Mechanism of *c-erbB* Transduction: Newly Released Transducing Viruses Retain Poly(A) Tracts of *erbB* Transcripts and Encode C-Terminally Intact *erbB* Proteins

M. A. RAINES,^{1,2†} N. J. MAIHLE,¹ C. MOSCOVICI,³ L. CRITTENDEN,⁴ AND H.-J. KUNG^{1*}

Department of Molecular Biology and Microbiology, School of Medicine, Case Western Reserve University, Cleveland, Ohio 44106¹; Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824²; Department of Tumor Virology, Veteran's Administration Hospital, Gainesville, Florida 32602³; and U.S. Department of Agriculture Regional Poultry Research Laboratory, East Lansing, Michigan 48823⁴

Received 20 January 1988/Accepted 23 March 1988

We have previously shown that avian leukosis virus (ALV) induces erythroblastosis by insertional activation of the *c-erbB* gene. In 25% of the ALV-induced leukemic samples we have analyzed, acute retroviruses that have captured the activated *erbB* oncogene were released. The unusually high frequency at which *erbB* transduction occurs makes this an ideal system for studying the mechanism of oncogene transduction. In addition, these leukemic samples provide a rich source for the isolation of novel *erbB*-transducing viruses. We report here our characterization of several new *erbB*-transducing proviruses. The 5' recombination points of all these viruses mapped to the same intron in which proviral insertions cluster, supporting the hypothesis that transduction begins with proviral insertion near the oncogene. The 3' recombination points usually occurred within the 3' untranslated region downstream from the termination codon of the *c-erbB* gene. Three of the *erbB*-containing proviruses were molecularly cloned and analyzed in detail. Two of them were capable of releasing acute viruses, and interestingly, both retained poly(A) tracts of *erbB* messages in their genomes. A stretch of six adenosine residues in the ALV *env* gene appeared to mediate the 3' recombination events required for the generation of these viruses. These data provide further insight into the mechanism by which oncogenes are transduced into retroviral genomes.

Avian leukosis virus (ALV) induces a variety of long-latency neoplasms in chickens (for reviews, see references 18, 24, and 34). Proviral insertional activation of host proto-oncogenes is a key step in ALV-induced oncogenesis. In ALV-induced erythroblastosis, *c-erbB*, a gene closely related to the human epidermal growth factor receptor, is the specific locus affected by proviral insertion (8, 25). We have previously shown that in this disease, the proviruses are clustered within a specific intron of *c-erbB*, which separates the exons coding for the extracellular ligand-binding domain from the transmembrane and intracellular kinase domains (25). A readthrough transcript that encompasses both viral genes and downstream *c-erbB* sequences is generated (10, 22). This transcript is processed by differential splicing and polyadenylation, giving rise to two major insertionally activated *c-erbB* (IA *c-erbB*) messages, 7.0 and 3.6 kilobases (kb) in size. These transcripts carry viral leader and *gag* sequences which are fused via splicing to the 3' half of the *c-erbB*-coding sequences. The resulting IA *c-erbB* proteins are truncated receptor molecules which lack a functional ligand-binding domain. While the majority of ALV-induced erythroleukemias we have studied were generated via proviral activation of the *c-erbB* gene, a fraction of the leukemias appear to carry processed or transduced *erbB* genes (8, 25). Oncogene transduction is a process whereby a host proto-oncogene is incorporated into a retroviral genome and is subsequently introduced into other cells. Remarkably, Miles and Robinson (21), using higher-titer ALV inocula,

have observed that as many as 50% of ALV-induced erythroblastosis samples are capable of releasing transducing *erbB* viruses. These figures are significantly higher than reported in other systems and make ALV-induced erythroblastosis a rich and appealing system for the study of the mechanism of oncogene transduction. At least 16 new *erbB*-transducing viruses have recently been identified (21 and this study). These new isolates, in contrast to most other transducing viruses, have a limited passage history, minimizing the probability of mutation during passage, thereby permitting investigation of the initial transduction process. In this report, we analyze the structure of the viral *erbB* junctions in several new transducing *erbB* proviruses. Three proviruses have been molecularly cloned and, in one case, sequenced in entirety. In contrast to the other extensively characterized *erbB*-transducing isolates, our studies reveal that most of our newly released viruses contain sequences encoding the entire carboxy terminus of the *c-erbB* protein. We have also found that identical ALV sequences are involved in recombination between two different isolates, and these sequences occur in the poly(A) tracts of the *erbB* transcripts. The implications of these structural features in *erbB* transduction will be discussed in the context of currently available models.

MATERIALS AND METHODS

Viruses, chickens, and induction of erythroleukemia. All of the ALV-induced erythroblastosis samples analyzed in this report were generated at the same time as those described in Raines et al. (25) and were induced by injecting 10² to 10³ infectious units of the RAV-1 strain of ALV into the peritoneum of 1-day-old line 15₁ chicks. Avian erythroblastosis virus (AEV)-induced erythroleukemia samples were simi-

* Corresponding author.

