MOLECULAR AND CELLULAR BIOLOGY, Aug. 1991, p. 4165–4176
0270-7306/91/084165-12$02.00/0

Vol. 11, No. 8

Analysis of cDNAs of the Proto-Oncogene c-*src*: Heterogeneity in 5′ Exons and Possible Mechanism for the Genesis of the 3′ End Of v-*src*

THAMBI DORAI,[1] JOAN B. LEVY,[2]† LILY KANG,[1] JOAN S. BRUGGE,[2] AND LU-HAI WANG[1]*

Department of Microbiology, Mount Sinai School of Medicine, New York, New York 10029-6574,[1] and Department of Microbiology, Howard Hughes Medical Institute, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6076[2]

Received 17 December 1990/Accepted 20 May 1991

To further characterize the gene structure of the proto-oncogene c-*src* and the mechanism for the genesis of the v-*src* sequence in Rous sarcoma virus, we have analyzed genomic and cDNA copies of the chicken c-*src* gene. From a cDNA library of chicken embryo fibroblasts, we isolated and sequenced several overlapping cDNA clones covering the full length of the 4-kb c-*src* mRNA. The cDNA sequence contains a 1.84-kb sequence downstream from the 1.6-kb pp60$^{c\text{-}src}$ coding region. An open reading frame of 217 amino acids, called *sdr* (*src* downstream region), was found 105 nucleotides from the termination codon for pp60$^{c\text{-}src}$. Within the 3′ noncoding region, a 39-bp sequence corresponding to the 3′ end of the RSV v-*src* was detected 660 bases downstream of the pp60$^{c\text{-}src}$ termination codon. The presence of this sequence in the c-*src* mRNA exon supports a model involving an RNA intermediate during transduction of the c-*src* sequence. The 5′ region of the c-*src* cDNA was determined by analyzing several cDNA clones generated by conventional cloning methods and by polymerase chain reaction. Sequences of these chicken embryo fibroblast clones plus two c-*src* cDNA clones isolated from a brain cDNA library show that there is considerable heterogeneity in sequences upstream from the c-*src* coding sequence. Within this region, which contains at least 300 nucleotides upstream of the translational initiation site in exon 2, there exist at least two exons in each cDNA which fall into five cDNA classes. Four unique 5′ exon sequences, designated exons UE1, UE2, UEX, and UEY, were observed. All of them are spliced to the previously characterized c-*src* exons 1 and 2 with the exception of type 2 cDNA. In type 2, the exon 1 is spliced to a novel downstream exon, designated exon 1a, which maps in the region of the c-*src* DNA defined previously as intron 1. Exon UE1 is rich in G+C content and is mapped at 7.8 kb upstream from exon 1. This exon is also present in the two cDNA clones from the brain cDNA library. Exon UE2 is located at 8.5 kb upstream from exon 1. The precise locations of exons UEX and UEY have not been determined, but both are more than 12 kb upstream from exon 1. The existence and exon arrangements of these 5′ cDNAs were further confirmed by RNase protection assays and polymerase chain reactions using specific primers. Our findings indicate that the heterogeneity in the 5′ sequences of the c-*src* mRNAs results from differential splicing and perhaps use of distinct initiation sites. All of these RNAs have the potential of coding for pp60$^{c\text{-}src}$, since their 5′ exons are all eventually joined to exon 2.

c-*src* is the cellular counterpart of the oncogene v-*src*, encoded by Rous sarcoma virus (RSV) (6, 33, 71). The c-*src* gene is one of the most extensively characterized proto-oncogenes among a large family of 50 or more proto-oncogenes known to date (26, 34, 38). Highly conserved through evolution, c-*src* is widely distributed in all metazoa and encodes a 60-kDa membrane-associated phosphoprotein exhibiting tyrosine-specific kinase activity (2, 13, 15, 39, 64).

The ubiquitous c-*src* mRNA is a 4-kb RNA species (24, 29, 52, 70, 73, 80, 81). Low levels of c-*src* mRNA are present in most chicken tissues and chicken embryo fibroblasts (CEF) (29, 70, 80, 81). Slightly elevated levels of c-*src* mRNA were observed in macrophages, monocytes, and spleen, thymus, and chromaffin cells (1, 23, 24, 29, 30, 54, 61). An exceptionally high level of the c-*src* protein (0.2 to 0.4% of the total protein) is present in platelets (27, 28). Studies of the expression of c-*src* in neural tissues has strongly implicated a role for c-*src* in the development and maturation of neurons. Eight- to tenfold higher levels of pp60$^{c\text{-}src}$ expression (compared with levels in CEF) were seen in the developing vertebrate nervous system and in the *Drosophila* nervous system (9–12, 14, 21, 48, 67, 69, 72). A unique role of pp60$^{c\text{-}src}$ in neuronal cells is highlighted by the finding that neurons and neuroblastoma cells contain an altered form of c-*src* having an increased kinase activity (7, 9–11). This neuronal form of c-*src* contains six additional amino acids in its regulatory domain, generated by alternative splicing to include an 18-bp miniexon (NI) located in intron 3 (45, 49). While both forms of pp60$^{c\text{-}src}$ are expressed in the central nervous system, only the unmodified pp60$^{c\text{-}src}$ is primarily detected in the peripheral nervous system (10, 11, 43, 47). Recently, another novel neuronal c-*src* exon (NII) has been found to be expressed along with NI, between c-*src* exons 3 and 4 in human brain (57).

Instead of the commonly observed 4-kb RNA, chicken skeletal muscle expresses a class of smaller c-*src* mRNA species of about 3 kb, which are generated by alternative splicing (80). Expression of these smaller forms of c-*src* mRNA in muscle commences at prehatching and persists thereafter. Expression of the ubiquitous 4-kb RNA and pp60$^{c\text{-}src}$ is detectable in embryonic muscle up to the prehatching stage and is permanently turned off at this point (17,

* Corresponding author.
† Present address: Department of Pharmacology, New York University Medical Center, New York, NY 10016.

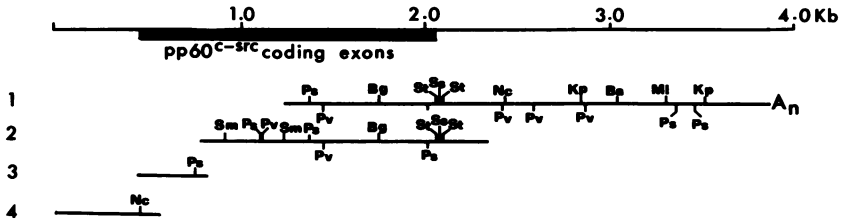

FIG. 1. Organization of the CEF c-*src* cDNA clones. The top line shows the scale for the 4.0-kb c-*src* mRNA. The various overlapping clones are depicted with a brief restriction map. The solid box indicates the coding region of $pp60^{c\text{-}src}$. Clones 1 and 2 were isolated from the oligo(dT)-primed cDNA library prepared from CEF. Clone 3 was isolated by amplifying the reverse transcripts from the CEF poly(A)$^+$ RNAs by using specific primers (see Materials and Methods). Clone 4 represents one of the cDNA clones (type 2b; Fig. 4) isolated by the RACE method. Restriction sites: Nc, *Nco*I; Ps, *Pst*I; Pv, *Pvu*II; Sm, *Sma*I; Bg, *Bgl*I; St, *Stu*I; Ss, *Sst*I; Kp, *Kpn*I; Ba, *Bam*HI; Ml, *Mlu*I. The total complexity of the four cDNA clones is 3.74 kb. Considering the poly(A) sequence of about 150 to 200 bases, the overlapping cDNAs can account for the observed full-length 4-kb c-*src* mRNA.

80). The 3-kb c-*src* RNAs are generated apparently by using different sites of initiation as well as an alternative scheme of splicing and polyadenylation (17). We have shown that the 3-kb c-*src* mRNAs lack the tyrosine kinase domain and instead code for a 24-kDa non-tyrosine kinase protein whose function is yet to be determined.

Human, chicken, and *Drosophila* c-*src* genomic DNAs covering the $pp60^{c\text{-}src}$ coding region have been isolated and characterized (25, 36, 53, 65, 74, 76). The $pp60^{c\text{-}src}$ coding sequences are distributed in 11 exons (75) and, in the case of neuronal c-*src*, in 12 to 13 exons (45, 49, 57). Comparison of the coding sequence of chicken c-*src* with that of RSV v-*src* revealed multiple internal point mutations in v-*src* (32). In addition, the carboxyl 19 amino acids of c-*src* were replaced by 12 unique amino acids in v-*src* (75). Interestingly, this new 3′ v-*src* sequence was found to be present at about 0.9 kb downstream of the last coding exon of c-*src* DNA (75).

The 4-kb c-*src* mRNA is about 2.4 kb larger than the 1.6-kb coding sequence for $pp60^{c\text{-}src}$. Studies of the c-*src* cDNA have so far been limited to the coding region (45, 49, 67). Little information is available about 5′ and 3′ noncoding sequences in the c-*src* mRNA. Previous definition of the 12 c-*src* exons was based on comparisons of the c-*src* DNA sequence with that of the v-*src* gene of RSV (75). Our previous study using specific c-*src* DNA probes for hybridization in Northern (RNA) analysis of the 4-kb mRNA indicated that most of the noncoding sequences are located in the 3′ end of the RNA molecule (80). We have embarked on an effort to characterize the full-length c-*src* cDNA in order to further define the c-*src* gene structure, to characterize the nature of these noncoding sequences, and to further understand the genesis of the RSV v-*src*. In this report, we describe the primary structure and exon organization of the 4-kb c-*src* mRNA from CEF. We also report evidence for diversity in the 5′ exons of CEF and brain c-*src* mRNAs. Finally, we describe the 3′ region downstream from the c-*src* coding sequences and the detection within that region a 217-amino-acid reading frame and the precursor for the 3′ end of the RSV v-*src*.

MATERIALS AND METHODS

Isolation of genomic c-*src* DNA fragments and Southern hybridization. The various fragments of the c-*src* DNA shown in Fig. 5B were isolated from appropriate restriction enzyme digests of molecular clones as described previously (76, 80). ^{32}P-labeled DNA probes were prepared by random-hexamer-primed synthesis (Prime-a-Gene; Promega Biotec) of gel-purified DNAs. Southern hybridization was performed as described previously (17).

