

1. **Akusjarvi, G., and U. Pettersson.** 1979. Sequence analysis of adenovirus DNA: complete nucleotide sequence of the 5' noncoding region of adenovirus 2 hexon messenger RNA. Cell **16**:841–850.
2. **Alwine, J. C., D. J. Kemp, and G. R. Stark.** 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Natl. Acad. Sci. U.S.A. **74**:5350–5354.
3. **Becker, A., and M. Gold.** 1975. Isolation of the bacteriophage lambda A-gene protein. Proc. Natl. Acad. Sci. U.S.A. **72**:581–585.
4. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science **196**:180–182.
5. **Bister, K., M. J. Hayman, and P. K. Vogt.** 1977. Defectiveness of avian myelocytomatosis virus MC29: isolation of long-term nonproducer cultures and analysis of virus-specific polypeptide synthesis. Virology **82**:431–448.
6. **Blattner, F. R., B. G. Williams, A. E. Blechl, K. D. Thompson, H. E. Faber, L.-A. Furlong, D. J. Grunwald, D. O. Kiefer, D. D. Moore, J. W. Schumm, E. L. Sheldon, and O. Smithies.** 1977. Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. Science **196**:161–169.
7. **Cooper, G. M., and P. E. Neiman.** 1981. Two distinct candidate transforming genes of lymphoid leukosis virus-induced neoplasms. Nature (London) **292**:857–858.
8. **Czernilofsky, A. P., W. DeLorbe, R. Swanstrom, H. E. Varmus, J. M. Bishop, E. Tischer, and H. M. Goodman.** 1980. The nucleotide sequence of a nontranslated but conserved domain at the 3' end of the avian sarcoma virus genome. Nucleic Acids Res. **13**:2967–2984.
9. **Dhar, R., W. L. McClements, L. W. Enquist, and G. F. Vande Woude.** 1980. Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. Proc. Natl. Acad. Sci. U.S.A. **77**:3937–3941.
10. **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.
11. **Edgell, M. H., and F. Polsky.** 1980. Use of preparative gel electrophoresis for DNA fragment isolation. Methods Enzymol. **65**:319–327.
12. **Fung, Y.-K. T., A. M. Fadly, L. B. 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. U.S.A. **78**:3418–3422.
13. **Hayward, W. S., S. B. Braverman, and S. M. Astrin.** 1980. Transcriptional products and DNA structure of endogenous avian proviruses. Cold Spring Harbor Symp. Quant. Biol. **44**:1111–1121.
14. **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) **290**:475–480.
15. **Hohn, B., and K. Murray.** 1977. Packaging of recombinant DNA molecules into bacteriophage particles in vitro. Proc. Natl. Acad. Sci. U.S.A. **74**:3259–3263.
16. **Ju, G., L. Boone, and A. M. Skalka.** 1980. Isolation and characterization of recombinant DNA clones of avian retroviruses: size heterogeneity and instability of the direct repeat. J. Virol. **33**:1026–1033.
17. **Ju, G., and A. M. Skalka.** 1980. Nucleotide sequence analysis of the long terminal repeat (LTR) of avian retroviruses: structural similarities with transposable elements. Cell **22**:379–386.
18. **Lautenberger, J. A., R. A. Schulz, C. F. Garon, P. N. Tsichlis, and T. Papas.** 1981. Molecular cloning of avian myelocytomatosis virus (MC29) transforming sequences. Proc. Natl. Acad. Sci. U.S.A. **78**:1518–1522.
19. **Leder, P., D. Temier, and L. Enquist.** 1977. EK2 derivatives of bacteriophage lambda useful in the cloning of

DNA from higher organisms: the λgtWES system. Science **196**:175–177.
20. **Lerner, M. R., J. A. Boyle, S. M. Mount, S. L. Wolin, and J. Steitz.** 1980. Are snRNPs involved in splicing? Nature (London) **283**:220–224.
21. **Maniatis, T., A. Jeffrey, and D. G. Kleid.** 1975. Nucleotide sequence of the rightward operator of phage lambda. Proc. Natl. Acad. Sci. U.S.A. **72**:1184–1188.