† Present address: Division of Hematology and Oncology, Department of Medicine, University of California, Los Angeles, CA 90024.

larly generated by using the ES-4 strain (kindly provided by P. K. Vogt) except that chicks were inoculated 2 weeks posthatching. Virus stocks derived from ALV-induced erythro leukemia samples were prepared by filtering 10% (wt/vol) liver homogenates (in 10% tryptose-phosphate broth) or were plasma samples. A portion (0.1 ml) of these stocks was injected intravenously into 13-day-old embryos or 1-day-old chicks from both 15₁ × 15I₄ or SPAFAS chicken lines. Erythroblastosis development and leukemic sample collection were as previously described (8, 25).

DNA and RNA analysis. DNA was isolated from tissues, digested with restriction enzymes, electrophoresed on agarose gels, transferred to nitrocellulose, hybridized to nick-translated DNA probes, and washed as previously described in Raines et al. (25).

Total cellular RNA was extracted from frozen tissue by using the guanidinium isothiocyanate-hot phenol method (19). Polyadenylated RNA was selected by oligo(dT) cellulose chromatography (1), electrophoresed in 1% formaldehyde-agarose gels (19), and electrotransferred to Gene-Screen (New England Nuclear Corp.). Hybridization to DNA probes and washing conditions were identical to those used for DNA blots.

Molecular cloning and nucleotide sequence analysis. Transduced proviruses were cloned from total cellular DNA derived from leukemic liver samples (9134 E1 and 9141) or infected chicken embryo fibroblasts (9134 S3). Partial *Eco*RI-digested DNAs were size selected on sucrose density gradients, ligated to *Eco*RI-*Bam*HI-digested EMBL-4 λ DNA (7), and packaged into bacteriophage. The appropriate recombinant phage were isolated, and their structures were established by restriction enzyme digestion and Southern blot analysis with ALV- and *erbB*-specific probes.

Appropriate fragments were subcloned into M13mp18 or M13mp19, and their nucleotide sequences were determined by using the dideoxy-chain termination method (27). DNA sequences were confirmed by generating overlapping clones and by determining the sequences of both strands.

RESULTS

Identification of the transducing proviruses. We have previously shown that the majority of erythroblastosis samples induced by ALV contain proviral insertions within the *c-erbB* locus (8, 25). The proviral insertion sites all map at or near the 5' end of the VB1 exon, an exon where homology with *v-erbB* begins. About 25% of the erythroblastosis samples, however, revealed the presence of *erbB*-transducing viruses. The present report is concerned specifically with the structural characterization of these proviruses.

Transducing *erbB* proviruses can be readily identified by their lack of introns. They have a restriction map more compatible with *c-erbB* cDNA or *v-erbB* than with the *c-erbB* genomic locus. Here, we present the *Bam*HI digestion data to illustrate the approach used to screen tumor samples for the presence of transducing proviruses. *Bam*HI cleaves *c-erbB* cDNA three times (Fig. 1C); an internal 0.72-kb *Bam*HI fragment (probe BB) was used as a probe. As shown in Fig. 1A, samples carrying the transduced *erbB* gene (lanes with tumor numbers) reveal an additional 0.72-kilobase-pair (kbp) fragment, whereas in an uninfected sample (lane N) or a sample with an insertionally activated *c-erbB* gene (lane IA), the exons covered by the BB probe remain in an unspliced form and are distributed over a 6.0-kb *Bam*HI fragment. Sample 9132 contains a *Bam*HI fragment slightly larger than 0.72 kb. More detailed analysis revealed

a structural rearrangement within the transduced *erbB* gene (data not shown).

The 0.72-kb *Bam*HI fragment represents a common diagnostic fragment for all the transducing *erbB* proviruses. To differentiate individual transducing proviruses, we used probe AB, which was derived from the 5' end (a 0.47-kbp *Apa*-*Bam* fragment) of the IA *c-erbB* cDNA (Fig. 1C) and which should detect the variable 5' virus-*erbB* junction fragments in the different proviruses. Indeed, junction fragments of different sizes (0.8 to 7.0 kb) are detected in transduced samples, and in several of them multiple bands are seen, suggesting that more than one transducing provirus is present in these samples.

Structure and expression of the transducing proviruses. We then analyzed the expression pattern of these transducing proviruses. A complex pattern of *erbB* transcripts is detected in many of the samples (Fig. 2A). This is in stark contrast to the simple pattern observed in leukemic samples containing an insertionally activated *c-erbB* locus (lane IA), in which only two transcripts, 7.0 and 3.6 kb in size, are seen (22). The complex expression pattern of the transducing proviruses, however, is not unexpected, since a given provirus can generate at least two transcripts (a genomic and a spliced message) and since several of the leukemic samples contain multiple transducing proviruses. Furthermore, some of them carry internal polyadenylation signals (see below), giving rise to additional *erbB* messages. Another striking difference between the transduced and insertionally activated samples is the relative level of *erbB* expression. Expression of *erbB* in the transduced *erbB* samples is, in general, about 10- to 20-fold higher than *erbB* expression in the IA *c-erbB* samples and is comparable to *erbB* expression in AEV-ES4-infected samples (lane AEV). This is most likely due to the fact that transduced *erbB* transcripts result from direct transcription from the 5' long terminal repeat (LTR), whereas additional readthrough through the 3' LTR is required to generate the IA *c-erbB* transcripts. This latter process has been estimated to occur in only 15% of 5' LTR-promoted viral transcripts (12).

