JOURNAL OF VIROLOGY, Sept. 1986, p. 556–563
0022-538X/86/090556-08$02.00/0

Vol. 59, No. 3

Transduction of c-*src* Coding and Intron Sequences by a Transformation-Defective Deletion Mutant of Rous Sarcoma Virus

MEEI-MEEI SOONG,[1]† SHINJI IIJIMA,‡ AND LU-HAI WANG*
The Rockefeller University, New York, New York 10021-6399

Received 6 March 1986/Accepted 13 May 1986

The mechanism of cellular *src* (c-*src*) transduction by a transformation-defective deletion mutant, *td*109, of Rous sarcoma virus was studied by sequence analysis of the recombinational junctions in three *td*109-derived recovered sarcoma viruses (rASVs). Our results show that two rASVs have been generated by recombination between *td*109 and c-*src* at the region between exons 1 and 2 defined previously. Significant homology between *td*109 and c-*src* sequences was present at the sites of recombination. The viral and c-*src* sequence junction of the third rASV was formed by splicing a cryptic donor site at the 5′ region of *env* of *td*109 to exon 1 of c-*src*. Various lengths of c-*src* internal intron 1 sequences were incorporated into all three rASV genomes, which resulted from activation of potential splice donor and acceptor sites. The incorporated intron 1 sequences were absent in the c-*src* mRNA, excluding its being the precursor for recombination with *td*109 and implying that initial recombinations most likely took place at the DNA level. A potential splice acceptor site within the incorporated intron 1 sequences in two rASVs was activated and was used for the *src* mRNA synthesis in infected cells. The normal *env* mRNA splice acceptor site was used for *src* mRNA synthesis for the third rASV.

A unique feature of acute oncogenic retroviruses is the presence of transforming genes in their genomes. These genes are obtained from the host by retroviruses via rare recombinational events (1). So far, the vast majority of the acute oncogenic retroviruses have been isolated from tumors developed in the field or in laboratory animals infected spontaneously or experimentally with nonacute retroviruses (42). However, even in the experimentally infected animals, the frequency for the conversion of a nonacute to an acute oncogenic virus is so low that the event cannot necessarily be reproduced under any given experimental conditions. This is not so with a series of transformation-defective (*td*) deletion mutants of Rous sarcoma virus (RSV) (14). These mutants, which all retain a portion of the RSV transforming gene *src*, are capable of recombining with cellular *src* (c-*src*) to regenerate transforming viruses at predictable frequencies (10, 13, 14, 46, 52, 53). Analyses of the genomic sequences of mutants and of recovered sarcoma viruses (rASVs) derived from them enabled us to conclude that retention of a portion of the 3′ *src* sequence in the mutant virus is both necessary and sufficient for the generation of rASVs (29, 48). Although retention of the 5′ *src* sequence is not essential for generation of rASVs, it appears to be important for producing nondefective rASVs (29, 48). All the rASVs derived from *td* viruses retaining both 5′ and 3′ *src* sequences are nondefective (50, 52–54), whereas all the rASVs derived from a *td* mutant, *td*109, which retains the 3′ *src* but lacks the 5′ *src*, and its upstream c-*src*-derived sequences, are replication defective due to deletions of various lengths of replicative genes (48). A simple explanation for these observations is that the *td* viruses containing both 5′ and 3′ *src* sequences are able to undergo homologous recombination with c-*src*, resulting in the insertion of the deleted sequences into the original position. By contrast, *td*109 would have to undergo nonhomologous recombination at the 5′ end with c-*src*, resulting in the loss of certain replicative sequences (48). This process is similar to the transduction of most c-*onc* sequences into retroviruses (1).

To understand better the mechanism of recombination between *td*109 and c-*src* and to explain the origin of deletions in *td*109-derived rASVs, we molecularly cloned and analyzed the nucleotide sequences of *td*109 and three rASVs derived from it. Our data provide evidence that initial recombination between *td*109 and c-*src* most likely occurred at the DNA level and was mediated by certain partially homologous sequences in the *td*109 genome and in c-*src*. Our results also show that some recombination junctions were formed by splicings involving cryptic donor and acceptor sites in the *td*109 genome and c-*src* sequences.

MATERIALS AND METHODS

Molecular cloning. Circular viral DNAs were isolated from infected cells as described previously (27, 50). Briefly, DNAs from 20 10-cm dishes of virus-infected cells were extracted by the procedure of Hirt (15); supernatnat DNAs were enriched for the circular forms by acidic phenol extraction (57) followed by passage through a Bio-Gel column (36). The circular DNA-enriched materials were checked for the presence and purity of viral DNAs by Southern blots (38). After confirmation, 2 to 3 μg of the DNA was digested with an appropriate restriction enzyme and was cloned into pBR322 DNA according to the standard procedure (22). *td*109 DNA was digested with *Eco*RI, and the resulting 2.5-kilobase (kb) *gag* and 3.6-kb *pol-env* subgenomic DNA fragments were cloned separately into the *Eco*RI site of pBR322. rASV3812 and rASV382 circular DNAs were digested with *Hin*dIII, and the linearized full-length genomes were cloned into the *Hin*dIII site of pBR322. Because of deletions in these rASV genomes, *Hin*dIII became a single-cut enzyme for their DNAs. rASV374 DNA was digested with *Eco*RI, and the linearized full-length genome was cloned into the *Eco*RI site of pBR322. Similarly, only the *Eco*RI sites in the U3 regions of the rASV374 DNA remain, because of its internal deletion. The potential DNA fragment

* Corresponding author.
† Present address: Department of Veterinary Pathology, University of Illinois, Urbana, IL 61801.
‡ Present address: Nagoya University, Nagoya, Japan.

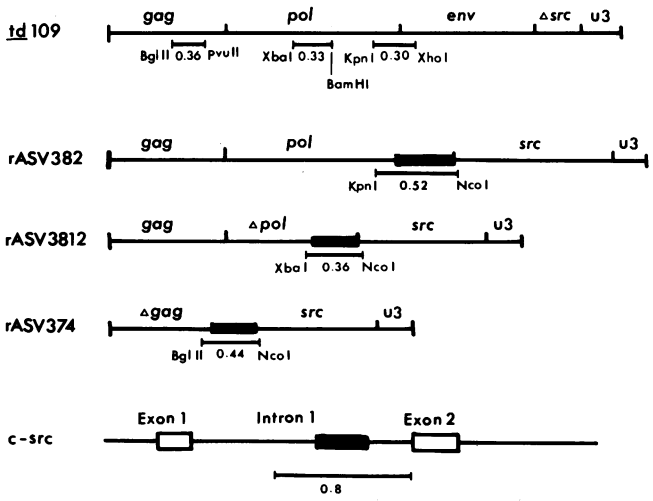

FIG. 1. Regions of viral genome and c-*src* DNAs sequenced. Viral genomes are shown as RNA molecules without poly(A) tails. The *src* in *td*109 represents the 3′ 296 nucleotides of *src* (29). *pol* and *gag* represent the remaining 5′ *pol* and 5′ *gag* sequences in rASV3812 and rASV374 genomes, respectively, as described in the text. Symbols: ▬, part of c-*src* intron 1 sequences; ——, regions and sizes (in kilobases) of viral genome and c-*src* DNA sequences, shown under viral genomes or c-*src* DNA. Enzymes used for preparing DNA fragments are shown. Corresponding positions of those restriction enzyme cleavage sites in PR-C RSV (33) as measured by number of nucleotides from the 5′ end of the viral RNA genome are as follows: *Bgl*II (no. 1630 in p27), *Pvu*II (no. 1987 in p12), *Xba*I (no. 3378 in *pol*), *Bam*HI (no. 3707 in *pol*), *Kpn*I (no. 4999 in *pol*), and *Xho*I (no. 5258 in *env*). *Nco*I cleaves at the initiation codon of *src*. The structure of the c-*src* gene is as defined previously (41).