Isolation of cDNA clones of c-*src* RNAs from CEF. Primary and secondary CEF cultures were prepared from 11-day-old chicken embryos and maintained as previously described (31). Total RNAs were extracted from secondary cultures of CEF (three passages) (80). Poly(A)$^+$ RNAs were selected by two passages through an oligo(dT)-cellulose column (78). Enrichment of the c-*src* mRNA in the preparation of the CEF poly(A)$^+$ RNA was done by velocity sucrose gradient sedimentation; c-*src* RNAs in each fraction were detected by using a 1.7-kb *Nco*I-*Nru*I DNA fragment of v-*src* as a probe in Northern analysis (17). A cDNA library was constructed in a λgt10 vector as described previously (17). Initially, cDNA clones (such as clone 1 in Fig. 1) were isolated by screening the library with the *Nco*I-*Nru*I fragment of v-*src* mentioned above. Later, another overlapping clone (clone 2) was obtained by rescreening the library with clone 1. To obtain cDNA clones containing the 5′ region of the c-*src* coding sequence, the polymerase chain reaction (PCR) method was used to amplify a defined region of the c-*src* mRNA after reverse transcription. The sequences of the primers used in PCR were 5′-GTCTGTCCTGTAGTGAG-3′, which was used for reverse transcription and amplification and is complementary to a region of c-*src* at the beginning of exon 4, and 5′-ACCATGGGGAGCAGCAA-3′, which was used for the *Taq* polymerase reaction and is located at the beginning of exon 2. Reverse transcription and PCR were performed in a one-step procedure, using avian myeloblastosis virus reverse transcriptase and Replinase (from *Thermus flavis*; NEN) in a buffer containing 50 mM Tris-HCl (pH 9.0), 20 mM $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, 200 μM deoxynucleoside triphosphates, 400 ng of each primer, and 2 μg of CEF poly(A)$^+$ RNA in a final volume of 50 μl. After annealing of the oligonucleotides to the RNA in the reaction buffer, all of the deoxynucleoside triphosphates, 5 U of reverse transcriptase, and 1.25 U of Replinase were added. cDNA synthesis was performed at 42°C for 40 min. The PCR cycles were set as follows: 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min. After amplification, the reaction mixture was run on a low-melting-point agarose gel to remove the unreacted oligonucleotide primers and to size select the PCR products corresponding to 300 nucleotides or longer. The ends of the PCR product were flushed with T4 DNA polymerase, phosphorylated with T4 polynucleotide kinase, and ligated to either *Eco*RV-cut pBluescript (Stratagene) or *Sma*I-cut M13mp19. Plasmid minipreps were screened for the size of the inserts, and selected clones were sequenced

by using a double-stranded DNA sequencing procedure. Clone 3 shown in Fig. 1 was obtained by this method.

A CEF cDNA library enriched for clones spanning the 5′ region of the c-*src* mRNA was also prepared, using a 194-bp *Hin*fI-*Hin*fI fragment of v-*src* (74, 79) as the primer for cDNA synthesis. This primer spans the 3′ two-thirds of exon 2 and 5′ two-thirds of exon 3. This 5′ CEF cDNA library was screened with a 243-bp *Acc*I-*Nco*I fragment of v-*src*, isolated from pTT107 (74, 80) and covering the entire exon 1 and part of the *env-src* intercistronic region. A 5′ c-*src* cDNA clone represented by type 2a in Fig. 3 and 4 was obtained in this manner.

Amplification and cloning of the 5′ end of the c-*src* mRNA. The original protocol (20) for the rapid amplification of cDNA ends (RACE) was adapted and modified to prepare a cDNA library representing the 5′ ends of the CEF c-*src* mRNAs. The oligonucleotide primer used for the first-strand cDNA synthesis contained 17 nucleotides complementary to the c-*src* exon 2 with the sequence 5′-CTGCGAGGCTGG GAATC-3′, designated 3′-RT. The 3′ amplification primer, a 28-mer designated 3′-AMP, contained 21 nucleotides complementary to the c-*src* exon 2 and was located 18 bases upstream of the sequence of 3′-RT. The sequence of the 3′-AMP primer is 5′-CTGAATTCGGGTGGCTCCAGGCT GCGCC-3′.

The reaction mixture for the first-strand synthesis with reverse transcriptase contained 50 mM Tris-HCl (pH 8.3, 42°C), 10 mM $MgCl_2$, 50 mM KCl, 1 mM dithiothreitol, 40 μg of actinomycin D per ml, 1,000 U of RNAsin per ml, 0.5 μg of 3′-RT (90 pmol), 0.5 mM spermidine, 25 μCi of [α-^{32}P]dCTP, 1.5 mM each dGTP, dATP, dTTP, and dCTP, and 25 U of avian myeloblastosis virus reverse transcriptase (Life Sciences) in a final volume of 40 μl. The poly(A)$^+$ RNA (2 μg), dissolved in water, was heated at 70°C for 5 min and quickly chilled on ice.

After addition of the buffer containing salt components for the reverse transcriptase reaction and the 3′-RT primer to the poly(A)$^+$ RNA, the mixture was heated again at 65°C for 2 min and allowed to cool slowly to room temperature before addition of the rest of the reaction components.

Reverse transcription was performed at 42°C for 2 h, followed by phenol-chloroform extraction and ethanol precipitation. The RNA-cDNA hybrids were dissolved in 20 μl of 10 mM Tris-HCl (pH 8.0)–300 mM NaCl–1 mM EDTA and passed through a column of Sepharose CL-4B (2-ml bed volume) to separate the excessive 3′-RT primers from the reaction product. One-drop fractions were collected, and radioactivity was monitored by Cerenkov counting. The fractions in void volume, which contained the ^{32}P-labeled RNA-cDNA hybrids, were collected and ethanol precipitated.

The cDNA products were dissolved in 15 μl of H_2O and tailed by terminal transferase in a final volume of 40 μl in 1× tailing buffer (Bethesda Research Laboratories) supplemented with 8 mM $MgCl_2$, 0.3 mM $ZnSO_4$, 6 μM dATP, 40 μCi of [α-^{32}P]dATP, and 20 U of terminal deoxynucleotidyltransferase (Bethesda Research Laboratories) at 37°C for 20 min. At the end of the reaction, the mixture was incubated at 65°C for 15 min, extracted with phenol-chloroform, and ethanol precipitated. Incorporation of [^{32}P]dATP was monitored to confirm the efficacy of the tailing reaction.

The oligo(A)-extended products were dissolved in 20 μl of H_2O, denatured at 95°C for 5 min, cooled to 72°C, and subjected to second-strand synthesis in a final volume of 50 μl containing 1× reaction buffer, 200 μM each dGTP, dATP, dTTP, and dCTP, 5 U of Amplitaq polymerase, and 25 pmol of the adapter primer (designated ADPR; 5′-GATCTA GAGTCGACATCGATTTTTTTTTTTTTT-3′) containing the restriction sites for *Xba*I, *Sal*I, and *Cla*I. After incubation of the reaction mixture for 40 min at 72°C, 40 pmol of the adapter oligonucleotide (designated AD; 5′-GATCTA GAGTCGACATCGAT-3′) and 25 pmol of the 3′-AMP oligonucleotide were added to the reaction mixture, which was overlaid with 30 μl of mineral oil and subjected to PCR as instructed by the manufacturer (Gene Amp kit; Perkin Elmer Cetus). The mixture was annealed at 54°C for 2 min and amplified in a Perkin Elmer DNA thermal cycler for 40 cycles, with a denaturation step at 94°C for 30 s, an annealing step at 55°C for 5 min, an extension step at 72°C for 3 min, and a final extension step at 72°C for 15 min.

Ten percent of the amplified DNAs was analyzed by electrophoresis in a 1.5% agarose gel to determine the size of the amplified products. The remaining PCR-amplified products were extracted with phenol-chloroform and ethanol precipitated. An aliquot of the final product was digested sequentially with *Eco*RI and *Sal*I and run in a 1.5% agarose gel, and the DNAs in the gel slice corresponding to 200 to 600 bp were recovered. These size-selected PCR products were cloned into the *Eco*RI and *Sal*I sites of M13mp18 and M13mp19. The clones containing c-*src* exon 1 were selected by Benton-Davis blotting, using an exon 1 probe derived from v-*src* as described above. Purified single-stranded DNAs or double-stranded replicative forms were sequenced by using Sequenase (United States Biochemical). A 5′ overlapping clone (clone 4 in Fig. 1) represents one of the clones isolated from this library.

A 2.6-kb *Eco*RI-*Hin*dIII fragment of the c-*src* DNA (probe 2 region in Fig. 5B) with which the PCR-amplified cDNA inserts hybridized was cloned into M13mp19, and its sequence was determined. Those DNA regions rich in GC residues were sequenced by using the *Taq* DNA polymerase. Nested deletions in c-*src* probe 2 clone were created by using either exonuclease III (35) or T4 DNA polymerase (16).

Isolation of the c-*src* cDNA clones from neural tissues. Construction of a cDNA library from 10-day-old embryonic chicken brains as well as isolation and sequencing of the coding region of pp60$^{c\text{-}src}$ from a 3.9-kb cDNA clone isolated from this library were described previously (45). Another independent cDNA clone was isolated later from the same library.

PCR analysis of the exon arrangement. To confirm the exon structure derived from the analysis of 5′ c-*src* cDNA clones, pairs of specific primers were used for PCR to assess the generation of expected products. Total cytoplasmic RNA (10 μg) was hybridized with the 3′-RT primer (see above) and reverse transcribed as described above for the RACE method. The cDNA products were subjected to second-strand synthesis by using (i) a primer (5′-ACAGAAGG GAAAGCAAC-3′) homologous to a sequence (positions 28 to 44) in exon UE2 and (ii) a primer (5′-CCCGCA GAAGGGGTGAG-3′) homologous to a sequence (positions 46 to 62) in exon UE1 (see Fig. 3).

Negative controls for the PCR assay included all of the components for the first-strand synthesis without reverse transcriptase and were carried out in parallel through the PCR steps. Conditions for the PCR amplification were the same as described for the RACE method except that 10 μCi of [α-^{32}P]dCTP (3,000 Ci/mmol) was added to the PCR cocktail. One-tenth of each PCR product was then separated on a nondenaturing 5% acrylamide gel and autoradiographed.