To verify that the observed novel transcripts originate from the transducing proviruses, viral extracts from some of these samples (i.e., 9134, 8660, 8669, and 9141) were made and inoculated into embryos. Of the four samples tested, all but one (9141) induced erythroblastosis. RNA from the resulting erythroblastosis samples was compared with RNA from tumors obtained with the original inoculum. As shown in Fig. 2B, *erbB*-related transcripts of similar size are observed in the original samples (9134 E1, 8660 E1, 8669 E1), as well as in the newly generated leukemic samples (E2). This conclusively documents the viral nature of these transcripts.

5' and 3' junctions of the transduced *c-erbB* gene. To probe the possible mechanism underlying the frequent transduction of the *c-erbB* gene, we first identified the 5' and 3' boundaries of the *erbB* sequences incorporated into these viruses. Leukemic samples with fewer transducing proviruses were chosen for this study. S1 protection analysis of the leukemic RNAs was performed by using genomic probes spanning both the 5' and 3' boundaries (M. A. Raines, Ph.D. thesis, Michigan State University, East Lansing, 1987, and data not shown). Protected fragments provided a measure of the extent of intron and 3'-untranslated sequences incorporated into the individual viral genomes. A composite diagram of the extent of the 5' and 3' *c-erbB* sequences incorporated into several of these transducing viruses, based on restriction enzyme mapping and S1 analysis, is given in Fig. 3.

FIG. 1. Identification of transducing *erbB* proviruses in ALV-induced erythroblastosis. (A) Genomic DNA isolated from ALV-induced leukemic liver tissue was digested with *Bam*HI and analyzed by transferring to nitrocellulose and hybridizing with probe BB. N, Normal uninfected line 15₁ chicken DNA; IA, sample containing a proviral insertionally activated *c-erbB*. Numbered lanes indicate samples from individual ALV-infected birds selected for the presence of transducing proviruses. (B) Southern transfer of genomic DNA as illustrated in panel A, hybridized with probe AB. The 8.4- and 8.0-kb bands marked as *c-erbB* represent the endogenous *c-erbB* locus. As previously reported (25), the *c-erbB* locus is polymorphic in the region detected by probe AB. Thus, some birds are homozygous for the 8.4-kb band, while others are heterozygous as characterized by the 8.4- and 8.0-kb fragments. (C) Restriction map of the 7.0-kb insertionally activated *c-erbB* cDNA (18). Probes BB and AB correspond to the *Bam*HI and *Apal*-to-*Bam*HI fragments of the cDNA. Probe T (cf. Fig. 2) is derived from a 1.7-kb *Apal*-to-*Sac*I fragment and contains most of the *erbB*-coding sequence (■). A, *Apal*; B, *Bam*HI; S, *Sac*I; 3' UT, 3'-untranslated region (▨); pA1, first polyadenylation signal in *c-erbB*; pA2, second (downstream) polyadenylation site in *c-erbB*; LTR and ALV-derived sequences (□).

Some of the proviral junctions (shown in boldface) were further confirmed and extended by cloning and sequencing as described below. All samples except 9132 contained an intact carboxy-terminal coding region, a structural feature

distinct from the existing AEV-H and AEV-ES4 isolates (5, 28, 36). The 3' junctions of four of these proviruses actually map downstream from the first polyadenylation signal of *erbB*, pA1.

Molecular cloning and sequence analysis of the transducing proviruses. S1 analysis provides only an approximate mea-

FIG. 2. Expression of *erbB*-transduced viral RNAs in ALV-induced erythroblastosis and secondary leukemias. RNA samples (2 μ g of poly(A) RNA) isolated from the original leukemic samples (panel A and E1 series in panel B), or from secondary leukemias (E2 series in panel B) were resolved on 1% formaldehyde-agarose gels, electrophoretically transferred to GeneScreen (New England Nuclear), and hybridized with probe T (Fig. 1C). E2 samples are from birds with acute erythroblastosis induced by injection of the E1 leukemic extract (see text). AEV, Leukemic liver samples from a bird infected with AEV-ES4. For other abbreviations, see the legend to Fig. 1.

