JOURNAL OF VIROLOGY, May 1987, p. 1731–1734
0022-538X/87/051731-04$02.00/0
Copyright © 1987, American Society for Microbiology
Vol. 61, No. 5

cDNA Cloning with a Retrovirus Expression Vector: Generation of a pp60$^{c\text{-}src}$ cDNA Clone

PAUL L. KAPLAN,†* SUZANNE SIMON, CHRISTINE A. CARTWRIGHT, AND WALTER ECKHART
The Salk Institute for Biological Studies, San Diego, California 92138

Received 28 October 1986/Accepted 9 February 1987

We used a murine retroviral expression vector, containing a genomic clone of the chicken c-*src* gene, a bacterial origin of replication, and a selectable marker, to remove 10 introns from the c-*src* gene. All 10 introns were removed accurately, and no mutations were introduced. The processed gene encoded a functional pp60$^{c\text{-}src}$ protein tyrosine kinase.

Retroviral vectors can be used to remove introns from cloned cellular genes (3, 11, 25, 27). The coding region of the chicken c-*src* gene contains 11 exons and 10 introns (29). Levy et al. (17) have constructed a cDNA clone of the c-*src* gene by combining restriction fragments of v-*src* genes, which lack introns, with a restriction fragment of the c-*src* gene containing the 3′ end of the coding region, where the sequences of the v-*src* and c-*src* genes diverge (29). One remaining intron was removed with a retrovirus (17). Piwnica-Worms et al. (20) used a similar procedure to remove a single intron from a hybrid *src* gene. We report here the use of a murine retroviral vector to remove all 10 introns from the coding region of the c-*src* gene. The cloning system was designed to minimize the number of times the gene was transcribed, thereby minimizing the possibilities for mutation (30). We verified the nucleotide sequence of the c-*src* coding region and the accurate removal of introns by DNA sequencing of two independently derived clones.

To generate a c-*src* cDNA clone, 20 μg of p*neo*MLV (Fig. 1A) containing an 8.1-kilobase-pair (kbp) c-*src* genomic clone (24) (pc-*src*$^{8.1g}$*neo*MLV) (Fig. 1B) was transfected onto 10^6 ψ-Am22b cells (8) in 9-cm dishes by the calcium phosphate coprecipitation method (13, 31). The precipitate was removed after 8 h, and the medium was harvested 40 h later. The primary RNA transcript of the plasmid was expected to be processed to remove introns and packaged into retrovirus particles. Samples of 2×10^6 NIH 3T3 cells in 9-cm dishes were infected with 5 ml each of the virus-containing supernatant (approximately 10^3 G418 CFU/ml). Twenty-four hours after infection, low-molecular-weight DNA containing circular proviral DNA molecules produced by reverse transcription was isolated (15). Because p*neo*MLV contains a bacterial origin of replication, this circular molecule can replicate as a bacterial plasmid. Therefore, the proviral DNA could be recovered directly in bacteria after a single step of transcription from transfected DNA and a single step of reverse transcription from viral RNA. Proviral DNA was extracted twice with phenol-chloroform and precipitated twice with ethanol. One-tenth of the resuspended pellet was used to transform *Escherichia coli* DH-5 to kanamycin resistance. Colonies were screened by hybridization (14) with the 8.1-kbp c-*src* *Bam*HI fragment labeled by nick translation (21). Two independent transfections produced 73 kanamycin-resistant colonies. Plasmid DNA from 32 clones was analyzed by restriction mapping. Thirty clones contained pc-*src*$^{8.1g}$*neo*MLV molecules, most likely resulting from carry-over of DNA from transfection of ψ-Am22b cells. (A carry-over of 0.01% would be sufficient to account for this number of colonies.) Plasmids from two clones contained a 3.1-kbp *Bam*HI fragment, the size expected for a processed c-*src* genomic fragment lacking introns (Fig. 1B). Analysis with eight other restriction enzymes supported this interpretation. Furthermore, both plasmids contained a new 0.59-kbp *Sma*I fragment predicted to be generated by circularization of linear proviral DNA after reverse transcription. Therefore, the two plasmids appeared to contain processed copies of the c-*src* gene in the form of a retrovirus replicative intermediate.