RNase protection assay. ^{32}P-labeled antisense RNA probes containing upstream exons UE2, UE1, 1a, and UEX were obtained by SP6 transcription of *Hin*dIII-linearized RACE clones (types 1 through 4), subcloned into pGEM4 (Promega Biotec). A total of 10^5 cpm of the riboprobe and 30 μg of total cytoplasmic CEF RNAs were used in the protection assay according to the protocol provided by the Ambion RPA assay kit except that hybridization was done at 55°C. The amounts of RNase needed to obtain the optimum signals were determined empirically. Yeast tRNAs served as the negative control. One-third of each RNase-treated and untreated sample was run on a 5% polyacrylamide gel containing Tris-borate-EDTA buffer and 7 M urea, and the gel was autoradiographed.

S1 nuclease analysis. The radiolabeled probe used in the S1 analysis was synthesized by using the single-stranded M13 template containing UE1, exon 1, and *src* coding sequences of the type 3b cDNA. The uniformly labeled M13 probe was generated by the Klenow reaction (New England BioLabs), using the M13 universal sequencing primer and [α-^{32}P]dCTP (3,000 Ci/mmol; Amersham). The radiolabeled single-stranded probe was isolated from a denaturing gel containing 4% acrylamide and 7 M urea. Probe (10^5 cpm per reaction) was hybridized with 20 μg of calf liver tRNA or 5 μg of poly(A$^+$) RNA from chicken embryonic brain or embryonic limb in a buffer containing 80% formamide, 40 mM sodium piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES; pH 6.4), 1 mM EDTA, and 0.4 M NaCl at 60°C for 12 h. Nine volumes of S1 nuclease buffer (0.03 M sodium acetate [pH 4.6], 0.05 M NaCl, 1 mM $ZnSO_4$, 0.5% glycerol) was added to the hybridization reaction mixtures, and digestions were allowed to proceed with 400 U of S1 nuclease (Bethesda Research Laboratories) for 1 h at 25°C. The products were separated on a 4% acrylamide–7 M urea denaturing gel.

RESULTS

Isolation and characterization of the c-*src* cDNAs. Our previous study by Northern analysis (80) showed that the commonly observed 4.0-kb c-*src* RNA is most likely the mRNA coding for the pp60$^{c\text{-}src}$, since it contains all of the coding exons equivalent to those of pp60$^{v\text{-}src}$. To analyze this 4.0-kb c-*src* mRNA further, particularly the 5′ and 3′ sequences outside the coding region, we constructed several cDNA libraries from CEF mRNAs by using different strategies. The approaches included conventional cDNA cloning using oligo(dT)$_{12-18}$ or sequence-specific primer and PCR amplification of defined regions of the c-*src* mRNA (20; see Materials and Methods). Representative clones isolated from these libraries (Fig. 1) are aligned to construct a full-length c-*src* cDNA. Sequencing revealed a stretch of A's at the 3′ end of the cDNA clone 1 with an appropriate polyadenylation signal. Clone 2 appears to have originated from internal initiation during reverse transcription. By Southern analysis (data not shown), the 3′ noncoding region of the 4-kb c-*src* mRNA was found to originate from a 4-kb region of the genomic c-*src* DNA between *Sac*I and *Bgl*II sites (probes 8 to 11; see Fig. 5B) immediately downstream of the last coding exon.

Sequences of the c-*src* cDNAs. A 3,520-nucleotide sequence was compiled from the overlapping cDNA clones 1 through 4. For simplicity, the previously published exon 1 and the c-*src* coding regions (75) are not depicted. The nucleotide and the derived 533-amino-acid sequences of pp60$^{c\text{-}src}$ are essentially identical to those previously published (75). However, we found a few minor differences, as noted in the legend to Fig. 2. As expected, the neuron-specific miniexon (45) is missing in our CEF-derived cDNAs.

The 1,599-nucleotide coding sequence of pp60$^{c\text{-}src}$ is followed by a 1.84-kb 3′ noncoding sequence starting from the 3′ portion of c-*src* exon 12 and ending with a poly(A) tail (Fig. 2). Surprisingly, there is an extended open reading frame of 217 amino acids in this region. We call this reading frame *sdr*, for *src* downstream region. A computer search for sequence homology with existing data did not find a meaningful homologous protein. A poly(A) addition signal (ATTAAA) is present 22 nucleotides upstream of the poly(A) addition site. The 3′ noncoding sequence also contains two ATTTA motifs at positions 2657 and 3310, which have been reported to confer instability to mRNAs (66). Finally, a 39-bp sequence which forms the 3′ end of pp60$^{v\text{-}src}$ is detected at 660 bases downstream from the termination codon of pp60$^{c\text{-}src}$ in the cDNA (Fig. 2 and 8).

Identification of the 5′ exons of c-*src* mRNAs. The cDNA sequence compiled from clones 1 through 4 accounts for only 3.5 kb of the c-*src* mRNA, while its established size is about 4 kb (70, 80, 81). Taking into account that there are 150 to 200 A residues at the 3′ poly(A) tail, one would predict that there are about 0.3 kb of the 5′ sequence missing in our cDNAs. Our primer extension analysis using two small DNA restriction fragments from c-*src* exon 2 as the primers and CEF poly(A)$^+$ mRNAs as templates showed that there were 270 to 300 nucleotides upstream of the c-*src* exon 1 and that there was some heterogeneity in the length of the extended products (data not shown). From the 194-bp (*Hin*fI) primer-extended cDNA library described above, a clone representing the type 2a cDNA (Fig. 3) was isolated which harbored an upstream exon of 268 bases, designated UE2. However, in this clone, exon 1 is not spliced to exon 2; instead, it is spliced to a novel exon, called 1a. To further investigate the potential heterogeneity of the 5′ c-*src* mRNA sequences, a PCR-amplified 5′ RACE cDNA library was prepared (see Materials and Methods). Screening of this library with the exon 1 probe yielded over 20 clones, 14 of which were sequenced. On the basis of their sequences and splicing patterns, we categorize them into five types (Fig. 3 and 4). Sequences of those cDNAs and two independent cDNA clones isolated from a brain cDNA library are shown in Fig. 3, and their exon organizations are depicted in Fig. 4.

In type 1, an upstream exon UE2 (incomplete in this clone) is spliced to exons 1 and 2. In 2a (Fig. 3), however, exon UE2 is spliced to exon 1, which in turn is spliced to exon 1a (incomplete in this clone), as mentioned above. In type 2b, represented by clone 4 (Fig. 1), exon 1a is spliced to exon 2. It is likely that clones 2a and 2b originate from the same type of c-*src* mRNA. They are so differentiated to account for the fact that they are independent clones isolated from different cDNA libraries. The same upstream exon UE2 is present in types 1, 2a, and 2b. In type 3a, an upstream exon designated UE1 is spliced to c-*src* exons 1 and 2 in the expected manner. UE1 contains only a short stretch of sequence, possibly resulting from premature termination during reverse transcription. The 5′ noncoding regions of the two brain-derived c-*src* cDNA clones designated type 3b were sequenced and compared with 5′ cDNA clones of CEF. The sequences of the c-*src* coding region of these clones have been reported previously (45). As shown in Fig. 3, the 5′ noncoding regions of those brain c-*src* cDNAs (type 3b) contain exon 1 and 296 nucleotides of upstream sequence. The most 3′ 27 nucleotides of this upstream sequence in the brain cDNA clones are identical to those of UE1 in type 3a clones from CEF. Hence, we have

FIG. 2. Nucleotide sequence of the 3′ portion of the c-*src* mRNA from CEF. Only the sequence from the 3′ half of exon 12 to the poly(A) tail is shown. Numbering of the amino acids is shown above numbering of the nucleotides; both are adapted from the previously published sequence (75). Few minor differences in the c-*src* coding region (not shown) between our sequence and that in reference 75 are noted; they are as follows. (i) In exon 1, the sequence CTGCTGTGG in reference 75 reads as CTGCTGGTGG in our sequence. (ii) The codon for threonine at position 301 in reference 75 is written as AAC, which actually codes for asparagine. Our cDNA sequence in that position is ACC, which codes for threonine. (iii) The amino acid in position 501 is lysine (AAG) in our sequence instead of arginine (AGG) as originally described. Of these differences in our cDNA sequence, the first is in accordance with v-*src* sequence published by Schwartz et al. (63) and the cDNA sequence of the alternatively spliced c-*src* mRNA from chicken skeletal muscle (17). The presence of lysine (AAG) at position 501 is typical for c-*src* and is supported by other independent studies on c-*src* and the sequence analysis of recovered avian sarcoma viruses (22, 41, 46, 51). The unique open reading frame called *sdr* (217 amino acids in length) lying downstream from the termination codon of $pp60^{c\text{-}src}$ is shown. Termination codons are depicted as either TER or AM. The poly(A) addition signal for the c-*src* mRNA, located 22 nucleotides upstream from the poly(A) tail, is shown with a broken underline. The 3′-terminal 10 amino acids of $pp60^{v\text{-}src}$, beginning with TGCAGGCC and ending with AGGTCG, located in the 3′ noncoding region is shown by bold letters with an underline. The eight-nucleotide sequence TGCAGGCC (the P box [18]; see also Fig. 8) present at the beginning of this sequence is repeated in exon 12 and is highlighted in a similar manner. The putative Q box (18; see also Fig. 8) with the sequence CTCAGTAG (Q′ in Fig. 8) is underlined in closed circles. The two mRNA instability motifs located in the 3′ sequence are indicated by open triangles.

designated the 296-bp upstream sequence of type 3b as UE1 and the mRNA species giving rise to these cDNA clones as type 3b, again to underscore the different origins of cDNA clones 3a and 3b. Again, it is likely that clones 3a and 3b arise from the same c-*src* mRNA. The unusual feature of the UE1 sequence is that it has a high G+C content of approximately 75%. Two other clones, designated types 4 and 5, contain upstream exons UEX and UEY, respectively, spliced to exons 1 and 2 (Fig. 4).