containing the recombination junction in each rASV (Fig. 1) as suggested by previous study (48) was prepared from the respective molecularly cloned viral DNAs. The gaps in the the termini of the DNA fragments were filled, and then the fragments were subcloned into the *Sma*I site in the polylinker region of M13mp8 replicative-form DNA (24). After initial sequencing and identification of the recombination sites, the corresponding DNA fragments from parental *td*109 (Fig. 1) were isolated and were subcloned into M13 by a similar method. The c-*src* intron 1 DNA fragments were prepared from the plasmid clones pFCI and pFCII, which have been described by Takeya and Hanafusa (41). These fragments were also subcloned into M13 DNA as above. For each DNA fragment subcloned into M13, clones with opposite orientations of the insert were isolated.

Nucleotide sequencing. For viral DNA subcloned in M13, the dideoxy method (32) was used to sequence pairs of clones with opposite polarity. For c-*src* DNA, sequencing methods of Maxam and Gilbert (23) and Sanger et al. (32) were used.

RNA blotting and hybridization. Poly(A)$^+$ RNAs from rASV3812-infected cells or normal chicken embryo fibroblasts (CEF) were prepared as described before (49, 51). Poly(A)$^+$ RNAs from 12-day-old chicken embryonic brain were isolated by sodium thiocyanate extraction and CsCl-cushioned centrifugation (12, 44). An appropriate amount of RNA from each source was analyzed by agarose gel electrophoresis and RNA blotting (43), and subsequent hybridization and washing were performed as described previously (51). The ^{32}P-labeled DNA probes used for hybridization included the 3′ *Pvu*II *src* fragment (40) and the 0.36-kb *Xba*I-to-*Nco*I fragment of rASV3812 DNA (Fig. 1). The RNA blots were hybridized sequentially with the intron 1 and the *src* probes.

S1 nuclease analysis. Poly(A)$^+$ RNAs from individual rASV-infected cells were isolated as described above. The recombination junction fragment of each rASV (Fig. 1) was prepared from the viral DNA in plasmid clone and was labeled with ^{32}P at its 5′ ends, using bacteriophage T4 polynucleotide kinase (22). The plus-strand DNA was removed partially by hybridizing the denatured [^{32}P]DNA with the single-stranded M13 DNA containing the homologous insert in opposite polarity from the plus-strand DNA, followed by gel separation of the hybrid complex from the unhybridized minus-strand DNA. The minus-strand DNA recovered from the gel was used as an S1 nuclease probe. Conditions for DNA-RNA hybridization and for S1 nuclease digestion were similar to the published methods (4, 11). A 10- to 20-μg amount of poly(A)$^+$ RNAs from rASV-infected cells was mixed with the S1 nuclease probe, heat denatured, and then precipitated in ethanol. The nucleic acids were redissolved in 10 to 15 μl of hybridization buffer containing 14 mM PIPES [piperazine-*N*-*N*′-bis(2-ethanesulfonic acid)], pH 6.5, 80% formamide, 0.4 N NaCl, and 1 mM EDTA; the mixture was sealed in a capillary. After heating at 85°C for 5 min, the mixture was incubated at 53°C for 15 h. The hybridization mixture was then ejected into 0.2 ml of S1 nuclease buffer containing 50 mM sodium acetate, pH 4.5, 0.25 N NaCl, 1 mM $ZnSO_4$, 50% glycerol, and 3,000 to 5,000 U of S1 nuclease per ml (Bethesda Research Laboratories, Inc., Gaithersburg, Md.); the mixture was incubated at 37°C for 30 min. A 10-μl volume of 0.2 M EDTA and 20 μl of 1 M Tris hydrochloride, pH 8, were added after digestion, and the mixture was extracted with buffered phenol and chloroform. The nucleic acids were concentrated by ethanol precipitation and then were analyzed in sequencing gels by using appropriate sequencing leaders as size markers.

RESULTS AND DISCUSSION

v-*src* and c-*src* sequences at the sites of recombination. Previous sequence analysis of the *td*109 genome, using defined DNA primers and purified viral RNA, enabled us to determine the precise *src* deletion in the mutant (29). *td*109 retains 296 nucleotides of the 3′ *src* sequence but lacks all of the 5′ *src* and 316 nucleotides of its upstream region, including the *src* mRNA splice acceptor site. In two other studies, the approximate extent of deletions and junctions of recombination in a series of *td*109-derived rASVs, including the three studied here, were mapped (47, 48). To further pursue the mechanisms for recombination between *td*109 and c-*src*, we set out to determine the *td*109-derived rASV and c-*src* sequences at the recombination junctions. We cloned a portion of the *td*109 genome and the complete genomes of rASV3812, rASV374, and rASV382. The genomic structure of the three rASVs (48) is shown in Fig. 1. The subgenomic DNA fragment presumed to contain the junction of recombination in each rASV genome was prepared. In each case, the fragment was flanked by an *Nco*I site at its 3′ end (*Nco*I cuts at the initiation codon of *src* [40, 41]) and by a 6-base restriction enzyme site closest to the 5′ deletion boundary at its 5′ end (Fig. 1). In addition, *td*109 DNA fragments corresponding to the recombination sites for the three rASVs, and the intron 1 region of c-*src* DNA, were also isolated (Fig. 1). The intron 1 region of c-*src* was chosen for analysis because we suspected that it was involved in the recombination with *td*109. The three rASVs contained all of

the normal 5′ *src*-specific oligonucleotides (48) and the N-terminal tryptic peptides (17), which suggested that the viruses had the normal *src* initiation codon located 10 nucleotides downstream from the beginning of exon 2 (41). Therefore, the c-*src* recombination sites must be located upstream from exon 2. All of the v-*src* and c-*src* DNA fragments of interest were subcloned into M13mp8 and sequenced.