FIG. 3. A summary of the 5' and 3' boundaries of the *c-erbB* gene present in the transducing viruses. This is a composite summary based on S1 protection analysis, restriction enzyme mapping, cloning, and sequencing. The numbers in boldface represent data confirmed by cloning and sequencing of the proviruses. 9134E1, Breakpoint due to deletion rather than recombination. The dotted line adjacent to 9141 indicates the 3' boundary of 9141 proviral DNA clones. The AEV-ES4 and AEV-H boundaries are based on published sequences (5, 11, 28, 36). 3' UT, 3'-untranslated region.

FIG. 4. Structure of *c-erbB*-transducing proviral clones 9141 (A), 9134 E1 (B), and 9134 S3 (C). Viral and *erbB* sequences were located by restriction enzyme digestion and Southern transfer analysis (31) with ^{32}P -labeled probes corresponding to either viral sequences ($\square$), *erbB*-coding sequences ($\blacksquare$), or 3'-untranslated sequence of *erbB* (▨). Only the *Bam*HI (B) restriction map is shown here. The 9134 E1 and 9134 S3 intron sequences (▨) were determined by DNA sequence analysis, and the precise junction between viral sequences and *erbB* sequences is delineated by the arrow above the diagram of each provirus (9134 E1 and 9134 S3). The amount of *erbB* untranslated sequence and 3' recombination points (arrow below provirus, 9134 E1 and 9134 S3) were similarly determined. Both 9134 E1 and S3 have suffered internal deletions in the *c-erbB* sequence; the extent and location of the deletions have been determined and are indicated. The extent of the *erbB* intron region contained in provirus 9141 was estimated by S1 protection analysis. The virus and *erbB* splice donor (SD) and splice acceptor (SA) sites are shown, as well as the cryptic splice donor ('SD'; 22) and the two *c-erbB* polyadenylation signals (pA1 and pA2; 10). 3' UT, 3'-untranslated region.

sure of the contiguous sequences present at the 5' and 3' ends of *erbB* and does not distinguish breakpoints due to recombination with ALV sequences, deletions, or other structural alterations. Furthermore, minor transduced viral components, if they exist, are not likely to be revealed by this method. These limitations prompted us to isolate individual proviral clones for detailed characterization. Genomic libraries of leukemic samples were used to isolate clones representing the transducing proviruses 9141 and 9134 E1. In addition, during further propagation of 9134 leukemic extracts, a sarcomagenic isolate, 9134 S3, was obtained. The isolation and oncogenicity of 9134 S3 will be described in detail elsewhere. Here, we describe the junction structures of these cloned proviruses, to clarify the transduction process. Detailed restriction mapping and sequence analysis of these proviruses were performed, and in two cases (9134 E1 and S3), all of the *erbB*-derived sequences, including the sites of recombination, were determined.

Structure of provirus 9141. Provirus 9141 probably represents the largest transduced genome thus far identified (Fig. 4A). It is at least 12 kb long and contains 6.0 kb of ALV

sequence at its 5' end, interrupted by a small deletion of 1.0 kb in the *env* region. The 5' junction point between ALV and *erbB* lies in the intron region about 150 bp 5' of the VB1 exon. The entire *erbB*-coding sequences downstream from VB1 are retained in a completely spliced form. Interestingly, virtually the entire 3'-untranslated region of the 7.0-kb IA *c-erbB* transcript (Fig. 1C) is transduced, suggesting that the transcript coterminal with the 7.0-kb IA *c-erbB* transcript is the precursor engaged in recombination. Our clone terminates at a site 0.3 kb upstream of the second polyadenylation signal (pA2). The upstream polyadenylation signal, pA1, is retained. Since both LTRs are required for integration, we presumed there would be a 3' LTR at the 3' end of this provirus. Exhaustive screening, as well as Southern analysis, of 9141 leukemic DNA failed to provide any supporting evidence. It is, thus, likely that the 9141 provirus represents a vestige of a transducing virus which has lost its 3' LTR following integration. The defective nature of the 9141 provirus, together with its unusually large size, explains why extracts of 9141 failed to induce acute erythroleukemia.

The defective provirus appears to be responsible for the

FIG. 5. Nucleotide sequence of the 5' virus-*erbB* junctions of 9134 E1 and 9134 S3. Sequences surrounding the 5' recombination point of provirus 9134 E1, including a portion of the viral *gag* sequences and all of the *erbB* intron sequences, are shown. Vertical lines indicate the junction between viral and intron sequences in AEV-ES4 (5, 11), AEV-H (36), 9134 S3, and 9134 E1. Only nucleotides divergent (between 9134 E1 and AEV-ES4 AEV-H) are indicated. *, Deletion; SA, splice acceptor site.

development of the original 9141 leukemia. Southern analysis (Fig. 1B) reveals that the 9141 provirus appears to be the only transduced genome in the 9141 leukemic sample. This provirus contains multiple splice sites; at least eight transcripts can be generated via differential processing. The predicted sizes match closely those identified in Northern analysis. The two major transcripts, the 7.0- and 3.6-kb messages, are virtually identical in size and content to the oncogenic transcripts arising from the insertionally activated *c-erbB* locus. It is, therefore, not surprising that provirus 9141 is leukemogenic, despite the fact that no infectious virus can be produced.