To confirm the nature of the c-*src* clones, we analyzed the DNA sequence of the coding region for the pp60$^{c\text{-}src}$ protein. We also analyzed the sequence of a c-*src* cDNA clone, p5H-2 (17), generously provided by H. Hanafusa. Appropriate restriction fragments were inserted into the replicative form of the coliphage vector M13mp19 for sequencing by the enzymatic method of Sanger et al. (22). The DNA sequences of the coding regions of all three clones were identical, confirming the assignment of intron-exon boundaries reported by Takeya and Hanafusa (29). All 10 introns were removed accurately from the genomic clone. The DNA sequences of the c-*src* cDNA clones differed from the revised sequence of chicken c-*src* DNA (17) at nucleotides 209 and 270. The differences at nucleotides 209 (here C) and 270 (here T) do not change the amino acids encoded by the respective codons.

To study the protein encoded by the cDNA clone, we inserted the coding sequence of the cDNA clone (Fig. 1B) into p*neo*MLV to create the plasmid pc-*src*^{wt}*neo*MLV. G418-resistant clones were isolated after transfection of 10^6 NIH 3T3 cells with 5 μg of pc-*src*^{wt}*neo*MLV. Proteins were immunoprecipitated from lysates of [^{35}S]methionine-labeled cultures by using an antibody, EC10 (18), that is specific for chicken pp60$^{c\text{-}src}$. Immunoprecipitates were tested for pp60$^{c\text{-}src}$ protein tyrosine kinase activity. The radiolabeled proteins were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Results from a representative clone are shown in Fig. 2. pp60$^{c\text{-}src}$ from transfected NIH 3T3 cells, radiolabeled in vivo with [^{35}S]methionine (Fig. 2, lane 3) or in vitro with [γ-^{32}P]ATP (lane 6), had the same mobility as pp60$^{c\text{-}src}$ from chicken embryo cells (lanes 2 and 5). The ratio of ^{32}P incorporation to ^{35}S incorporation in pp60$^{c\text{-}src}$ was the same for the recombinant protein and the authentic chicken pp60$^{c\text{-}src}$, suggesting that the two proteins

* Corresponding author.
† Present address: Division of Cancer Genetics, Dana-Farber Cancer Institute, Boston, MA 02115.