Figure 2 shows the nucleotide sequence of the 0.36-kb *Xba*I-to-*Nco*I fragment (Fig. 1) of rASV3812. Comparison of this sequence with those of the 0.33-kb *Xba*I-to-*Bam*HI fragment in *td*109 *pol* gene (Fig. 1) and with the intron 1 sequence of c-*src* revealed the junction of recombination (Fig. 2). The rASV3812 and *td*109 sequences match up to the position corresponding to nucleotide 3557 in the *pol* gene of PR-C RSV (33). Subsequently, the rASV sequence matched instead with a c-*src* intron 1 sequence beginning 662 nucleotides upstream from exon 2 and continuing for 174 nucleotides before skipping the remaining 488 nucleotides of the 3′ intron 1 sequence and joining to the c-*src* exon 2 sequence. At the site of recombination, a significant sequence homology was found between *td*109 and c-*src* (Fig. 2). The crossover point is apparently located within the CTCCAC sequence, which is identical between *td*109 and c-*src*. This sequence and its upstream homology probably mediated the recombination. The 3′ boundary of the 174-base-pair internal segment of c-*src* intron 1 was marked by a potential splice donor signal sequence GGAGGTAT on c-*src* DNA (Fig. 3). The joining of this intron 1 sequence to exon 2 apparently

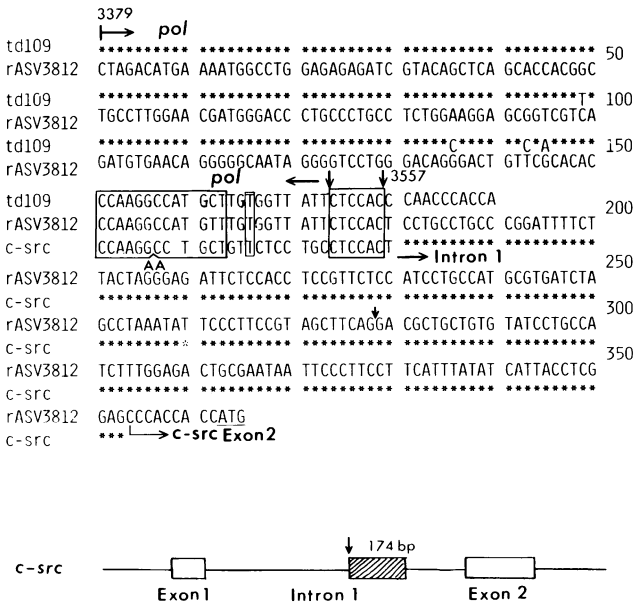

FIG. 2. Recombination junction in the rASV3812 genome. The nucleotide sequence of the 0.36-kb DNA fragment of rASV3812 is compared with that of the 0.33-kb *td*109 DNA fragment and with the 3′ region of c-*src* intron 1. Numbers at beginning of sequence and at recombination site indicate corresponding positions in PR-C RSV. The sequences of *td*109 and c-*src* with significant homology are boxed. The AA dinucleotide shown below the box is part of the GGAACC sequence of the c-*src* intron 1. Symbols: *, *td*109 and c-*src* sequences identical to that of rASV3812; ↓ ↓, putative recombination site; ↓ within intron 1 sequence, potential splice acceptor site; ▨, region of c-*src* intron 1 incorporated into rASV3812 genome (size is in base pairs). A portion of the c-*src* gene structure is shown below the intron 1 sequence.

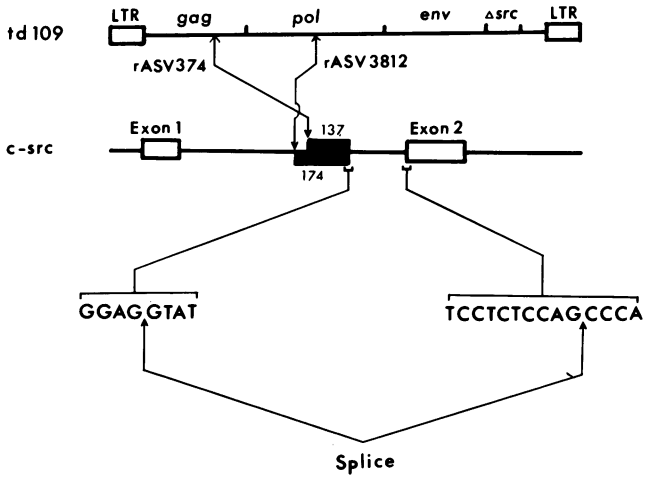

FIG. 3. Mechanism for the formation of recombination junctions in rASV3812 and rASV374 genomes. Initial recombination between *td*109 and c-*src* are presumed to occur at the DNA level. The 3′ *gag* and middle *pol* regions of *td*109 recombined with c-*src* at the indicated sites to form the 5′ recombination junctions for rASV374 and rASV3812, respectively. A cryptic splice donor site in intron 1 downstream from the recombination site was activated and spliced to exon 2, resulting in the incorporation of intron 1 sequences (▬) into the rASV genomes. Sizes of the incorporated intron regions are shown in number of nucleotides. The sequences of the cryptic splice donor site and the normal exon 2 splice acceptor site are shown.

occurred by splicing the potential donor site to the normal exon 2 acceptor site (Fig. 3).

The sequence of the 0.44-kb rASV374 DNA fragment is compared with that of the 0.36-kb 3′ *gag* DNA fragment of *td*109 and with the c-*src* intron 1 sequence (Fig. 4). The recombination in rASV374 appears to have occurred in the middle of p12 of *td*109, corresponding to nucleotide 1934 in PR-C RSV (33). Downstream from this region, the rASV374 sequence matches with c-*src* intron 1 sequence beginning 625 nucleotides upstream from exon 2 and continuing for 137 nucleotides before skipping the remaining intron 1 sequence and joining to exon 2. As in the case of rASV3812, similar sequence homology (Fig. 4) between *td*109 and c-*src* was observed at the site of recombination and probably promoted the recombination in the generation of rASV374. The sequence data suggest that the recombination junction is located within the TC dinucleotide (Fig. 4). Interestingly, the same cryptic splice donor site of c-*src* intron 1 mentioned above was used to join the 3′ boundary of the intron 1 sequence in rASV374 to exon 2 of c-*src*.