The structure of provirus 9134 E1. The 9134 E1 provirus (Fig. 4B) is the predominant viral component of the original 9134 erythroblastosis sample and of leukemic samples generated through subsequent passages of 9134 extracts (Fig. 2B, lanes 9134). It has one of the smallest leukemic viral genomes thus far identified, with a total size of 4.5 kb. We have sequenced this virus in its entirety. The 5' ALV sequence is only 559 bp long, but it contains both regions previously implicated in efficient packaging of ALV RNA (17, 29). The breakpoint is situated 161 bp downstream of the *gag* AUG, near the second encapsidation signal (23). The ALV sequence is juxtaposed next to 326 bp of *erbB* intron sequence with no significant homology at the junction. As shown in Fig. 5, the *erbB* intron sequence retained in 9134 E1 is larger than in AEV-ES4 and AEV-H and differs at several nucleotide positions. We have also sequenced the relevant portion of an intron from a *c-erbB* genomic clone derived from line 15₁ chicken DNA and have found that this sequence is identical to the intron sequence associated with 9134 E1. Thus, the intron sequence of 9134 E1 is a faithful copy of the *c-erbB* locus from line 15₁ chickens. The *erbB*-coding sequence is also identical to that of the IA *c-erbB* cDNA, indicating that no point mutations have occurred. The most intriguing aspect of this provirus perhaps lies in its 3' end, where 929 bp of the 3'-untranslated region of the *c-erbB* locus adjoins 42 bp of the *env*-coding region of ALV (Fig. 4B). The recombination point is within a stretch of 23 adenosines (interrupted by 1 thymidine). This oligo(A) sequence is apparently derived from the polyadenylation sequence of an *erbB* transcript which used the pA2 signal. We have also sequenced the *env* gene of the RAV-1 provirus that was originally used to induce erythroblastosis. Most interestingly, the point of recombination lies within a stretch of six adenosines found in the RAV-1 genome. These

six adenosines can thus potentially serve as a template for homologous recombination during reverse transcription. A comparison of the entire 3'-untranslated region of 9134 E1 with the IA *c-erbB* cDNA sequence (10) reveals that the untranslated region has also sustained a large deletion of 3227 bp, including the pA1 signal. As a consequence, the virus is more compact and contains one internal polyadenylation signal (pA2). The 9134 E1 provirus is fully infectious and leukemogenic (Fig. 2B), despite the fact that there is an internal polyadenylation signal, pA2.

Structure of provirus 9134 S3. The 9134 S3 virus (Fig. 4C) was isolated from secondary leukemic extracts of birds infected with 9134 extracts. Although 9134 S3 was not detected in the original leukemic sample by DNA or RNA analysis, it became apparent upon selection for fibroblast transformation and sarcomagenesis. The 9134 S3 virus readily transforms fibroblasts and induces sarcomas. The structure and transforming properties of the 9134 S3 and 9134 E1 proviruses are distinct. The 9134 S3 virus has a much longer 5' extension of viral sequence (Fig. 4C). It includes intact *gag* and *pol* genes and a portion of *env*; the 5' viral sequences are followed by *erbB*-derived intron sequences. The 5' junction point contains an overlapping pentanucleotide, CACAG, common to both *env* sequences of ALV and to *erbB* intron sequences. These pentanucleotides are likely to be involved in homologous recombination. Most strikingly, the *erbB* sequence of 9134 S3 also terminates within a polyadenylated stretch and the 3' recombination point occurs in ALV *env* sequences identical to those shown for 9134 E1 (Fig. 4B). One important difference, however, is that the polyadenylation tract in 9134 S3 is derived from an *erbB* transcript which uses the pA1 rather than the pA2 signal. This suggests that 9134 E1 and 9134 S3 evolved through independent recombination events and that the homologous oligo(A) sequence may represent a frequent site of recombination.

DISCUSSION

Mechanisms of oncogene transduction. Viral oncogene transduction is a well-documented phenomenon, but the mechanism is still not fully understood. Several models have been proposed (e.g., references 3, 9, 13, 26, and 32). A multistep model that has received general acceptance (32) contains the following features: (i) proviral integration within or upstream of a proto-oncogene in the same transcriptional orientation; (ii) deletion of the 3' proviral DNA including the 3' LTR, thereby fusing the proto-oncogene to the viral transcriptional unit; (iii) transcription from the 5' LTR promoter to generate a fusion transcript encompassing both proviral and proto-oncogene sequences; (iv) packaging of the chimeric transcripts and wild-type viral genomes as heterodimers into virus particles; and (v) template switching by reverse transcriptase between the heterodimers during viral DNA synthesis, thereby restoring 3' viral sequences to the chimeric molecule. The overall consequences of this process are the incorporation of proto-oncogene sequences into the internal portion of the viral genome and retention of the terminal *cis*-acting viral elements essential for replication. This model predicts that the 5' joining of viral and cellular sequences occurs at the DNA level (step 2) and the 3' joining occurs at the RNA level (step 5). Most of the transducing viral genomes that have been analyzed support this model. They have 5' recombination points which lie within intron regions of a proto-oncogene (thus retaining a splice acceptor site) and 3' recombination points that are