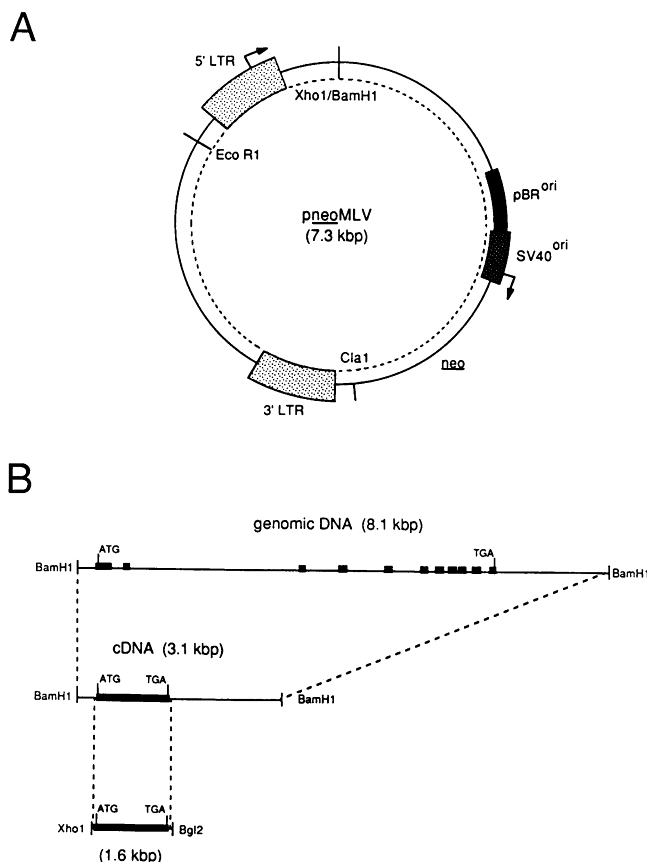

FIG. 1. DNAs used in the isolation and analysis of pp60$^{c\text{-}src}$ clones. (A) Plasmid p*neo*MLV. The plasmid was constructed by joining the following restriction fragments. (i) *Eco*RI to *Bam*HI of modified pEVC4 (A. D. Miller and I. M. Verma, personal communication), containing cellular flanking sequences and the 5′ Moloney murine sarcoma virus long terminal repeat. pEVC4 was derived from an infectious proviral clone of Moloney murine sarcoma virus (inserted at the *Eco*RI site of pBR322) by deleting sequences between the outermost viral *Pst*I sites. The 3′ long terminal repeat (contained in a *Cla*I to *Eco*RI fragment) was subsequently replaced with a homologous fragment containing the 3′ long terminal repeat of a Moloney murine leukemia virus clone. pEVC4 was modified by the successive insertion of synthetic linkers at the single remaining proviral *Pst*I site, creating the sequence 5′-CTGCAGCCCTCGAGG CGGATCCGCCTCGAGGGCTCGAG-3′ (5′-*Pst*I-*Xho*I-*Bam*HI-*Xho*I-*Pst*I-3′). (ii) *Bam*HI to *Cla*I of a modified pSV2*neo* (28) containing the pBR322 origin of DNA replication, the simian virus 40 early promoter, and the *neo* gene (pSV2*neo* was modified by inserting a *Bam*HI linker into the *Hin*cII site in the *amp* gene and a *Cla*I linker into the *Sma*I site downstream of the putative *neo* terminator codon). (iii) *Cla*I to *Eco*RI of pEVC4 containing the 3′ long terminal repeat and cellular flanking sequences of Moloney murine leukemia virus. (B) Chicken c-*src* genes. The 8.1-kbp *Bam*HI fragment obtained from plasmid pCS12.13 (23) is shown on the top line, with the 11 coding exons indicated by boxes. The locations of the initiator and terminator codons are indicated. The 3.1-kbp *Bam*HI fragment recovered in *E. coli* and its structure relative to the genomic clone are shown in the middle line. Its generation is described in the text. A 1.8-kbp fragment from the 5′ *Bam*HI site to a *Sac*I site 5 base pairs 3′ to the terminator codon was cloned into M13 for sequencing. After the insertion of an *Xho*I linker at the initiator codon (*Nco*I site) using the synthetic adapter 5′-CATGGTCTCGAGAC-3′ and a *Bgl*II linker at the *Sac*I site, a 1.6-kbp fragment shown in the lower line was cloned into p*neo*MLV, creating pc-*src*^{wt}*neo*MLV, which was used for expression studies.

had the same specific activity. The specific activities, as measured by phosphorylation of exogenous substrates added to the protein kinase reactions, were also the same (data not shown).

We further characterized pp60$^{c\text{-}src}$ radiolabeled in vivo with $^{32}P_i$ and immunoprecipitated with antibody EC10 by partial proteolytic mapping with *Staphylococcus aureus* V8 protease (6) (Fig. 2). The ^{32}P-labeled pp60$^{c\text{-}src}$ from transfected mouse cells (Fig. 2, lane 8) and from chicken cells (lane 7) had the same mobility. The recombinant protein had the same four major V8 phosphopeptides as chicken embryo cell pp60$^{c\text{-}src}$ (compare lane 10 with lane 9). For both proteins, peptides V1, V3, and V4 contained phosphoserine and peptide V2 contained phosphotyrosine (data not shown).