Mapping of the upstream exons in the c-*src* locus. To locate the origins of the upstream c-*src* exons UE1, UE2, UEX, and UEY, each of these cDNA clones was used as a probe to hybridize with Southern blots containing a panel of c-*src* genomic DNA fragments including probes 1 through 5. Dissection of the c-*src* locus to generate these probes is shown in Fig. 5B. The results of Southern analysis indicated that types 1, 2a, 2b, and 3a hybridized to c-*src* probes 5 (which contains exon 1) and 2 (data not shown). This finding suggests that sequences 5′ to exon 1 in these clones map to a region about 8 to 8.5 kb upstream from exon 1. Type 4 and 5 cDNAs had negligible hybridization to any c-*src* DNA except the exon 1-containing probe 5 (data not shown). Therefore, UEX and UEY must be derived from a region more than 12 kb upstream from the exon 1. Their precise origins await further analysis.

Since most of the cDNAs contain exons UE1 and UE2 and both exons hybridize to c-*src* probe 2, this region of the c-*src* locus, defined by a 2.6-kb *Eco*RI-*Hin*dIII fragment (Fig. 5B), was subcloned and sequenced. The 5′ 1.5-kb sequence of this genomic fragment is shown in Fig. 5A. The entire exon UE2 is located within this genomic sequence at about 400 nucleotides from the 5′ end (bold letters in Fig. 5A). Exon UE1 is mapped in a region further downstream (underlined sequences in Fig. 5A). UE1 is flanked by typical splice donor and acceptor sequences. A comparison of the UE1 sequences in the neuronal c-*src* cDNA (type 3b) with the corresponding genomic sequences shown in Fig. 5A revealed that the first 15 nucleotides of this cDNA clone diverged from the genomic sequences precisely at a splice acceptor site. This result suggests that the first 15 nucleo-

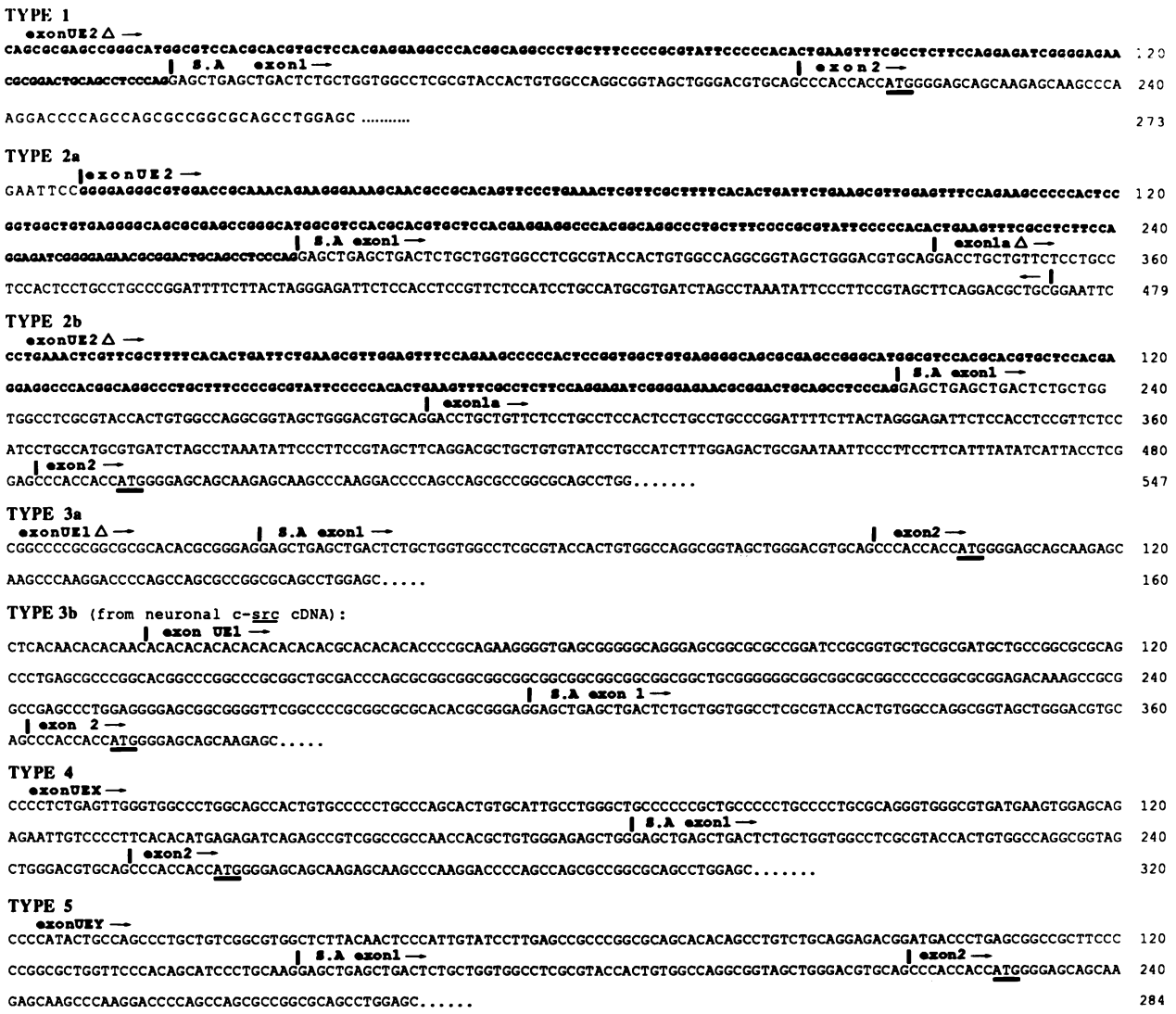

FIG. 3. Sequences of the various types of cDNA clones representing the 5′ ends of the c-*src* mRNAs. Of the sequences shown, types 1, 2b, 3a, 4, and 5, were isolated by the RACE method. The cDNA clone designated type 2a was originally isolated from a 5′-primer-extended CEF cDNA library. The sequence designated type 3b is the 5′ noncoding sequence derived from the brain c-*src* cDNAs isolated by Levy et al. (45). Types 1 and 2 contain a common upstream exon (UE2) spliced to exon 1. These two types are differentiated by the presence of exon 1a in type 2. Clones 2a and 2b could arise from the same c-*src* mRNA and hence could belong to the same type, but they are differentiated since they have been isolated by two independent cloning strategies. Clones 3a and 3b are distinguished for the same reason. The boundary of upstream exon UE1 (for the neuronal c-*src* cDNA [type 3b]) is identified on the basis of a comparison with the corresponding region of the genomic c-*src* DNA (see Fig. 5A). The possible origin of the first 15 nucleotides in the neuronal c-*src* cDNA is discussed in the text. Note that the boundary shown for exon UE2 is the beginning of the corresponding cDNA clone, and it does not necessarily show the 5′ end of that exon. Exons UEX and UEY of types 4 and 5, respectively, have not been mapped on the c-*src* DNA but are at least 12 kb upstream from the exon 1. The initiation codon (ATG) for pp60$^{c\text{-}src}$ in all of these 5′ cDNA clones is highlighted by an underline in the sequence. S.A, splice acceptor.

tides most likely are derived from an exon further upstream. This 15-nucleotide sequence is not present in the probe 2 genomic region that has been sequenced and must therefore have originated from a region further upstream. Exon UE2 lacks a splice acceptor site at its 5′ junction and is not preceded by one in its upstream vicinity of the genomic DNA sequence. This suggests that it could represent the most 5′ exon for this type of the c-*src* mRNA. Its 3′ end does have a splice donor site. The locations of exons UE2 and UE1 in the c-*src* locus are shown in Fig. 5B.

Are these 5′ RACE clones authentic? The results showing the 5′ sequence diversity and the exon structure rely on sequences of the 5′ c-*src* cDNAs. To confirm the existence of heterogeneous c-*src* mRNAs and the predicted 5′ exon arrangement, we performed an analysis of the PCR products by using specific primers and RNase protection experiments. Total CEF RNAs were directly analyzed by PCR for the presence of a contiguous arrangement of exons UE2, UE1, 1, 1a, and 2. The 5′ PCR primers are chosen from the 5′ sequences of UE2 and UE1, respectively, and the 3′ primer (3′-RT) is located in exon 2. The exon arrangements of types 1 and 2, which differ by the insertion of exon 1a, would

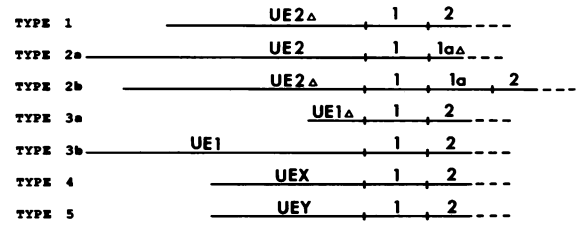

FIG. 4. Schematic illustration of the organization of the 5′ exons of c-*src* mRNAs. Upstream exon UE2 in types 1 and 2b is truncated and is designated UE2Δ. For the same reason, exon 1a in type 2a is designated 1aΔ. Upstream exon UE1 is found in both type 3a from CEF and type 3b from brain c-*src* cDNA. Clone 3a contains only a short stretch of the 3′ sequence of UE1 and is designated UE1Δ.

predict the generation of 620- and 422-bp PCR products. This is in fact the case (Fig. 6A, lane 1). Similarly, the type 3 exon structure was supported by the generation of a 425-bp PCR product (lane 3). Negative controls gave no detectable products due to nonspecific amplification. The RNase protection assays were performed with CEF RNAs and antisense RNA probes synthesized from the clones representing upstream exons UE2 (with or without exon 1a), UE1, and UEX, respectively (Fig. 3). The results are shown in Fig. 6B. The probes were made longer than the c-*src* exon sequences to allow us to distinguish them from the protected fragments. All of the exon probes were protected fully by the CEF RNAs, confirming their representation in the c-*src* mRNAs. The UE2 probe without exon 1a was better protected reproducibly, presumably reflecting the higher abun-