situated within an exon region (for a review, see reference 35) or, in one case, within a poly(A) tract (14). Several interesting exceptions, however, have been noted. For instance, in some recently described viral *myc* and *src* variants (15, 20), the 5' recombination point coincides with an RNA splice junction. This suggests that RNA splicing across the viral and cellular sequences may serve as an alternative pathway for gene fusion (i.e., step 2). The observation of Herman and Coffin (13), that readthrough transcripts composed of the provirus and downstream cellular sequences can be efficiently packaged, suggests another interesting possibility. This readthrough transcript could serve as a precursor for recombination at both the 5' and 3' ends, presumably during reverse transcription. This model is attractive in its simplicity since no proviral deletion step need be invoked. It also predicts a higher frequency of oncogene transduction in cases in which readthrough transcription of the proto-oncogene is a necessary step, such as in *c-erbB* activation. In the following paragraphs, we discuss our structural analysis of the *erbB*-transducing viruses described in this report in the context of these models.

Mechanism of *c-erbB* transduction. (i) **Insertional activation as the first step in *c-erbB* transduction.** All the transduction models described above postulate that proviral insertion near cellular oncogenes is the first step involved in oncogene transduction. Yet, for most of the transducing viral oncogenes, it is difficult to experimentally recreate this rare evolutionary event. In ALV-induced erythroblastosis, both insertional activation and transduction are frequently observed, providing a unique opportunity to examine this issue. Our findings, that all the proviruses involved in the insertional activation of *erbB* are in the same transcriptional orientation as the *erbB* gene and that they cluster in an intron region which coincides with the 5' junction of all the transducing *erbB* viruses analyzed in this report, lend strong support to this thesis.

(ii) **Formation of the 5' junction: deletion of proviral and *erbB* sequences at the DNA level.** As discussed above, the second step in oncogene transduction involves the fusion of viral and *erbB* sequences into a single transcriptional unit. This is where the proposed models differ most significantly. The model of Herman and Coffin (13), which suggests that a readthrough virus-*erbB* transcript can serve as a precursor to transduction, is appealing and may account for the unusually high frequency of *erbB* transduction. The large size of the primary virus-*erbB* readthrough transcript makes it unlikely that such a transcript would be efficiently packaged. However, a partially processed *erbB* transcript (i.e., spliced *erbB* gene and unspliced viral genes) could provide a good substrate for subsequent recombination events. The model which suggests that 5' fusion can be accomplished by RNA splicing is also attractive for *erbB* transduction since *c-erbB* has been shown to be activated through splicing and generates virus-*erbB* fusion transcripts in all leukemic samples. We, however, have no evidence that this is the case. Our data, especially those based on direct cloning and sequence analysis, clearly indicate that most, if not all, transducing *erbB* viruses retain intron sequences and a 5' splice acceptor site. We interpret this to mean that efficient packaging of ALV-*erbB* transcripts requires a viral "encapsidation signal," which at least in part, resides in a region between the splice donor site and the 5' junction of 9134 E1 (23). Our results are consistent with the original transduction model (32), which proposed a fusion transcript generated by the deletion of proviral and *erbB* intron sequences at the DNA level. Examination of structures at the 5' junction, however,

does not indicate a unified mechanism for such a process. With 9134 S3, homologous recombination involving a pentanucleotide has occurred. By contrast, no obvious homology is detected at this junction in provirus 9134 E1.

(iii) **Formation of the 3' junction: template switching within poly(A) tracts by reverse transcriptase.** Once a deletion of the 3' proviral sequence has occurred, the ALV-*erbB* fusion transcript can be generated by initiation at the 5' viral LTR promoter and termination at one of two *erbB* polyadenylation sites, pA1 or pA2. By virtue of maintaining the encapsidation sequence, this fusion message should be effectively packaged into virions together with the wild-type ALV RNA, known to be amply abundant in the leukemic cells that harbor more than one provirus. Upon subsequent infection, a switch between RNA templates by reverse transcriptase permits the incorporation of 3' viral sequence into the fusion transcript. It is most interesting that in two different isolates (9134 E1 and 9134 S3), the template switch occurs within the poly(A) tract of the *erbB* transcripts and at an identical site in the ALV genome. The six adenosine residues present in the ALV *env* gene apparently facilitate the switching process. While we cannot completely rule out the possibility that 9134 E1 and 9134 S3 arose through a common ancestral virus, the distinct 5' junctions and the different poly(A) tracts used in 3' recombination strongly suggest that they evolved through independent transduction events. The presence of a poly(A) tract within a retroviral genome as demonstrated here, together with a similar finding in the PRCII virus (14), provides strong evidence that the 3' recombination event takes place at the RNA level. Both 9134 E1 and S3 contain internal polyadenylation signals, a feature that in theory should impede the transcription of viral genomic RNA. Literature reports on this point, however, are not converging. Some internal polyadenylation signals appear to be detrimental to virus production, whereas others have no effect (2, 16, 30, 33). We have found that in several cases, including 9134 E1-infected cells, RNA transcripts prematurely terminating at the internal polyadenylation site could be identified (unpublished data), yet this virus is still produced at moderately high levels.