We characterized the phosphorylation sites of the recombinant pp60$^{c\text{-}src}$ by two-dimensional tryptic phosphopeptide analysis (Fig. 3). pp60$^{c\text{-}src}$ from chicken embryo cells (Fig. 3A) contained the four major phosphopeptides identified previously (5). Peptides 1 and 2 contained phosphotyrosine (data not shown). Both peptides contain the phosphorylated amino acid tyrosine 527 (5, 9). Peptides 3 and 4 contained phosphoserine (data not shown). Peptide 4 contains serine 17 (7, 19). The identity of the phosphorylated serine residue in peptide 3 is not known. The tryptic phosphopeptides of pc-*src*^{wt}*neo*MLV-encoded chicken pp60$^{c\text{-}src}$ expressed in

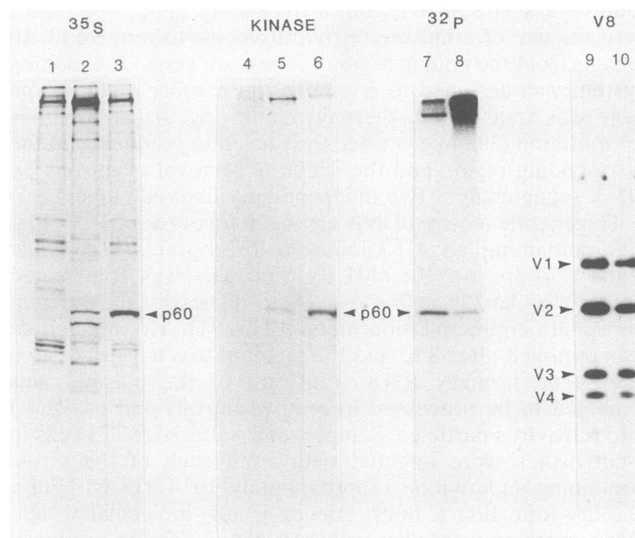

FIG. 2. Characterization of pp60$^{c\text{-}src}$ encoded by the c-*src* cDNA clone. Cells were metabolically labeled with 100 μCi of [^{35}S]methionine or 2.5 mCi of $^{32}P_i$ per ml for 18 h as previously described (4, 5). Proteins were immunoprecipitated from cell lysates with monoclonal antibody EC10 (18) (antibody excess). Immunoprecipitates of [^{35}S]methionine-labeled proteins were divided into two equal parts, one of which was phosphorylated in vitro with [γ-^{32}P]ATP (4, 10) (lanes 4 through 6). ^{32}P-labeled proteins (lanes 7 and 8) were excised from the gel and digested with *S. aureus* V8 protease, and the peptides were separated on a 12% sodium dodecyl sulfate-polyacrylamide gel (lanes 9 and 10). [^{35}S]methionine-containing gels were impregnated with 2,5-diphenyloxazole (2) before being dried. Exposure times with presensitized film were: lanes 1 to 3, 6.5 h at −70°C; lanes 4 to 6, 5 h at room temperature; lanes 7 and 8, 45 min with an intensifying screen (16) at −70°C; lanes 9 and 10, 13 h with an intensifying screen at −70°C. Lanes 1 and 4, Control NIH 3T3 cells; lanes 2, 5, 7, and 9, chicken embryo cells (4 days after explantation of 9- to 10-day embryos); lanes 3, 6, 8, and 10: NIH 3T3 cells expressing c-*src* encoded by pc-*src*^{wt}*neo*MLV.