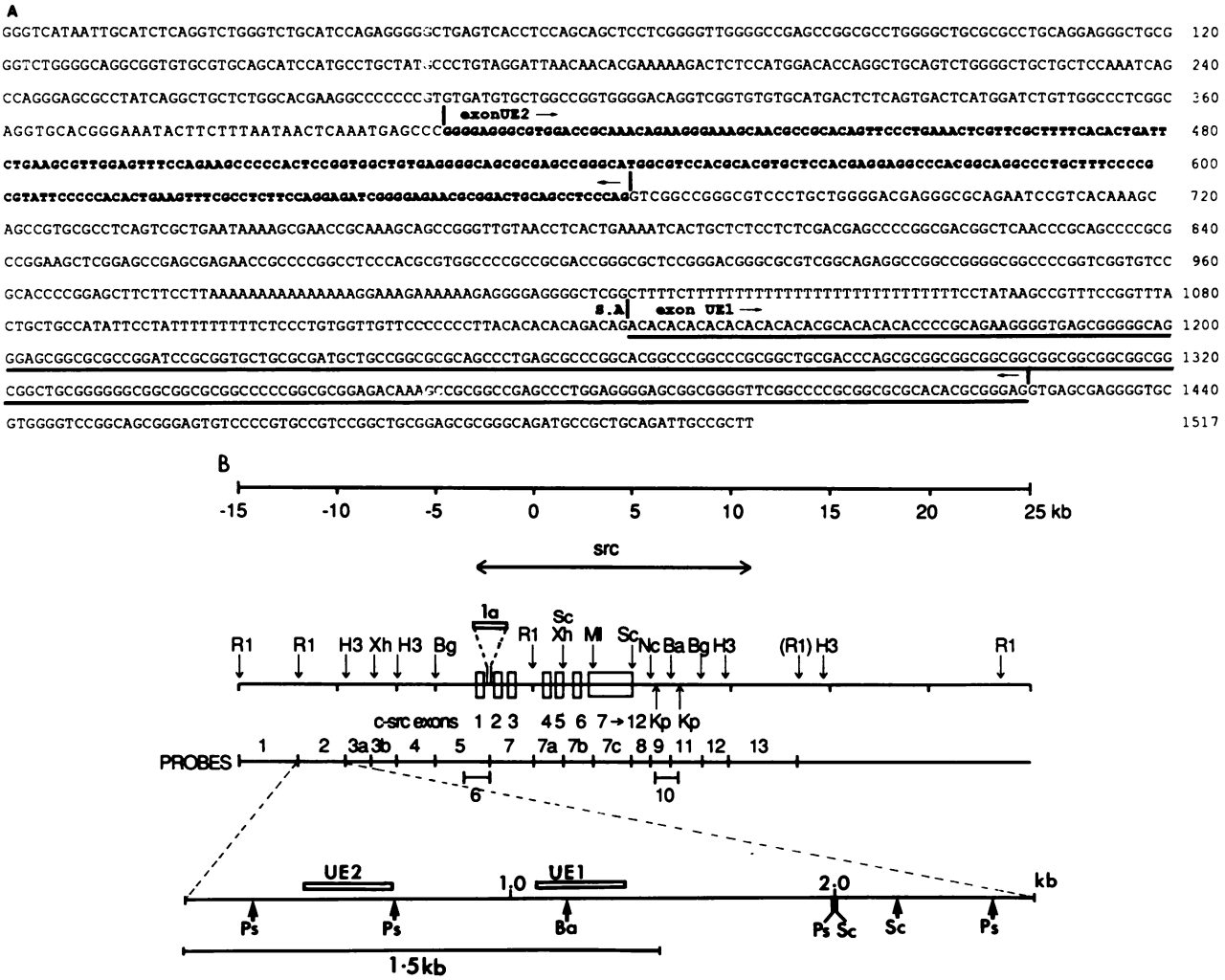

FIG. 5. (A) Sequence of the c-*src* genomic DNA region containing exons UE1 and UE2. The sequence of the 5′ 1.5 kb of the 2.6-kb *Eco*RI-*Hin*dIII probe 2 region is shown. Exon UE2 is shown in bold type, and exon UE1 is underlined. Exon UE1 is flanked by consensus splice donor and splice acceptor (S.A) signal sequences. Exon UE2 is bound at its 3′ junction by a splice donor site, but its 5′ end does not abut a splice acceptor site. Instead, an atypical TATA sequence (TAATAA) is present 19 bases upstream. In addition, a typical AP1 binding sequence is present 330 bases upstream (not shown). (B) Diagrammatic sketch showing the organization of the c-*src* locus and its 5′ exons. The horizontal arrow defines the $pp60^{c\text{-}src}$ coding region. For simplicity, exons 7 to 12 are boxed together without showing the small intron regions. The restriction sites are taken from published studies (76, 80). The *Eco*RI site shown in parentheses is the artificial cloning site. The c-*src* locus is divided into regions of probes 1 through 13 for Southern analysis in exon mapping. The locations of exon UE2 and exon UE1 are shown as open boxes above the expanded c-*src* probe 2 DNA. The location of exon 1a is highlighted in a similar manner above the intron 1 region. However, these exons are not drawn to scale. Restriction sites: Ba, *Bam*HI; Bg, *Bgl*II; RI, *Eco*RI; H3, *Hin*dIII; Ml, *Mlu*I; Nc, *Nco*I; Ps, *Pst*I; Sc, *Sac*I; Xh, *Xho*I.

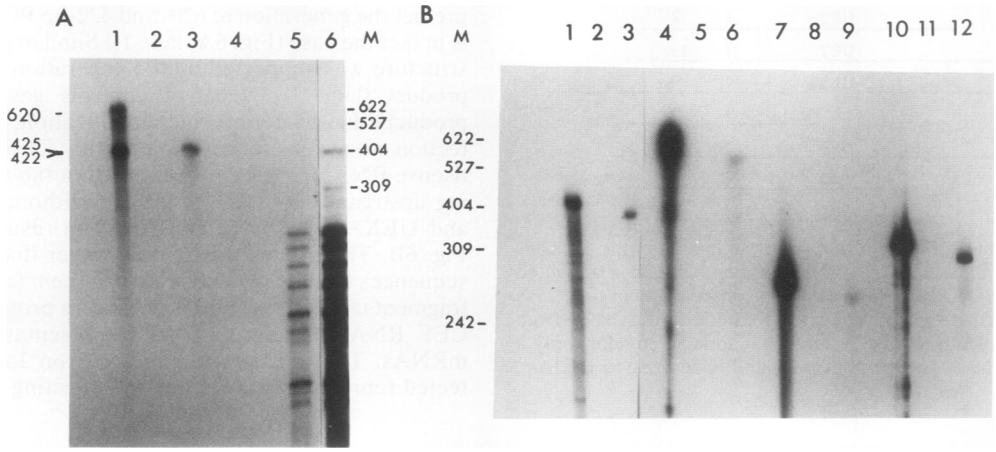

FIG. 6. PCR and RNase protection analysis of c-*src* 5′ exons. (A) PCR analysis of exon arrangement. Samples (10 μg) of total CEF RNAs were used for PCR amplification using specific pairs of 5′ c-*src* exon primers. The PCR products were analyzed by nondenaturing gel electrophoresis. The PCR reaction mixtures contained a small amount of [α-^{32}P]dCTP to allow direct visualization of the amplified products by autoradiography. Lanes: 1, PCR reaction using a 5′ primer from exon UE2 and the 3′-RT primer; 3, PCR reaction using a 5′ primer from exon UE1 and the 3′-RT primer; 2 and 4, negative controls for primers UE2 and UE1, respectively; M, molecular weight marker (pBR322 DNA digested with *Msp*I and end filled with [α-^{32}P]dCTP, using Klenow fragment of DNA polymerase I). Sizes are indicated in nucleotides. (B) RNase protection assay. For each assay, 30 μg of total RNA from CEF was hybridized to [α-^{32}P]UTP-labeled antisense transcripts from the RACE clones. The hybrid molecules were subjected to RNase digestion, and the protected fragments were resolved on a denaturing gel. Lanes: 1, 4, 7, and 10, labeled probes containing intact antisense transcripts of cDNA clones represented by types 1, 2b, 3a, and 4; 3, 6, 9, and 12, protected fragments from the cDNA clones of type 1, 2b, 3a, and 4; 2, 5, 8, and 11, RNase-digested antisense transcripts (in the presence of yeast tRNA) of the aforementioned clones. Lane M is as described for panel A. Each antisense transcript contains the indicated exon sequences in addition to the 27-nucleotide vector sequences.

dance of this mRNA type. Taken together, these results argue strongly for the authenticity of these upstream exons and their patterns as shown in Fig. 4.

Is UE1 neuronal cell specific? Although UE1-containing cDNAs are present in CEF, one cannot rule out the neuronal cell-specific expression of this exon, since CEF are derived from multiple tissue precursors. Therefore, experiments were undertaken to examine the tissue specificity of this 5′ cDNA sequence. S1 nuclease analysis was performed with the poly(A)$^+$ RNAs isolated from both chicken embryonic brain and limb tissues. The probe used in those studies (equivalent to type 3b cDNA) is shown diagrammatically in Fig. 7A. This 714-bp probe contains UE1, exon 1, and the first 322 bp of the *src* coding sequence. If the UE1 sequence is present in the c-*src* mRNAs analyzed, a fully protected fragment of 684 nucleotides would be generated after S1 digestion. The result of this analysis is shown in Fig. 7B. The radiolabeled probe that had not been treated with S1 is shown in lane 1. A protected fragment was not observed in the tRNA control reaction (lane 2). A 684-nucleotide fragment was protected after S1 digestion in samples of poly(A)$^+$ RNAs isolated from both embryonic limb (lanes 3 and 5) and brain (lanes 4 and 6) tissues, although there appeared to be more protection by the brain RNAs, consistent with its higher c-*src* RNA abundance (45, 80). This result suggests that the UE1 sequence is present in c-*src* mRNAs synthesized in neuronal and nonneuronal tissues. Similar results were observed with a probe that contained only the 5′ noncoding sequences (data not shown).

DISCUSSION

In this report, we describe the cloning and sequence of the 4-kb c-*src* mRNAs from CEF and brain, providing for the first time a characterization of the 5′ and 3′ noncoding exons. Since previous structural studies of c-*src* mRNA were performed on cDNA such as clones generated from retroviral vectors (used to remove introns from cloned genomic DNA), only a limited characterization of pp60$^{c\text{-}src}$ coding sequences has been presented (40, 46, 56). The 5′ noncoding sequences presented in this study provide evidence suggesting that there is considerable heterogeneity in the 5′ ends of the c-*src* mRNAs. Four types of 5′ exon sequences, UE1, UE2, UEX, and UEY, were observed. Distinct splicing events generate several c-*src* mRNAs by adding those alternative 5′ exons to a common exon (exon 1). A fifth class of c-*src* mRNA is produced by a unique combinatorial splicing of the upstream exon UE2 with a novel exon 1a arising from a region of c-*src* previously defined as intron 1. Characterization of these upstream exons has been made possible by the use of RACE, a PCR-based cDNA cloning technique (20). We believe that the observed heterogeneity of the 5′ ends of c-*src* mRNAs reflect the true complex organization and expression of the c-*src* gene rather than a result of the PCR cloning artifact for the following reasons: (i) occurrence of the same 5′ cDNA clones from both CEF and the brain c-*src* cDNA libraries prepared by independent cloning strategies, (ii) precise splicing of those 5′ c-*src* exons as predicted from the genomic sequence of the c-*src* DNA, (iii) incorporation of a similar 5′ exon into recovered viruses in the spontaneous c-*src* transduction (68), and (iv) the fact that the direct PCR amplification and RNase protection assay show that the expected size and exon contiguity are indeed represented in the CEF c-*src* mRNA species.