It is also significant that unlike AEV-H and AEV-ES4, all but one of these newly transducing *erbB* proviruses retain the C-terminal *erbB* coding sequences, including the putative major tyrosine autophosphorylation sites (6, 22). This further suggests that loss of the C terminus of *erbB* is not required for *c-erbB* transduction and is more likely imposed by selection for fibroblast transformation (4, 21).

In summary, our analysis of six newly generated *erbB* transducing viruses reveals that the 5' junction point often occurs in a specific intron region of *erbB* and retains a functional splice acceptor site. This suggests that the proviral insertionally activated *erbB* transcripts, despite their abundance in leukemic cells, are probably not the immediate precursors for transduction. Their 3' junction points reside primarily in the 3' untranslated region and, in two cases, within polyadenylation tracts of the *erbB* transcripts. This reinforces the view that the second recombination event occurs at the RNA level and is facilitated by homologous sequences.

ACKNOWLEDGMENTS

We thank C. Boerkoel for insightful discussion, J. M. Bishop for communicating AEV-ES4 sequences before publication, and Karen Fatica for careful preparation of the manuscript.

Nita J. Maihle is a recipient of a National Institute of Health postdoctoral fellowship. This work is supported by a Public Health

Service grant to C.M. from the National Institutes of Health (CA 10697), by the Medical Research Service of the Veterans Administration, and by a Public Health Service grant to H.-J.K. (CA 39207). This work was also supported by the Case Western Reserve University Cancer Center Core grant (P30 CA43703). H.-J.K. is a recipient of an American Cancer Society Faculty Research Award.