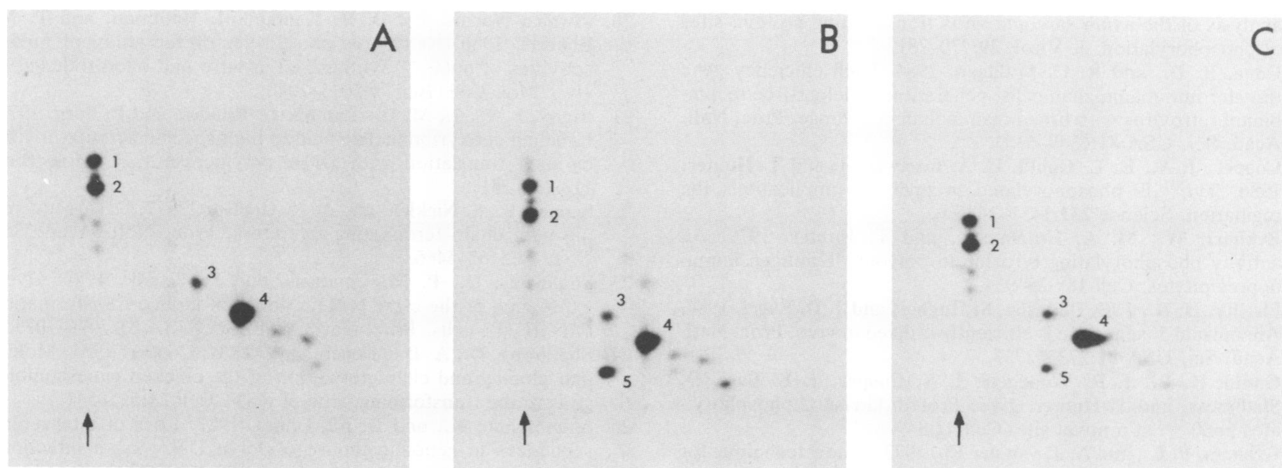

FIG. 3. Tryptic phosphopeptides of chicken and recombinant pp60$^{c\text{-}src}$. pp60$^{c\text{-}src}$ from chicken embryo cells (4 days after explantation of 9- to 10-day embryos) or from NIH 3T3 cells expressing the cDNA clone were extracted from the gel shown in Fig. 3, lanes 7 and 8. The proteins were oxidized with performic acid, digested with trypsin, and separated on cellulose thin-layer plates by electrophoresis at pH 1.9 in the horizontal direction (anode on the left) and chromatography in the vertical direction (1,5). The origin is indicated with an arrow. The identity of the numbered peptides is discussed in the text. (A) pp60$^{c\text{-}src}$ from chicken embryo cells; (B) pp60$^{c\text{-}src}$ from NIH 3T3 cells expressing c-*src* encoded by pc-*src*^{wt}*neo*MLV; (C) mixture of pp60$^{c\text{-}src}$ peptides from panels A and B. Cerenkov counts per minute loaded: (A) 480; (B) 490; (C) 480 of panel A and 490 of panel B. Presensitized film was exposed for 4 days (A and B) or 2 days (C) at −70°C with an intensifying screen (16).

NIH 3T3 cells are shown in Fig. 3B. The peptides were similar to those shown in Fig. 3A; however, there was an additional serine-containing phosphopeptide (peptide 5). Other experiments (not shown) identified serine 12 as the phosphorylated amino acid in peptide 5. This site is phosphorylated in chicken, but not mouse, pp60$^{c\text{-}src}$ in NIH 3T3 cells expressing both proteins (12) and in pp60$^{c\text{-}src}$ from chicken embryo cells treated with agents that activate protein kinase C (12). A mixture of tryptic peptides from the authentic and recombinant pp60$^{c\text{-}src}$ proteins (Fig. 3C) confirmed the identity of phosphopeptides 1 through 4 from the two proteins. Therefore, by three criteria (in vitro protein kinase activity, partial peptide mapping with *S. aureus* V8 protease, and two-dimensional tryptic phosphopeptide mapping) the protein encoded by the cDNA clone of chicken c-*src* was indistinguishable from the pp60$^{c\text{-}src}$ expressed in normal chicken embryo cells.

We isolated c-*src* cDNA clones directly in bacterial plasmids after transfection with proviral DNA. The amount of proviral DNA available for transfection is limited by the virus titer and the necessity of isolating DNA during the first round of infection. The amount of proviral DNA might be increased by using other packaging cell lines or by ligating linear proviral DNA molecules to form circular molecules before transfection.

The accurate removal of all 10 introns from the genomic c-*src* clone suggests that all the information necessary for the accurate removal of the introns was present in the 8.1-kbp genomic fragment. The size of the gene that can be processed in this way may be limited only by the size of the RNA that can be incorporated into a retrovirus particle. That limit has not been firmly established. The genome size of Moloney murine leukemia virus is slightly larger than 8.3 kilobases (26), suggesting that p*neo*MLV could accommodate processed inserts of at least 4.2 kbp.

The availability of cDNA clones of c-*src* will facilitate studies of the structure and function of the c-*src* gene. For example, it should now be possible to characterize the c-*src* promoter and the structure of the RNA products of the gene. We are presently using the cDNA clone to study the effects of mutations on the protein kinase activity and transforming ability of pp60$^{c\text{-}src}$.