While the existence of exons upstream from exon 1 was predicted from earlier studies and primer extension analyses, the presence of an exon (1a) between exons 1 and 2 was not expected. A subsequent sequence comparison revealed that exon 1a has been identified previously in our analysis of recovered avian sarcoma viruses (rASVs) as a cryptic exon

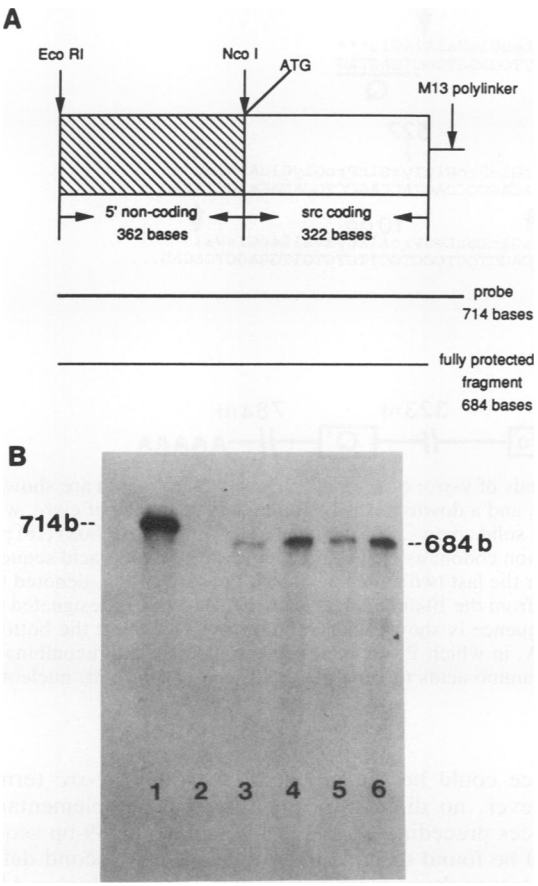

FIG. 7. S1 nuclease analysis of the UE1 sequence in the type 3b c-*src* clone. (A) The single-stranded probe (714 nucleotides in length) used in S1 nuclease analysis containing the entire 5′ noncoding region, including UE1, exon 1, the first 322 nucleotides of the *src* coding region, and the M13 polylinker sequences. The size of the fully protected fragment after S1 nuclease digestion is also indicated (684 nucleotides). In panel B, lanes 2 to 6 represent the results of S1 nuclease analysis after hybridization of 20 μg of calf intestinal tRNA (lane 2) or 5 μg of the various poly(A)$^+$ RNAs (lanes 3 to 6) with the radiolabeled probe (10^5 cpm per reaction) and digestion the products with 400 U of S1 nuclease at 25°C for 1 h (see Materials and Methods). The protected fragments generated after S1 nuclease digestion were analyzed on a 4% polyacrylamide gel–7 M urea denaturing gel and visualized by autoradiography. Lanes: 1, probe alone without S1 nuclease; 2, calf intestinal tRNA; 3 and 5, two independent isolates of chicken embryonic limb poly(A)$^+$ RNA; 4 and 6, two independent isolates of chicken embryonic brain poly(A)$^+$ RNA. b, bases.

in the intron 1 region with bona fide splice signals (68). Part or all of its sequence was transduced into three rASVs derived from a transformation-defective deletion mutant, *td*109 (68). The entire exon 1a was transduced into an rASV by using exactly the same splice sites (68). Current study on the 5′ c-*src* cDNA clones indicates that this presumed dormant exon is actually expressed among the CEF c-*src* mRNAs. This exon, however, cannot be part of the coding sequence since it contains termination codons in all three reading frames.

Alternative 5′ exons have been shown in other proto-oncogenes such as c-*abl* (3–5), *Drosophila* epidermal growth factor receptor homolog (62), and chicken c-*ets-1* (44). However, those alternative exons are in the coding regions of the respective proteins and hence are thought to influence the functions of these proto-oncogene products by forming different N-termini. The 5′ noncoding exons of the c-*src* mRNA may play an important role in regulating the expression of pp60$^{c\text{-}src}$ by influencing the stability of the c-*src* mRNA and its translational efficiency (8, 19, 42, 58, 59). The results from the RNase protection assay suggest that the mRNA species containing exon UE2 without exon 1a is the most abundant and that the species containing exon UE1 is the least abundant mRNA is CEFs. Whether these c-*src* mRNAs are expressed in a tissue-specific manner is an important issue that remains to be resolved. It would also be interesting to determine whether this 5′ exon heterogeneity is present in c-*src* mRNAs from mammalian species as well.

Our previous study of the muscle-specific 3-kb c-*src* mRNAs implied that expression of this mRNA (versus expression of the 4-kb pp60$^{c\text{-}src}$ mRNA) involves control at the levels of initiation, splicing, and polyadenylation (17). Although we have not definitely identified the initiation sites, that observation coupled with results for the heterogeneous 5′ cDNAs described here strongly suggests that there are multiple promoter sites for the c-*src* gene. The observed 5′ sequence heterogeneity of c-*src* mRNAs likely results from differential initiation and splicing. Since the 5′ end of the type 2a cDNA does not correspond to a splice acceptor site, we suspect that it may represent the 5′ terminus. Examination of its upstream c-*src* DNA sequence reveals an atypical TATA-like sequence 19 bases upstream of the 5′ end of the cDNA. However, no typical CCAAT box is found in its appropriate location relative to the presumed initiation site. Nevertheless, an AP1 binding site is present at 330 bases upstream of the exon UE2 (sequence not shown). Proof of this sequence as a promoter would require direct demonstration of its promoting activity in the initiation of transcription. Type 3b cDNA apparently contains at least one further upstream exon derived from a region 5′ to the c-*src* probe 2. Therefore, the promoter for this cDNA would have to be located upstream of the probe 2 region of the c-*src* DNA. Our data also suggest that exons UEX and UEY are derived from the region(s) at least 12 kb upstream from exon 1. Since the first step in the RACE method of amplifying the 5′ ends of an mRNA is reverse transcription, which is in effect a primer extension analysis, we consider the 5′ RACE cDNA clones to be a representative population of the products obtained by primer extension. Obviously, some of the clones analyzed did not reach the 5′ termini. Factors governing those premature terminations are unknown. While definitive physical and functional evidence for multiple c-*src* promoters awaits further study, we feel that our present data provide evidence that strongly suggests this possibility. However, at this point we cannot rule out the possibility that a single promoter gives rise to a short leader sequence that is then spliced alternatively to exons UEX, UEY, UE1, and UE2.

An unexpected finding in the 3′ sequence of the c-*src* mRNA is the presence of an open reading frame called *sdr*. Whether this reading frame is expressed as a spliced subgenomic mRNA in a tissue-specific manner remains to be seen. Interestingly, a computer search revealed that a small region of *sdr* reading frame is partly homologous to that of *sur*, which was shown previously to arise by alternative RNA splicing from the c-*src* locus in chicken muscle at around the time of hatching (17). The significance of these observations is not clear.

Possible mechanism for the generation of the 3′ v-*src*

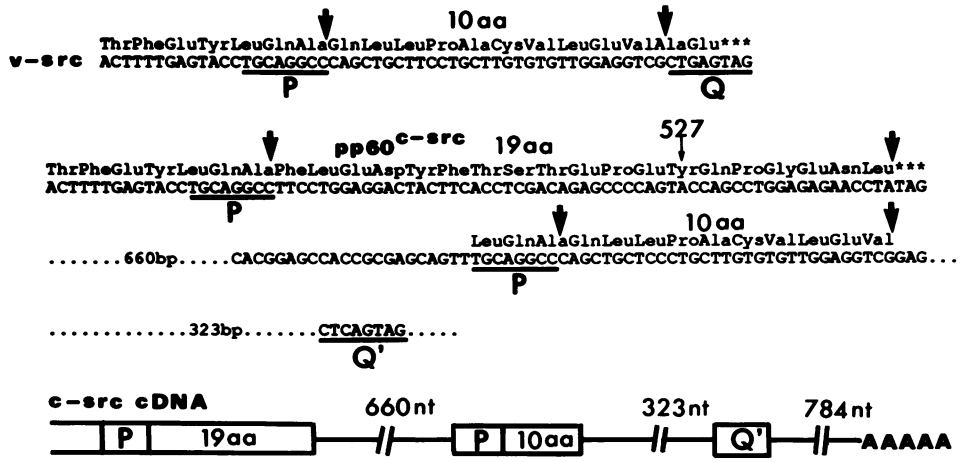

FIG. 8. Model for the genesis of the 3′ end of v-*src*. Sequences of the 3′ ends of v-*src*, c-*src*, and its downstream region are shown. The carboxyl-terminal 10 amino acids (aa) of v-*src*, the 19 amino acids of pp60$^{c\text{-}src}$, and a downstream 10-amino-acid sequence of c-*src*, which is thought to be the precursor for the last 10 amino acids of v-*src*, are bounded by solid arrows. The sequence TGCAGGCC (P box) (18) present 57 nucleotides upstream and 660 nucleotides downstream of the c-*src* termination codon, as well as preceding the 10-amino-acid sequence of v-*src*, is underlined. In the v-*src* sequence, the last eight nucleotides coding for the last two amino acids and the stop codon, denoted Q (18), are underlined. A similar sequence, CTCAGTAG, located 323 bp downstream from the 10-amino-acid sequence of c-*src*, is designated Q′ and underlined. The position of tyrosine 527 in the 3′ end of the c-*src* cDNA sequence is shown with a thin arrow. Shown at the bottom is a schematic illustration of the sequence motifs in the 3′ half of the c-*src* cDNA, in which P box is thought to mediate the recombination to replace the last 19 amino acids of pp60$^{c\text{-}src}$ with a new set of downstream 10 amino acids to generate the 3′ end of v-*src*. nt, nucleotides.