LITERATURE CITED

- Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA* **69**:1408-1412.
- Bandyopadhyay, P. K., and H. M. Temin. 1984. Expression of complete chicken thymidine kinase gene inserted in a retrovirus vector. *Mol. Cell. Biol.* **4**:749-754.
- Besmer, P., J. E. Murphy, P. C. George, F. Qui, P. J. Bergold, L. Lederman, H. W. Snyder, D. Bradeur, E. E. Zuckerman, and W. D. Hardy. 1986. A new acute transforming feline retrovirus and relationship of its oncogene v-kit with the protein kinase gene family. *Nature (London)* **320**:415-421.
- Beug, H., M. J. Hayman, M. B. Raines, H.-J. Kung, and B. Vennström. 1986. Rous-associated virus 1-induced erythroleukemic cells exhibit a weakly transformed phenotype in vitro and release *c-erbB*-containing retroviruses unable to transform fibroblasts. *J. Virol.* **57**:1127-1138.
- Choi, O.-R., C. Trainor, T. Graf, H. Beug, and J. D. Engel. 1986. A single amino acid substitution in *v-erbB* confers a thermolabile phenotype to ts167 avian erythroblastosis virus-transformed erythroid cells. *Mol. Cell. Biol.* **6**:1751-1759.
- Downard, J., P. Parker, and M. Waterfield. 1984. Autophosphorylation sites in the epidermal growth factor receptor. *Nature (London)* **311**:483-485.
- Frischauf, A., H. Lehrach, A. Poustka, and N. Murray. 1983. Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* **170**:827-842.
- Fung, Y. K., W. G. Lewis, L. B. Crittenden, and H.-J. Kung. 1983. Activation of the cellular oncogenic *c-erbB* by LTR insertion: molecular basis for induction of erythroblastosis by avian leukosis virus. *Cell* **33**:357-368.
- Goldfarb, M. P., and R. A. Weinberg. 1981. Generation of novel, biologically active Harvey sarcoma viruses via apparent illegitimate recombination. *J. Virol.* **38**:136-150.
- Goodwin, R. G., F. M. Rottman, T. Callaghan, H.-J. Kung, P. A. Maroney, and T. W. Nilsen. 1986. *c-erbB* activation in avian leukosis virus-induced erythroblastosis: multiple epidermal growth factor receptor mRNAs are generated by alternative RNA processing. *Mol. Cell. Biol.* **6**:3128-3133.
- Henry, C., M. Coquillaud, S. Saule, D. Stehelin, and B. Debuire. 1985. The four c-terminal amino acids of the *v-erbA* polypeptide are encoded by an intronic sequence of the *v-erbB* oncogene. *Virology* **140**:179-182.
- Herman, S. A., and J. M. Coffin. 1986. Differential transcription from the long terminal repeats of integrated avian leukosis virus DNA. *J. Virol.* **60**:497-505.
- Herman, S. A., and J. M. Coffin. 1987. Efficient packaging of read-through RNA in ALV: implications for oncogene transduction. *Science* **236**:845-848.
- Huang, C. C., N. Hay, and J. M. Bishop. 1986. The role of RNA molecules in transduction of the proto-oncogene *c-fps*. *Cell* **44**:935-940.
- Ikawa, S., K. Hagino-Yamagishi, S. Kawai, T. Yamamoto, and K. Toyoshima. 1986. Activation of the cellular *src* gene by transducing retrovirus. *Mol. Cell. Biol.* **6**:2420-2428.
- Joyner, A. L., and A. Bernstein. 1983. Retrovirus transduction: generation of infectious retroviruses expressing dominant and selectable genes is associated with in vivo recombination and deletion events. *Mol. Cell. Biol.* **3**:2180-2190.
- Katz, R. A., R. W. Terry, and A. M. Skalka. 1986. A conserved *cis*-acting sequence in the 5' leader of avian sarcoma virus RNA is required for packaging. *J. Virol.* **59**:163-167.
- Kung, H.-J., and N. J. Maihle. 1986. Molecular basis of oncogenesis by non-acute avian retroviruses, p. 77-99. In G. F. de Boer (ed.), *Avian leukosis*. Martinus Nijhoff Publishing, Boston.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Martin, P., C. Henry, F. Ferre, C. Bechade, A. Begue, C. Calothy, B. Debuire, D. Stehelin, and S. Saule. 1986. Characterization of a *myc*-containing retrovirus generated by propagation of an MH2 viral subgenomic RNA. *J. Virol.* **57**:1191-1194.
- Miles, B. D., and H. L. Robinson. 1985. High-frequency transduction of *c-erbB* in avian leukosis virus-induced erythroblastosis. *J. Virol.* **54**:295-303.
- Nilsen, T. W., P. A. Maroney, R. G. Goodwin, F. Rottman, L. B. Crittenden, M. A. Raines, and H.-J. Kung. 1985. *c-erbB* activation in ALV-induced erythroblastosis: novel RNA processing and promoter-insertion result in the expression of an amino-truncated EGF receptor. *Cell* **41**:719-726.
- Pugatsch, T., and D. W. Stacey. 1983. Identification of a sequence likely to be required for avian retroviral packaging. *Virology* **128**:505-511.
- Purchase, H. G. 1986. The pathogenesis and pathology of neoplasms caused by avian leukosis viruses, p. 171-196. In G. F. de Boer (ed.), *Avian leukosis*. Martinus Nijhoff Publishing, Boston.
- Raines, M. A., W. G. Lewis, L. B. Crittenden, and H.-J. Kung. 1985. *c-erbB* activation in avian leukosis virus induced erythroblastosis: clustered integration sites and the arrangement of ALV provirus in the *c-erbB* alleles. *Proc. Natl. Acad. Sci. USA* **82**:2287-2291.
- Roebroek, A. J. M., J. A. Schalken, C. Onnekink, H. P. J. Bloemers, and W. J. M. Van de Ven. 1987. Structure of the feline *c-fes/fps* proto-oncogene: genesis of a retroviral oncogene. *J. Virol.* **61**:2009-2016.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
- Sealy, L., M. L. Privalsky, G. Moscovici, C. Moscovici, and J. M. Bishop. 1983. Site-specific mutagenesis of avian erythroblastosis virus: *erbB* is required for oncogenicity. *Virology* **130**:155-178.
- Shank, P. R., and M. Linial. 1980. Avian oncovirus mutant (SE21Q1b) deficient in genomic RNA: characterization of a deletion in the provirus. *J. Virol.* **36**:450-456.
- Shimotohno, K., and H. M. Temin. 1981. Formation of infectious progeny virus after insertion of herpes simplex thymidine kinase into DNA of an avian retrovirus. *Cell* **26**:67-78.
- Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**:503-517.
- Swanstrom, R., R. C. Parker, H. E. Varmus, and J. M. Bishop. 1983. Transduction of a cellular oncogene: the genesis of Rous sarcoma virus. *Proc. Natl. Acad. Sci. USA* **80**:2519-2523.
- Sylla, B. S., and H. M. Temin. 1986. Activation of oncogenicity of the *c-rel* proto-oncogene. *Mol. Cell. Biol.* **6**:4709-4716.
- Teich, N., J. Wyke, T. Mak, A. Bernstein, and W. Hardy. 1984. Pathogenesis of retrovirus-induced diseases, p. 785-998. In R. Weiss, N. Teich, H. E. Varmus, and J. Coffin (ed.), *RNA tumor viruses*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Varmus, H. E. 1984. The molecular genetics of cellular oncogenes. *Annu. Rev. Genet.* **18**:553-612.
- Yamamoto, T., T. Nishida, N. Miyajima, S. Kawai, T. Ooi, and K. Toyoshima. 1983. The *erbB* gene of avian erythroblastosis virus is a member of the *src* gene family. *Cell* **35**:71-78.