The mechanism for the generation of rASV3812 and rASV374 is shown in Fig. 3. The recombinations between *td*109 and c-*src* occurred at multiple sites and apparently involved short stretches of homologous sequences. A 5/5- and a 10/11-nucleotide homology between FBJ murine leukemia virus genome and c-*fos* was seen at the 5′ and 3′ recombination sites, respectively (45). However, it cannot be generalized that such sequence homology is always needed for recombination between a viral genome and c-*onc* sequences. For example, no similar homology is present at the recombination junctions of *src* in SR-A or PR-C RSV (33, 41), although some homology is present between the *pol-env* junction of avian retroviruses and the region of c-*src* DNA corresponding to the 5′ recombination junction of v-*src* in BH RSV and in RSV29 (8, 21). In addition, there are only patchy sequence similarities between c-*fps* and the viral

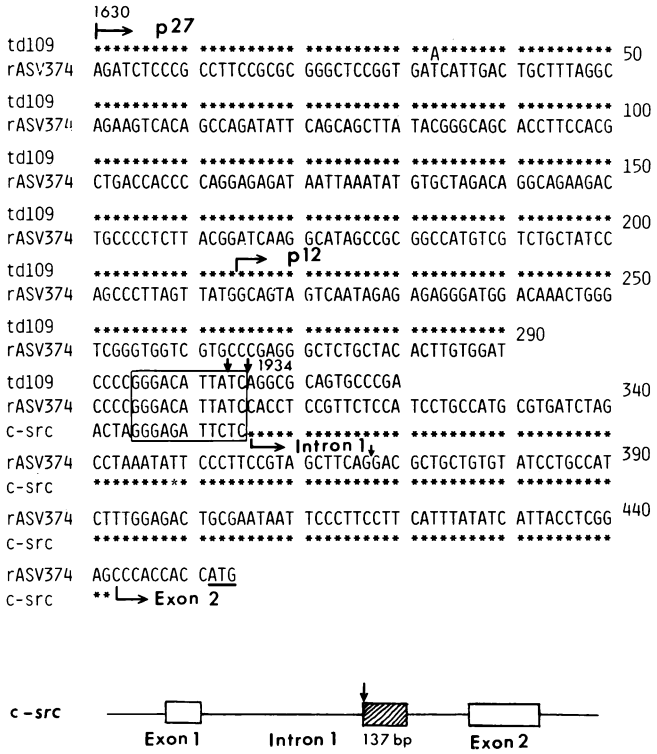

FIG. 4. Recombination junction in the rASV374 genome. The nucleotide sequence of the rASV374 0.44-kb DNA fragment is compared with that of the *td*109 0.36-kb fragment and with the 3′ region of c-*src* intron 1. All symbols are the same as in Fig. 2.

genome near the 5′ recombination sites of *fps* in Fujinami sarcoma virus and in PRCII (16, 35). Nevertheless, the lack of homology could be due to sequence divergence or deletion of the original recombination junction, particularly if the virus has been passaged extensively following the initial transduction and if there is no selective pressure for the conservation of those junction sequences.

The cryptic splice donor site in c-*src* intron 1 described above was activated probably by joining the 5′ *td*109 sequence to c-*src*. This potential donor site was not detected in a previous analysis of the c-*src* mRNA (39). However, use of this potential splice donor site may be tissue specific. Splicing of this donor site to c-*src* exon 2 removed 488 nucleotides of the 3′ intron 1 sequence and resulted in incorporation of 174 and 134 nucleotides of internal intron 1 sequences, in addition to the $p60^{c\text{-}src}$ coding region, into rASV3812 and rASV374, respectively. It is not clear whether removal of those 3′ intron 1 sequences of c-*src* occurred at the initial stages of recombination or subsequent to the formation of rASVs.

Comparison of the sequence of the 0.52-kb rASV382 DNA fragment with those of the relevant regions in *td*109 (the 0.3-kb *pol-env* DNA fragment) and c-*src* (exon 1 and intron 1) revealed an even more complicated sequence organization on the rASV382 genome (Fig. 5 and 6). The 5′ *env* sequence of *td*109 corresponding to nucleotide 5236 (46 nucleotides downstream from the termination codon of *pol*) in PR-C RSV (33) was fused to exon 1 of c-*src*, which was in turn joined to intron 1 sequence 687 nucleotides upstream from exon 2. The intron 1 sequence continues for 199 nucleotides and then joins to exon 2 by skipping 488 nucleotides of the 3′ intron 1 sequences, and by using the same potential splice donor site observed for the generation of the two rASVs described above. The *env*-exon 1 junction apparently was formed by splicing between *td*109 and c-*src*, using a cryptic splice donor signal, GGGGGTAA, located 43 nucleotides 3′ to the *pol* termination codon of *td*109 and the normal c-*src* exon 1 acceptor site. This potential viral splice donor site apparently is not used in the processing of viral RNAs, since no viral RNAs corresponding to this splice site have been observed (4). It is not clear how this site was activated in the presumed *td*109-c-*src* hybrid molecule. Interestingly, Nilson et al. (28) observed that the same potential donor site of an avian leukosis virus was used for splicing to c-*erb*B, resulting in the joining of the viral sequence to c-*erb*B at the same point. Activation of the same cryptic viral splice donor site in these totally independent transductional events implies that although this site appears to be suppressed in the regular processing of viral RNA transcripts, it can be activated if the viral sequence is truncated or joined to a foreign sequence. The secondary structure of the mRNA precursor can affect the pattern of splicing (37).

The joining of exon 1 to the 5′ end of the internal intron 1 sequence transduced into the rASV382 genome appears to be formed by splicing between exon 1 donor site and a cryptic acceptor sequence, TTCTTCTCCCAAGGAAC, located 700 nucleotides upstream from exon 2 (Fig. 6). Similar to the cryptic intron 1 splice donor site described above, this

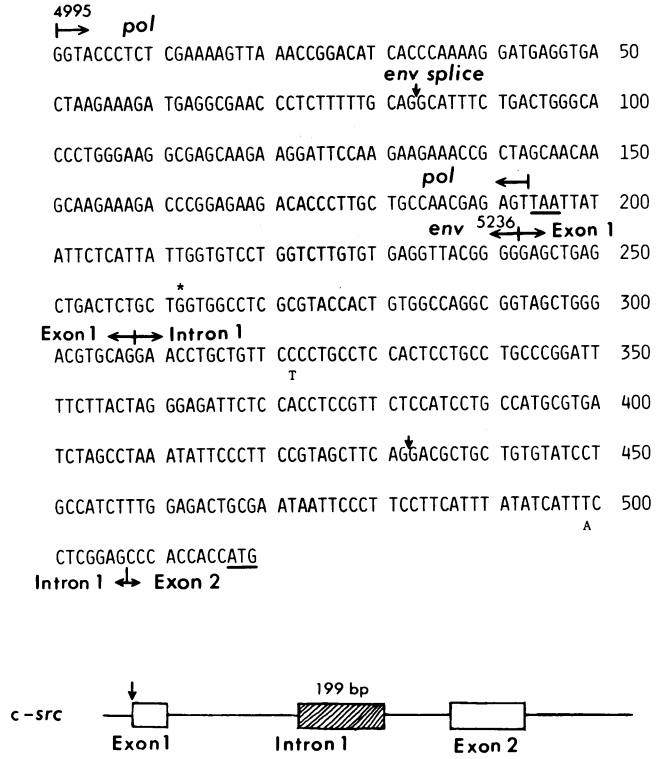

FIG. 5. Recombination junction in the rASV382 genome. The nucleotide sequence of rASV382 0.52-kb DNA fragment was compared with those of the *td*109 0.30-kb fragment and c-*src* exon 1 and intron 1 regions. The rASV382 sequence is shown with the differences from either *td*109 or c-*src* indicated below it. The nucleotide G* is missing in SR-A, but is present in PR-C RSV (33, 41). By comparing the rASV382 sequence with those of its parents, the domains of sequences can be defined as shown. The *td*109-c-*src* recombination junction is between *env* and exon 1. Symbols are the same as in Fig. 2.

potential acceptor site was not observed previously in the analysis of c-*src* mRNA (39). The intron 1-exon 2 joining in rASV382 involved the same cryptic splice donor site in intron 1 as for the other two rASVs. A stretch of 199 nucleotides of the c-*src* intron 1 sequence was incorporated into the rASV382 genome.