sequence. Formation of the 3′ end of RSV v-*src* appears to be rather complex, since the last 36 nucleotides are not contiguous with the upstream c-*src* sequence (75). Previous sequence analysis of the c-*src* DNA by Takeya and Hanafusa (75) showed that a 39-bp sequence corresponding to the 3′-terminal region of v-*src* was present in c-*src* DNA about 900 bp downstream from the termination codon of pp60$^{c\text{-}src}$. This finding suggested that a deletion removing sequences between the 39-bp region and the 3′ end of pp60$^{c\text{-}src}$ must occur during or after the recombination between the viral and cellular DNA. Our cDNA sequence indicates that the same 39-bp sequence of v-*src* is located 660 nucleotides downstream from the termination codon of pp60$^{c\text{-}src}$. The fact that this 3′-terminal v-*src* sequence is present in the c-*src* mRNA exon is consistent with the model of c-*src* transduction involving an RNA intermediate. The octanucleotide sequence TGCAGGCC is present in exon 12 and repeated at the 5′ eight nucleotides of the 39-bp sequence (box P; Fig. 8). This direct repeat could mediate joining of the 39-bp sequence to the c-*src* exon 12, resulting in the replacement of the original terminal 19 amino acids of pp60$^{c\text{-}src}$ with the 39-bp sequence. Our previous study of several *src* deletion mutants strongly suggested that direct repeats play a role in mediating the generation of those mutants (55). This most likely occurs at the step of negative-strand DNA synthesis by reverse transcriptase (18, 55, 77). The same process could mediate the deletion of sequences between exon 12 and the 39-bp sequence in an RNA intermediate molecule during or after the initial transduction of the c-*src* sequence. The last eight nucleotides of v-*src*, including the termination codon (CTGAGTAG; box Q in Fig. 8), are probably generated by a separate recombination event between the new c-*src* 3′ sequence and the 3′ region of the avian leukosis virus (ALV) genome as proposed previously (18). A similar sequence, CTCAGTAG (box Q′; Fig. 8), is present 323 nucleotides downstream from the 39-bp sequence. This raises the question of whether this Q′ sequence could be the precursor for the 3′ v-*src* terminus. However, no direct repeat or inverse complementary sequences preceding Q′ and at the end of the 39-bp sequence could be found to potentially mediate this second deletion. This leaves the recombination with the transducing ALV as a favorable mechanism for the formation of the v-*src* 3′ terminus. The fact that the same Q box is present in the 3′ regions of Y73, MH2, UR2, and CT10 (18, 50) viruses, which have captured different proto-oncogenes independently, is highly suggestive that the 3′ ends of their genomes have been derived from the respective ALV helper viruses. But the exact mechanism by which the progenitor ALV recombined with the c-*src* to generate the last two amino acids and the stop codon still remains unclear. Recombination within exon 12 and the presence of these sequence motifs in the c-*src* cDNA suggest that the right-hand recombination in the transduction of c-*src* involves an RNA molecule as the intermediate. Our detection of the 39-bp v-*src* sequence in the c-*src* cDNA lends further support for the model. The RNA-mediated model was also suggested for the transduction of the c-*fps* and c-*erbB* proto-oncogenes in which 3′ recombination appeared to have occurred within the poly(A) tracts of their mRNAs (37, 60). However, our data cannot rule out the possibility that the 3′ recombinational events occurred at the DNA level.

ACKNOWLEDGMENTS

We thank Hidesaburo Hanafusa and Song-Muh Jong for useful discussions and critical reading of the manuscript, Mingxu Lu for assistance in preparation of the figures, Jianmin Chen for pointing out the *sdr* reading frame, Marie-Cecil Raynal and Frances Smith for discussion of the PCR analysis of the 5′ cDNAs, Kenneth B. Marcu and A. Nepveu for help in the S1 nuclease analysis, and Anita Spingarn for help with preparation of the manuscript.

This work was supported by Public Health Service grant CA 49400 from the National Institutes of Health to L.-H.W. and Public Health Service grant CA 47572 from the National Institutes of Health to J.S.B.

REFERENCES

1. **Barnekow, A., and M. Gessler.** 1986. Activation of pp60$^{c\text{-}src}$ kinase during differentiation of monomyelocytic cells in vitro. EMBO J. **5**:701–705.
2. **Barnekow, A., and M. Schartl.** 1984. Cellular *src* gene product detected in the freshwater *Spongilla lacustris*. Mol. Cell. Biol. **4**:1179–1181.
3. **Ben-Neriah, Y., A. Bernards, M. Paskind, G. Q. Daley, and D. Baltimore.** 1986. Alternative 5′ exons in the c-*abl* mRNA. Cell **44**:577–586.
4. **Bernards, A., M. Paskind, and D. Baltimore.** 1988. Four murine c-*abl* mRNAs arise by usage of two transcriptional promoters and alternative splicing. Oncogene **2**:297–304.
5. **Bernards, A., C. M. Rubin, C. A. Westbrook, M. Paskind, and D. Baltimore.** 1987. The first intron in the human c-*abl* gene is at least 200 kilobases long and is a target for translocations in chronic myelogenous leukemia. Mol. Cell. Biol. **7**:3231–3236.
6. **Bishop, J. M., and H. E. Varmus.** 1985. Functions and origins of retroviral transforming genes, p. 249–356. *In* R. Weiss, N. Teich, H. E. Varmus, and J. Coffin (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
7. **Bolen, J. B., N. Rosen, and M. A. Israel.** 1985. Increased pp60$^{c\text{-}src}$ tyrosyl kinase activity in human neuroblastomas is associated with amino-terminal tyrosine phosphorylation of the src gene product. Proc. Natl. Acad. Sci. USA **82**:7275–7279.
8. **Brawerman, G.** 1987. Determinants of mRNA stability. Cell **48**:5–6.
9. **Brugge, J. S., P. Cotton, A. Lustig, W. Yonemoto, L. Lipsich, P. Coussens, J. N. Barrett, D. Nonner, and R. W. Keane.** 1987. Characterization of the altered form of the c-*src* gene product in neuronal cells. Genes Dev. **1**:287–296.
10. **Brugge, J. S., P. C. Cotton, A. E. Queral, J. N. Barrett, D. Nonner, and R. W. Keane.** 1985. Neurones express high levels of a structurally modified activated form of pp60$^{c\text{-}src}$. Nature (London) **316**:554–557.
11. **Cartwright, C. A., R. Simontov, W. M. Cowan, T. Hunter, and W. Eckhart.** 1988. pp60$^{c\text{-}src}$ expression in the developing rat brain. Proc. Natl. Acad. Sci. USA **85**:3348–3352.
12. **Cartwright, C. A., R. Simontov, P. O. Kaplan, T. Hunter, and W. Eckhart.** 1987. Alterations in pp60$^{c\text{-}src}$ accompany differentiation of neurons from rat embryo striatum. Mol. Cell. Biol. **7**:1830–1840.
13. **Collet, M. S., A. F. Purchio, and R. L. Erikson.** 1980. Avian sarcoma virus transforming protein pp60$^{c\text{-}src}$ shows protein tyrosine kinase activity specific for tyrosine. Nature (London) **285**:167–169.
14. **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.
15. **Courtneidge, S. A., A. D. Levinson, and J. M. Bishop.** 1980. The protein encoded by the transforming gene of avian sarcoma virus (pp60^{src}) and a homologous protein in normal cells (pp60^{protosrc}) are associated with the plasma membrane. Proc. Natl. Acad. Sci. USA **77**:3783–3787.
16. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. Plasmid **13**:31–40.
17. **Dorai, T., and L.-H. Wang.** 1990. An alternative non-tyrosine protein kinase product of the c-*src* gene in chicken skeletal muscle. Mol. Cell. Biol. **10**:4068–4079.
18. **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.
19. **Eick, D., M. Piechaczyk, B. Henglein, J. M. Blanchard, B. Traub, E. Kofler, S. Wiest, G. M. Lenoir, and G. W. Bornkamm.** 1985. Aberrant c-*myc* mRNAs of Burkitt lymphoma cells have longer half lives. EMBO J. **4**:3717–3725.
20. **Frohman, M. A., M. K. Dush, and G. R. Martin.** 1988. Rapid production of full length cDNAs from rare transcripts: amplification using a single gene specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85**:8998–9002.
21. **Fults, D. W., A. C. Towle, J. M. Lauder, and P. F. Maness.** 1985. pp60$^{c\text{-}src}$ in the developing cerebellum. Mol. Cell. Biol. **5**:27–32.
22. **Garber, E. A., and H. Hanafusa.** 1987. NH_2-terminal sequences of two *src* proteins that cause aberrant transformation. Proc. Natl. Acad. Sci. USA **84**:80–84.
23. **Gee, C. E., J. Griffen, L. Sastre, L. J. Miller, T. A. Springer, H. Piwnica-Worms, and T. M. Roberts.** 1986. Differentiation of myeloid cells is accompanied by increased levels of pp60$^{c\text{-}src}$ protein and kinase activity. Proc. Natl. Acad. Sci. USA **83**: 5131–5135.
24. **Gessler, M., and A. Barnekow.** 1984. Differential expression of the cellular oncogenes c-*src* and c-*yes* in embryonal and adult chicken tissues. Biosci. Rep. **4**:757–770.
25. **Gibbs, C. P., A. Tanaka, S. K. Anderson, J. Radul, J. Boar, A. Ridgway, H.-J. Kung, and D. J. Fujita.** 1985. Isolation and structural mapping of a human c-*src* gene homologous to the transforming v-*src* of Rous sarcoma virus. J. Virol. **53**:19–24.
26. **Golden, A., and J. S. Brugge.** 1988. The src oncogene, p. 149–173. *In* E. P. Reddy, A. M. Skalka, and T. Curran (ed.), The oncogene handbook. Elsevier Science Publishers, New York.
27. **Golden, A., and J. S. Brugge.** 1989. Thrombin treatment induces rapid changes in tyrosine phosphorylation in platelets. Proc. Natl. Acad. Sci. USA **86**:901–905.
28. **Golden, A., S. P. Nemeth, and J. S. Brugge.** 1986. Blood platelets express high levels of the pp60$^{c\text{-}src}$ specific tyrosine kinase activity. Proc. Natl. Acad. Sci. USA **83**:852–856.
29. **Gonda, T. J., D. K. Sheiness, and J. M. Bishop.** 1982. Transcripts from the cellular homologs of retroviral oncogenes: distribution among chicken tissues. Mol. Cell. Biol. **2**:617–624.
30. **Grandori, C., and H. Hanafusa.** 1988. pp60$^{c\text{-}src}$ is complexed with a cellular protein in subcellular compartments involved in exocytosis. J. Cell Biol. **107**:2125–2135.
31. **Hanafusa, H.** 1969. Rapid transformation of cells by Rous sarcoma virus. Proc. Natl. Acad. Sci. USA **63**:318–325.
32. **Hanafusa, H.** 1987. Activation of the c-*src* gene, p. 100–105. *In* P. Kahn and T. Graf (ed.), Oncogenes and growth control. Springer Verlag KG, Heidelberg.
33. **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.
34. **Hanks, S. K., A. M. Quinn, and T. Hunter.** 1988. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science **241**:42–52.
35. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
36. **Hoffman-Falk, H., P. Einat, B.-Z. Shilo, and F. M. Hoffmann.** 1983. Drosophila melanogaster DNA clones homologous to vertebrate oncogenes: evidence for a common ancestor to the src and abl cellular genes. Cell **32**:589–598.
37. **Huang, C.-C., N. Hay, and J. M. Bishop.** 1986. The role of RNA molecules in transduction of proto-oncogene c-*fps*. Cell **44**:935–940.
38. **Hunter, T., and J. A. Cooper.** 1985. Protein-tyrosine kinases. Annu. Rev. Biochem. **54**:897–930.
39. **Hunter, T., and B. M. Sefton.** 1980. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. Proc. Natl. Acad. Sci. USA **77**:1131–1135.
40. **Kaplan, P. L., S. Simon, C. A. Cartwright, and W. Eckhart.** 1987. cDNA cloning with a retrovirus expression vector: generation of a pp60$^{c\text{-}src}$ cDNA clone. J. Virol. **61**:1731–1734.
41. **Kato, J.-Y., T. Takeya, C. Grandori, H. Iba, J. B. Levy, and H. Hanafusa.** 1986. Amino acid substitutions sufficient to convert the nontransforming protein pp60$^{c\text{-}src}$ to a transforming protein. Mol. Cell. Biol. **6**:4155–4160.
42. **Kozak, M.** 1986. Regulation of protein synthesis in virus infected animal cells. Adv. Virus. Res. **31**:229–292.
43. **LeBeau, J. M., O. D. Weistler, and G. Walter.** 1987. An altered form of pp60$^{c\text{-}src}$ is expressed primarily in the central nervous system. Mol. Cell. Biol. **7**:4115–4117.
44. **Leprince, D., M. Duterque-Coquilland, R.-P. Li, C. Henry, A. Fluorens, B. Debuire, and D. Stehelin.** 1988. Alternative splicing