22. **Maxam, A., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
23. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
24. **Mellon, P. A., A. Pawson, K. Bister, G. S. Martin, and P. H. Duesberg.** 1978. Specific RNA sequences and gene products of MC29 avian acute leukemia virus. Proc. Natl. Acad. Sci. U.S.A. **75**:5874–5878.
25. **Murray, N. E., W. J. Brammer, and K. Murray.** 1977. Lambdoid phages that simplify the recovery of in vitro recombinants. Mol. Gen. Genet. **150**:53–61.
26. **Neel, B. G., W. S. Hayward, H. L. Robinson, J. Fang, and S. M. Astrin.** 1981. Avian leukosis virus-induced tumors have common proviral integration sites and synthesize discrete new RNAs: oncogenesis by promoter insertion. Cell **23**:323–334.
27. **Payne, G. S., J. M. Bishop, and H. E. Varmus.** 1982. Multiple arrangements of viral DNA and an activated host oncogene in bursal lymphomas. Nature (London) **295**:209–214.
28. **Payne, G. S., S. A. Courtneidge, L. B. Crittenden, A. M. Fadly, J. M. Bishop, and H. E. Varmus.** 1981. Analyses of avian leukosis virus DNA and RNA in bursal tumors: viral gene expression is not required for maintenance of the tumor state. Cell **23**:311–322.
29. **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.
30. **Rogers, J., and R. Wall.** 1980. A mechanism for RNA splicing. Proc. Natl. Acad. Sci. U.S.A. **77**:1877–1879.
31. **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) **282**:452–455.
32. **Sheiness, D. K., and J. M. Bishop.** 1979. DNA and RNA from uninfected vertebrate cells contains nucleotide sequences related to the putative transforming gene of avian myelocytomatosis virus. J. Virol. **31**:514–521.
33. **Sheiness, D. K., S. H. Hughes, E. Stubblefield, and J. M. Bishop.** 1980. The vertebrate homologue of the putative transforming gene of avian myelocytomatosis virus: characteristics of the DNA locus and its RNA transcript. Virology **105**:415–424.
34. **Shenk, T. E., C. Rhodes, P. W. J. Rigby, and P. Berg.** 1975. Biochemical method for mapping mutational alterations in DNA with S1 nuclease: the location of deletions and temperature-sensitive mutations in simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **72**:989–993.
35. **Shimotohno, K., S. Mizutani, and H. M. Temin.** 1980. Sequence of retrovirus provirus resembles that of bacterial transposable elements. Nature (London) **285**:550–554.
36. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
37. **Swanstrom, R., H. E. Varmus, and J. M. Bishop.** 1982. Nucleotide sequence of the 5′ noncoding region and part of the *gag* gene of Rous sarcoma virus. J. Virol. **41**:535–541.
38. **Takeya, T., H. Hanafusa, R. P. Junghans, G. Ju, and A. M. Skalka.** 1981. Comparison between the viral transforming gene (*src*) of recovered avian sarcoma virus and its cellular homolog. Mol. Cell. Biol. **1**:1024–1037.
39. **Thomas, P.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. U.S.A. **77**:5201–5205.
40. **Van Beveren, C., J. G. Goddard, A. Berns, and I. M. Verma.** 1980. Structure of the Moloney murine leukemia viral DNA: nucleotide sequence of the 5′ long terminal repeat and adjacent cellular sequences. Proc. Natl. Acad. Sci. U.S.A. **77**:3307–3311.
41. **Yamamoto, T., G. Jay, and I. Pastan.** 1980. Unusual features in the nucleotide sequence of a cDNA clone derived from the constant region of avian sarcoma virus messenger RNA. Proc. Natl. Acad. Sci. U.S.A. **77**:176–180.
42. **Zain, S., J. Sambrook, R. J. Roberts, W. Keller, M. Fried, and A. R. Dunn.** 1979. Nucleotide sequence analysis of leader segments in a cloned copy of adenovirus 2 fiber mRNA. Cell **16**:851–861.