We could not discern whether the *env*-exon 1 junction resulted from a primary or secondary event of recombination between *td*109 and c-*src*. For example, the initial recombination could have occurred between a region downstream from the 5′ *env* site in *td*109 and a region upstream from exon 1 in c-*src*, and the *env*-exon 1 junction was subsequently formed by splicing of the *td*109-c-*src* primary RNA transcript. Additionally, an intact or 3′-truncated *td*109 provirus integrated upstream from the c-*src* gene could also promote the synthesis of a *td*109-c-*src* readthrough transcript, which upon splicing gave rise to the recombination junctions in the rASV382 genome. In a separate study, we showed that in avian sarcoma virus UR2, the *gag* and *ros* junction had most likely been formed by a splicing event (26). Similarly, it was proposed that the 5′ recombination junction between the transforming gene, *rel*, and the viral replicative sequence in avian reticuloendotheliosis virus strain T, was a result of splicing (55).

A few nucleotide differences within the homologous regions were found between the rASVs and their parental *td*109 and c-*src* sequences (Fig. 2, 3 and 5). The differences may have resulted from divergence after the recombination, particularly since there is no selection pressure for the conservation of protein functions encoded by the remaining *gag* and *pol* sequences in rASV374 and rASV3812. The incorporated intron 1 sequences of c-*src* have no obvious function either. No sequence variations were found between

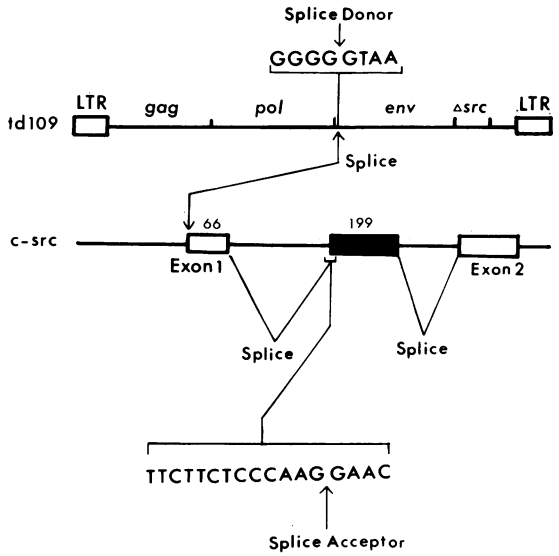

FIG. 6. Mechanism for formation of recombination junctions in the rASV382 genome. The *td*109-c-*src* junction was formed by splicing between a cryptic splice donor site in the 5′ *env* of *td*109 genome and the splice acceptor site of c-*src* exon 1. A cryptic splice acceptor and a donor site in c-*src* intron 1 were activated for splicing, resulting in the incorporation of a segment of intron 1 (▬) into the rASV382 genome. Sizes of exon 1 and the incorporated intron 1 region are shown in base pairs. Nucleotide sequences of the cryptic donor site in the *td*109 genome and the cryptic acceptor site in c-*src* are shown.

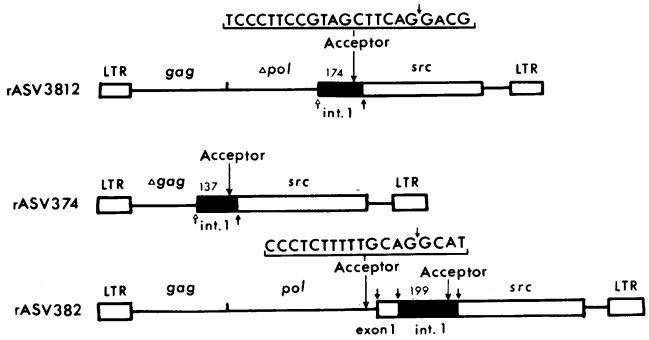

FIG. 7. Structure of three rASV genomes. Intron (int)-derived sequences are shown (▬) and their sizes are shown as number of nucleotides. The *td*109 and c-*src* sequence junctions formed presumably by recombination (△) and those by splicing (▲) are indicated. The sequences flanking a potential splice acceptor site within the transduced intron 1 sequence and the *env* mRNA splice acceptor site are shown. Potential acceptor site is located 84 nucleotides upstream from initiation of *src* in the three rASV genomes. The *env* mRNA splice acceptor site is located 159 nucleotides upstream from the exon 1 of *src* in rASV382.

the *pol* region of *td*109 and that of rASV382, which was shown to encode a functional reverse transcriptase (48).

***src* mRNA splicing acceptor sites in rASV genomes.** The *src* mRNA splice acceptor site of SR-A RSV was deleted in *td*109 and was not regained from c-*src* in the three rASV genomes. Previous analyses of the viral RNAs in *td*109-derived rASV-infected cells showed that the spliced *src* mRNAs of rASV3812 and rASV374 had a size indistinguishable from that of SR-A and that the *src* mRNA of rASV382 was larger than that of SR-A (47, 48). It was suggested previously that the *env* gene splice acceptor site could be used for the formation of *src* message in rASV382-infected cells (47, 48). However, new splice acceptor sites must be created in rASV3812 and rASV374 genomes for the v-*src* mRNA formation. Examination of the sequences upstream from the *src* coding region in the three rASV genomes revealed that there is a potential acceptor site within the intron 1-derived sequence located 84 nucleotides upstream from the AUG codon of *src* (Fig. 7). To see if this potential acceptor site is used for splicing of *src* mRNA from these rASVs, we prepared ^{32}P-labeled recombination junction DNA fragments (Fig. 1) from individual rASVs and performed S1 nuclease experiments. The results (Fig. 8) demonstrate that a 434-nucleotide DNA fragment was generated in rASV382, and that a series of fragments with the top band at the position of 85 nucleotides were generated in rASV374. Results similar to those of the rASV374 experiment were obtained with rASV3812 (data not shown). The spliced 5′ leader sequence is expected to protect two nucleotides 5′ to both of the acceptor sites. The series of DNA fragments separated from one another by one nucleotide in the experiments with rASV3812 and rASV374 were probably generated due to melting of the DNA-RNA hybrid at the ends, followed by digestion of the upaired DNA terminus by S1 nuclease. The minor bands at position of 93 to 95 nucleotides were also seen in rASV374 and rASV3812 (Fig. 7 and data not shown). The origin of these bands is not clear.