within the chicken c-*ets*-1 locus: implications for transduction within the E26 retrovirus c-*ets* proto-oncogene. J. Virol. **62:** 3233–3241.

45. **Levy, J. B., T. Dorai, L.-H. Wang, and J. S. Brugge.** 1987. The structurally distinct form of pp60$^{c\text{-}src}$ detected in neuronal cells is encoded by a unique c-*src* mRNA. Mol. Cell. Biol. **7:**4142–4145.
46. **Levy, J. B., H. Iba, and H. Hanafusa.** 1986. Activation of the transforming potential of pp60$^{c\text{-}src}$ by a single amino acid change. Proc. Natl. Acad. Sci. USA **83:**4228–4232.
47. **Maness, P. F.** 1986. pp60$^{c\text{-}src}$ encoded by the proto-oncogene c-*src* is a product of sensory neurons. J. Neurosci. Res. **16:**127–139.
48. **Maness, P. F., M. Aubry, C. G. Shores, L. Frame, and K. H. Pfenninger.** 1988. c-*src* gene product in developing rat brain is enriched in nerve growth cone membranes. Proc. Natl. Acad. Sci. USA **85:**5001–5005.
49. **Martinez, R., B. Mathey-Prevot, A. Bernards, and D. Baltimore.** 1987. Neuronal pp60$^{c\text{-}src}$ contains a six amino acid insertion relative to its non-neuronal counterpart. Science **237:**411–415.
50. **Mayer, B. J., M. Hamaguchi, and H. Hanafusa.** 1988. A novel viral oncogene with structural similarity to phospholipase C. Nature (London) **332:**272–275.
51. **Mayer, B. J., R. Jove, J. F. Krane, F. Poirier, G. Calothy, and H. Hanafusa.** 1986. Genetic lesions involved in temperature sensitivity of the *src* gene products of four Rous sarcoma virus mutants. J. Virol. **60:**858–867.
52. **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.
53. **Parker, R. C., H. E. Varmus, and J. M. Bishop.** 1981. Cellular homologue (c-src) of the transforming gene of Rous sarcoma virus: isolation, mapping and transcriptional analysis of c-src and flanking regions. Proc. Natl. Acad. Sci. USA **78:**5842–5846.
54. **Parsons, S. J., and C. E. Creutz.** 1986. pp60$^{c\text{-}src}$ activity detected in the chromaffin granule membrane. Biochem. Biophys. Res. Commun. **134:**736–742.
55. **Parvin, J. D., and L.-H. Wang.** 1984. Mechanism 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.
56. **Piwnica-Worms, H., D. R. Kaplan, M. Whitman, and T. M. Roberts.** 1986. Retrovirus shuttle vector for study of kinase activities of pp60$^{c\text{-}src}$ synthesized in vitro and overproduced in vivo. Mol. Cell. Biol. **6:**2033–2040.
57. **Pyper, J. M., and J. B. Bolen.** 1990. Identification of a novel neuronal c-*src* exon expressed in human brain. Mol. Cell. Biol. **10:**2035–2040.
58. **Rabbits, P. H., A. Forster, M. A. Stinson, and T. H. Rabbitts.** 1985. Truncation of exon 1 from c-myc gene results in prolonged mRNA stability. EMBO J. **4:**3727–3733.
59. **Raghow, R.** 1987. Regulation of mRNA turnover in eukaryotes. Trends. Biochem. Sci. **12:**358–360.
60. **Raines, M. A., N. J. Maihle, C. Moscovici, L. Crittenden, and H.-J. Kung.** 1988. Mechanism of c-*erbB* transduction: newly released transducing virus retain poly(A) tracts of *erbB* transcripts and encode C-terminally intact *erbB* proteins. J. Virol. **62:**2437–2443.
61. **Schartl, M., and A. Barnekow.** 1984. Differential expression of the cellular *src* gene during vertebrate development. Dev. Biol. **105:**415–422.
62. **Schejter, E. D., D. Segal, L. Glazer, and B.-Z. Shilo.** 1986. Alternative 5′ exons and the tissue specific expression of the Drosophila EGF receptor homolog transcripts. Cell **46:**1091–1101.
63. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32:**853–869.
64. **Sefton, B. M., T. Hunter, and K. Beemon.** 1980. Relationship of the polypeptide products of the transforming gene of Rous sarcoma virus and the homologous gene of vertebrates. Proc. Natl. Acad. Sci. USA **72:**2059–2063.
65. **Shalloway, D., A. D. Zelenetz, and G. M. Cooper.** 1981. Molecular cloning and characterization of the chicken gene homologous to the transforming gene of Rous sarcoma virus. Cell **24:**531–541.
66. **Shaw, G., and R. Kamen.** 1986. A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell **46:**659–667.
67. **Simon, M. A., B. Drees, T. Kornberg, and J. M. Bishop.** 1985. The nucleotide sequence and the tissue specific expression of Drosophila c-*src*. Cell **42:**831–840.
68. **Soong, M.-M., S. Iijima, and L.-H. Wang.** 1986. Transduction of c-*src* coding and intron sequences by a transformation defective deletion mutant of Rous sarcoma virus. J. Virol. **59:**556–563.
69. **Sorge, L. K., B. T. Levy, and P. F. Maness.** 1984. pp60$^{c\text{-}src}$ is developmentally regulated in the neural retina. Cell **36:**249–257.
70. **Spector, D. H., H. E. Varmus, and J. M. Bishop.** 1978. Characteristics of cellular RNA related to transforming gene of avian sarcoma viruses. Cell **13:**381–386.
71. **Stehelin, D., H. E. Varmus, J. M. Bishop, and P. K. Vogt.** 1976. DNA related to the transforming gene(s) of avian sarcoma viruses is present in the normal avian DNA. Nature(London) **260:**170–173.
72. **Sudol, M., A. Alvarez-Buylla, and H. Hanafusa.** 1988. Differential developmental expression of c-*yes* and c-*src* protein in cerebellum. Oncogene Res. **2:**345–355.
73. **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.
74. **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 encodes for gp37 and pp60$^{c\text{-}src}$. J. Virol. **44:**1–11.
75. **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.
76. **Takeya, T., H. Hanafusa, R. P. Junghans, G. Ju, and A. M. Skalka.** 1981. Comparison between the viral transforming gene (*src*) of recovered avian sarcoma virus and its cellular homolog. Mol. Cell. Biol. **1:**1024–1037.
77. **Wang, L.-H.** 1987. The mechanism of transduction of proto-oncogene c-*src* by avian retroviruses. Mutat. Res. **186:**135–147.
78. **Wang, L.-H., and P. Duesberg.** 1974. Properties and location of poly(A) in Rous sarcoma virus RNA. J. Virol. **14:**1515–1529.
79. **Wang, L.-H., B. Edelstein, and B. J. Mayer.** 1984. Induction of tumors and generation of recovered sarcoma viruses by, and mapping deletions in, two molecularly cloned *src* deletion mutants. J. Virol. **50:**904–913.
80. **Wang, L.-H., S. Iijima, T. Dorai, and B. Lin.** 1987. Regulation of the expression of proto-oncogene c-*src* by alternative RNA splicing in chicken skeletal muscle. Oncogene Res. **1:**43–59.
81. **Wang, S.-Y., W. S. Hayward, and H. Hanafusa.** 1977. Genetic variation in the RNA transcripts of endogenous virus genes in uninfected chicken cells. J. Virol. **24:**64–73.