The S1 nuclease experiments confirmed our prediction that the *env* acceptor site would be used for rASV382 and that the new potential acceptor site in the incorporated intron 1 sequence would be used for rASV374 and rASV3812 in the *src* mRNA splicing. Since no detectable amount of the

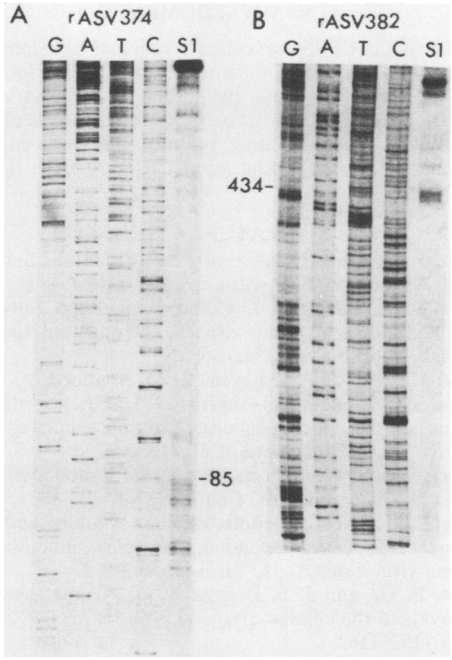

FIG. 8. S1 nuclease analysis of the rASV374 and rASV382 *src* mRNAs. The 5'-labeled minus-strand DNAs of the 0.44-kb and the 0.52-kb recombination junction-containing fragments (Fig. 1) of rASV374 and rASV382, respectively, were used as S1 probes. Conditions are as described in the text. The sequence leaders of the rASV382 0.52-kb DNA and *td*109 0.30-kb DNA were used as size markers for the analysis of rASV382 and rASV374 RNAs, respectively. The sizes of the relevant band positions are indicated.

85-nucleotide fragment was seen in rASV382 RNA, it appeared that the *env* acceptor site was strongly preferred over the new acceptor site in this virus. This new acceptor site was not functioning in the formation of the 4-kb c-*src* mRNA as shown previously by S1 nuclease analysis (39) and by our RNA blotting and hybridization experiments (Fig. 9).

Initial step of recombination between *td*109 and c-*src*. The data show that various lengths of c-*src* sequence defined previously as the intron 1 region (41) were incorporated into the rASV genomes. The intron 1 region was defined by comparing the c-*src* DNA sequence with that of v-*src* (41), and this definition is supported by S1 nuclease analysis of the c-*src* mRNA (39). Our results imply that initial recombinations between *td*109 and c-*src* have taken place at the DNA level. To assess this hypothesis, we prepared a probe containing the transduced c-*src* intron 1 sequences from rASV3812 and examined it to see whether the c-*src* mRNA contained these sequences. The most commonly observed c-*src* 4-kb mRNA from chicken embryonic brain (Fig. 9) or from CEF (not shown) hybridized strongly with a v-*src* probe, but not with the intron 1 probe, which in the positive control detected both the rASV3812 genomic and spliced *src* mRNAs. The result with brain c-*src* RNA is presented here since the c-*src* RNA level is higher in the embryonic brain than in any other tissue (26; L.-H. Wang, unpublished data); therefore, its use should increase the chance of detecting even a small amount of the intron 1 probe sequence present in the mRNA. These results provide evidence that initial recombinations between *td*109 and c-*src* took place at the DNA level. Our data agree with previous evidence which suggested that transduction of c-*src* into avian leukosis virus in the generation of RSV took place by recombination at the DNA level, since the 5' recombination site was located 16 nucleotides upstream from the acceptor site of c-*src* exon 1 (39, 41). The 5' recombination between avian myeloblastosis virus-associated helper virus and c-*myb* also occurred in the intron region of c-*myb* (19). Precursors or certain rare forms of c-*src* mRNA containing those intron sequences could have participated in the recombinations. The level of those RNAs may be so low that it was not detected by our analysis. Nevertheless, the precursor mRNA could exist only in the nucleus and has a very short half-life. Therefore, it is spatially and kinetically unfavorable for it to form a complex with the viral genome to be packaged for subsequent recombination during reverse transcription.

Overall, our data suggest strongly that transduction of c-*src* can occur through recombination at the DNA level, by involving short stretches of homologous sequences or by splicing between the viral genome and c-*src* sequences. It is very difficult to decide experimentally the stage of recombination between the viral genome and c-*src* during the viral replication cycle. But it could occur during or after the integration of the viral DNA genome. The recombination junctions observed in these rASVs represent the final products of what may be a multiple-stepped process of recombination between the *td* virus and c-*src*. However, it is likely that the 5' v-*src*–c-*src* sequence junctions observed in rASV3812 and rASV374 represent the primary event of recombination. By contrast, the corresponding junction in rASV382 is more likely formed after the initial recombination event. In the latter case, the initial event could be either merely the integration of the *td*109 provirus upstream from c-*src*, or a true recombination between *td*109 and c-*src*. Our

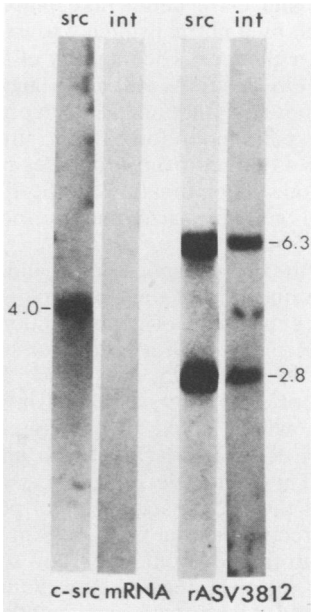

FIG. 9. Hybridization of c-*src* and rASV3812 RNAs with v-*src* and c-*src* intron 1 probes. A five-μg amount of poly(A)$^+$ RNA from an rASV3812-infected nonproducer CEF clone and 10 μg of RNA from 12-day-old chicken embryonic brain were analyzed. The *src* and intron probes indicate the v-*src* and c-*src* intron 1-derived probes described in the text. Exposure time with the rASV3812 RNAs was fourfold shorter than that for c-*src* RNA. Sizes of RNAs are shown in kilobases.

current data only concern the 5′ recombination. Obviously, a second 3′ recombination is needed to obtain the essential viral replicative sequences. The only major differenc between *td*109 and other avian leukosis viruses such as Rous-associated viruses 1 and 2 (ATCC VR-334 and ATCC VR-657) is the presence of a certain 3′ *src* sequence in the *td*109 genome. The chance of a 5′ recombination between a viral genome and c-*src* should be even among these viruses. The 3′ *src* sequence present in *td*109 apparently is responsible for the great difference in the frequency of c-*src* transduction into its genome. The 3′ *src* presumably promoted the 3′ recombination between the initial *td*109-c-*src* recombinant intermediate and the parental *td*109 to obtain the 3′ viral sequences essential for viral replication. The 3′ *src* could also play a role in conferring the transforming potential to the otherwise nontransforming c-*src* gene newly obtained.

The genomic structure of the three rASVs is summarized in Fig. 7. All of them have incorporated a segment of c-*src* internal intron 1 sequence into their genomes resulting from activation of potential splice donor and acceptor sites. Although no other *td*109-derived rASVs have been sequenced, previous analysis by RNase T_1 oligonucleotide fingerprinting did not reveal unexpected oligonucleotides in their genomes, in contrast to the three rASVs studied here, which were shown to contain one to three new oligonucleotides (48). The significance of the intron 1 sequence in rASV382 is not clear, since it is not needed for the *src* mRNA formation (as in the cases of rASV3812 and rASV374), or for viral replication, because apparently it is missing in the genomes of several other *td*109-derived rASVs. Recombinations for the generation of those *td*109-derived rASVs might have occurred in such a way that only a small amount of sequences other than the $p60^{c\text{-}src}$ coding region were incorporated. For example, recombinations between *td*109 and c-*src* could take place upstream from exon 1, followed by a normal splicing to exon 2, eliminating all the intron 1 sequences. The majority of the *td*109-derived rASVs contain *env* deletions (48). This suggests that recombination at the *pol-env* junction has been preferred. Whether there is a corresponding preferred c-*src* site is not clear. As expected, rASVs derived from *td* viruses capable of undergoing homologous recombination at the 5′ region of c-*src* have a normal genomic size and do not contain extra sequences (50, 52–54).

Detection of the cryptic splice donor and acceptor sites in c-*src* raises the question of their physiological significance. Tissue-specific expression of gene functions by alternative splicing of mRNAs has been reported for several eucaryotic genes (2, 6, 7, 9, 18, 20, 25, 30, 31, 34, 56). We detected multiple forms of smaller c-*src*-related mRNAs in chicken muscle tissues which could be generated by alternative splicing of c-*src* RNA transcript (Iijima and Wang, unpublished data). These muscle-specific c-*src*-related RNAs, lacking most of the coding sequences of $p60^{c\text{-}src}$, cannot be precursors for recombination with *td*109 in the generation of rASVs, although they do contain certain sequences derived from the c-*src* intron 1 region. Embryonic neural tissues have been found to express significantly high levels of $p60^{c\text{-}src}$ kinase activity (5) and of c-*src* mRNA (26). The N-terminal polypeptide fragments of $p60^{c\text{-}src}$ from neural tissues had a slower electrophoretic mobility than those from CEF or other tissues (J. Brugge, personal communication). It would be interesting to see whether those potential splice donor and acceptor sites in the c-*src* intron 1 region are activated in a tissue-specific manner.

ACKNOWLEDGMENTS

We thank Boris Lin for excellent technical assistance, and H. Hanafusa and R. Jove for comments on the manuscript.

This work was supported by Public Health Service grants CA29339 and CA18213 from the National Institutes of Health to L.H.W. M.M.S. was a visiting scientist and was supported by Public Health Service Training grant CA24930 from the National Cancer Institute.

LITERATURE CITED

1. **Bishop, J. M., and H. Varmus.** 1982. Functions and origins of retroviral transforming genes. p. 999–1108. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Molecular biology of tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
2. **Breitbart, R. E., H. T. Nguyen, R. M. Medford, A. T. Destree, V. Mandavi, and B. Nadal-Ginard.** 1985. Intricate combinational patterns of exon splicing generate multiple regulated troponin T isoforms from a single gene. Cell **41**:67–82.
3. **Broome, S., and W. Gilbert.** 1985. Rous sarcoma virus encodes a transcriptional activator. Cell **40**:537–546.
4. **Chang, L.-J., and C. M. Stoltzfus.** 1985. Cloning and nucleotide sequences of cDNAs spanning the splice junctions of Rous sarcoma virus mRNAs. J. Virol. **53**:969–972.
5. **Cotton, P. C., and J. S. Brugge.** 1983. Neural tissues express high levels of the cellular *src* gene product $pp60^{c\text{-}src}$. Mol. Cell. Biol. **3**:1157–1162.
6. **Crabtree, G. R., and J. Kant.** 1982. Organization of the rat γ-fibrinogen gene: alternative mRNA splice patterns produce the γA and γB (γ′) chains of fibrinogen. Cell **31**:159–166.
7. **de Ferra, F., H. Engh, L. Hudson, J. Kamholz, C. Puckett, S. Molineaux, and R. A. Lazzarini.** 1986. Alternative splicing accounts for four forms of myelin basic protein. Cell **43**:721–727.
8. **Dutta, A., L.-H. Wang, T. Hanafusa, and H. Hanafusa.** 1985. Partial nucleotide sequence of Rous sarcoma virus-29 provides evidence that the original Rous sarcoma virus was replication defective. J. Virol. **55**:728–735.
9. **Early, P., J. Rogers, M. Davis, K. Calame, M. Bond, R. Wall, and L. Hood.** 1980. Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. Cell **20**:313–319.
10. **Enrietto, P. J., L. N. Payne, and J. A. Wyke.** 1983. Analysis of the pathogenicity of transformation-defective partial deletion mutants of avian sarcoma virus: characterization of recovered viruses which encode novel *src*-specific proteins. Virology **127**:397–411.
11. **Fradin, A., R. Jove, C. Hemenway, H. D. Keiser, J. Manley, and C. Prives.** 1984. Splicing pathway of SV40 mRNAs in *X. laevis* oocytes differ in their requirement for snRNPs. Cell **37**:927–936.
12. **Glisin, V., R. Crkvenjakov, and C. Byus.** 1974. Ribonucleic acid isolated by cesium chloride centrifugation. Biochemistry **13**:2633–2637.
13. **Halpern, C. C., W. S. Hayward, and H. Hanafusa.** 1979. Characterization of some isolates of newly recovered avian sarcoma virus. J. Virol. **29**:91–101.
14. **Hanafusa, H., C. C. Halpern, D. L. Buchhagen, and S. Kawai.** 1977. Recovery of avian sarcoma virus from tumors induced by transformation-defective mutants. J. Exp. Med. **146**:1735–1747.
15. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–369.
16. **Huang, C.-C., C. Hammond, and J. M. Bishop.** 1985. Nucleotide sequence and topography of chicken c-*fps* genesis of a retroviral oncogene encoding a tyrosine-specific protein kinase. J. Mol. Biol. **181**:175–186.
17. **Karess, R. E., and H. Hanafusa.** 1981. Viral and cellular *src* genes contribute to the structure of recovered avian sarcoma virus transforming protein. Cell **24**:155–164.
18. **King, G. R., and J. J. Piatigorsky.** 1983. Alternative RNA splicing of the murine αA-crystalline gene: protein-coding information within an intron. Cell **32**:707–712.
19. **Klempnauer, K.-H., and J. M. Bishop.** 1983. Transduction of c-*myb* into myeloblastosis virus: locating points of recombination within the cellular gene. J. Virol. **48**:565–572.

20. **Kress, M., D. Glaros, G. Khoury, and G. Jay.** 1983. Alternative RNA splicing expression of the H-2K gene. Nature (London) **306**:602–604.
21. **Lerner, T. L., and H. Hanafusa.** 1984. DNA sequence of the Bryan high-titer strain of Rous sarcoma virus: extent of *env* deletion and possible genealogical relationship with other viral strains. J. Virol. **49**:549–556.
22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
24. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digested restriction fragments. Gene **19**:269–276.
25. **Nabashima, Y., Y. Kurijama-Fuji, M. Muramatsu, and K. Ogata.** 1984. Alternative transcription and two modes of splicing result in two myosin light chains from one gene. Nature (London) **308**:333–338.
26. **Neckameyer, W. S., M. Shibuya, M.-T. Hsu, and L.-H. Wang.** 1986. Proto-oncogene c-*ros* codes for a molecule with structural features common to those of growth factor receptors and displays tissue-specific and developmentally regulated expression. Mol. Cell. Biol. **6**:1478–1486.
27. **Neckameyer, W. S., and L.-H. Wang.** 1984. Molecular cloning and characterization of avian sarcoma virus UR2 and comparison of its transforming sequence with those of other avian sarcoma viruses. J. Virol. **50**:914–921.
28. **Nilson, T. W., P. A. Nuroney, R. G. Goodwin, F. M. Rottman, L. B. Crittenden, M. A. Raines, and H.-J. Kung.** 1985. c-*erb*B activation in ALV-induced erythroblastosis: novel RNA processing and promoter insertion result in expression of an amino-terminal EGF receptor. Cell **41**:719–726.
29. **Parvin, J. D., and L.-H. Wang.** 1984. Mechanisms for the generation of *src*-deletion mutants and recovered sarcoma viruses: identification of viral sequences involved in *src* deletions and in recombination with c-*src* sequences. Virology **138**:236–245.
30. **Rosenfeld, M. G., S. G. Amara, and R. M. Evans.** 1984. Alternative RNA processing: determining neuronal phenotype. Science **225**:1315–1320.
31. **Rozek, C. E., and N. Davidson.** 1983. *Drosophila* has one myosine heavy chain gene with three developmentally regulated transcripts. Cell **32**:23–34.
32. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
33. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
34. **Schwartzbauer, J. E., J. W. Tamkun, I. R. Lemischka, and R. O. Hynes.** 1983. Three different fibronectin mRNAs arise by alternative splicing within the coding region. Cell **35**:421–431.
35. **Shibuya, M., and H. Hanafusa.** 1982. Nucleotide sequence of Fujinami sarcoma virus: evolutionary relationship of its transforming gene with transforming genes of other sarcoma viruses. Cell **32**:787–795.
36. **Shibuya, M., L.-H. Wang, and H. Hanafusa.** 1982. Molecular cloning of the Fujinami sarcoma virus genome and its comparison with sequences of other related transforming viruses. J. Virol. **42**:1007–1016.
37. **Solnick, D.** 1986. Alternative splicing caused by RNA secondary structure. Cell **43**:667–676.
38. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
39. **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.
40. **Takeya, T., R. A. Feldman, and H. Hanafusa.** 1982. DNA sequence of the viral and cellular *src* gene of chickens. I. Complete nucleotide sequence of an *Eco*RI fragment of recovered avian sarcoma virus which codes for gp37 and pp60^{src}. J. Virol. **44**:1–11.
41. **Takeya, T., and H. Hanafusa.** 1983. Structure and sequence of the cellular gene homologous to the RSV *src* gene and the mechanism for generating the transforming virus. Cell **32**:881–890.
42. **Teich, N.** 1982. Taxonomy of retroviruses, p. 25–207. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Molecular biology of tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
43. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **70**:5201–5205.
44. **Ullrich, A., J. Shine, J. Chirgwin, R. Pietet, E. Tischer, W. J. Rutter, and H. M. Goodman.** 1977. Rat insulin genes: construction of plasmids containing the coding sequences. Science **196**:1313–1319.
45. **Van Beveren, C., F. van Straaten, T. Curran, R. Moeller, and I. M. Verma.** 1983. Analysis of FBJ-MuSV provirus and c-*fos* (mouse) gene reveals that viral and cellular *fos* gene products have different carboxy termini. Cell **32**:1241–1255.
46. **Vigne, R., M. Breitman, C. Moscovici, and P. K. Vogt.** 1979. Reinstitution of fibroblast-transforming ability in *src* deletion mutants of avian sarcoma virus during animal passage. Virology **93**:413–426.
47. **Wang, L.-H.** 1985. Deletion in the 3′ *pol* sequence correlates with aberration of RNA expression in certain replication-defective avian sarcoma viruses. J. Virol. **54**:446–459.
48. **Wang, L.-H., M. Beckson, S. M. Anderson, and H. Hanafusa.** 1984. Identification of the viral sequence required for the generation of recovered avian sarcoma viruses and characterization of a series of replication-defective recovered avian sarcoma viruses. J. Virol. **49**:881–891.
49. **Wang, L.-H., and P. Duesberg.** 1974. Properties and location of Poly(A) in Rous sarcoma virus RNA. J. Virol. **14**:1515–1529.
50. **Wang, L.-H., B. Edelstein, and B. J. Mayer.** 1984. Induction of tumors and generation of recovered sarcoma viruses by, and mapping of deletions in, two moleculary cloned *src* deletion mutants. J. Virol. **50**:904–913.
51. **Wang, L.-H., R. Feldman, M. Shibuya, H. Hanafusa, M. F. D. Notter, and P. C. Balduzzi.** 1981. Genetic structure, transforming sequence, and gene product of avian sarcoma virus UR1. J. Virol. **40**:258–267.
52. **Wang, L.-H., C. C. Halpern, M. Nadel, and H. Hanafusa.** 1978. Recombination between viral and cellular sequences generated transforming sarcoma virus. Proc. Natl. Acad. Sci. USA **75**: 5812–5816.
53. **Wang, L.-H., C. Moscovici, R. E. Karess, and H. Hanafusa.** 1979. Analysis of the *src* gene of sarcoma viruses generated by recombination between transformation-defective mutants and quail cellular sequences. J. Virol. **32**:546–556.
54. **Wang, L.-H., P. Snyder, T. Hanafusa, and H. Hanafusa.** 1980. Evidence for the common origin of viral and cellular sequences involved in sarcomagenic transformation. J. Virol. **35**:52–64.
55. **Wilhelmsen, K. C., K. Eggleton, and H. M. Temin.** 1984. Nucleic acid sequence of the oncogene v-*rel* in reticuloendotheliosis virus strain T and its cellular homolog, the proto-onocogene c-*rel*. J. Virol. **52**:172–182.
56. **Young, R. A., O. Hagenbuchle, and U. Schibler.** 1981. A single mouse α-amylase gene specifies two different tissue-specific mRNA. Cell **23**:451–458.
57. **Zasloff, M., G. D. Ginder, and G. Felsenfeld.** 1978. A new method for the purification of covalently closed circular DNA molecules. Nucleic Acids Res. **5**:1139–1152.