We thank David Shalloway for providing us with the genomic clone of chicken c-*src*, Inder Verma and Dusty Miller for plasmid pEVC4, Joan Levy and Hidesaburo Hanafusa for their c-*src* cDNA clone, and Clair Kelley and Bart Sefton for providing chicken cells. We also thank Joan Brugge and Sarah Parsons for antisera, and Gerry Weinmaster for advice on DNA sequencing. We are grateful to Gary Glenn, Tony Hunter, David Meek, and Bart Sefton for comments on the manuscript.

P.L.K. was the recipient of a Damon Runyon-Walter Winchell Fellowship (DRG-664). This work was supported by Public Health Service grants CA-09370, CA-13884, CA-14195, and CA-01040 from the National Cancer Institute and by a grant from the Weingart Foundation.

LITERATURE CITED

1. **Beemon, K., and T. Hunter.** 1978. Characterization of Rous sarcoma virus *src* gene products synthesized in vitro. J. Virol. **28**:551–566.
2. **Bonner, W. M., and R. A. Laskey.** 1974. A film detection method for tritium-labeled proteins and nucleotides in polyacrylamide gels. Eur. J. Biochem. **46**:83–88.
3. **Brown, A. M. C., R. S. Wildin, T. J. Prendergast, and H. E. Varmus.** 1986. A retrovirus vector expressing the putative mammary oncogene *int*-1 causes partial transformation of a mammary epithelial cell line. Cell **46**:1001–1009.
4. **Cartwright, C. A., M. A. Hutchinson, and W. Eckhart.** 1985. Structural and functional modification of pp60$^{c\text{-}src}$ associated with polyoma middle tumor antigen from infected or transformed cells. Mol. Cell. Biol. **5**:2647–2652.
5. **Cartwright, C. A., P. L. Kaplan, J. A. Copper, T. Hunter, and W. Eckhart.** 1986. Altered sites of tyrosine phosphorylation in pp60$^{c\text{-}src}$ associated with polyomavirus middle tumor antigen. Mol. Cell. Biol. **6**:1562–1570.
6. **Cleveland, D. W., S. G. Fisher, M. W. Kirschner, and U. K. Laemmli.** 1977. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. J. Biol. Chem. **252**:1102–1106.
7. **Collett, M. S., E. Erikson, and R. L. Erikson.** 1979. Structural

analysis of the avian sarcoma virus transforming protein: sites of phosphorylation. J. Virol. **29**:770–781.

8. **Cone, R. D., and R. C. Mulligan.** 1984. High efficiency gene transfer into mammalian cells: generation of helper-free recombinant retrovirus with broad mammalian host range. Proc. Natl. Acad. Sci. USA **81**:6349–6353.
9. **Cooper, J. A., K. L. Gould, C. A. Cartwright, and T. Hunter.** 1986. Tyr^{527} is phosphorylated in pp60$^{c\text{-}src}$: implications for regulation. Science **231**:1431–1433.
10. **Eckhart, W., M. A. Hutchinson, and T. Hunter.** 1979. An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates. Cell **18**:925–934.
11. **Fischer, H. D., J. B. Dodgson, S. Hughes, and J. D. Engel.** 1984. An unusual 5′ sequence is efficiently utilized in vivo. Proc. Natl. Acad. Sci. USA **81**:2733–2737.
12. **Gould, K. L., J. R. Woodgett, J. A. Cooper, J. E. Buss, D. Shalloway, and T. Hunter.** 1985. Protein kinase C phosphorylates pp60$^{c\text{-}src}$ at a novel site. Cell **42**:849–857.
13. **Graham, F. L., and A. J. van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456–467.
14. **Grunstein, M., and D. S. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain specific genes. Proc. Natl. Acad. Sci. USA **72**:3961–3965.
15. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–369.
16. **Laskey, R. A., and A. D. Mills.** 1977. Enhanced autoradiographic detection of ^{32}P and ^{125}I using intensifying screens and hypersensitized film. FEBS Lett. **82**:314–316.
17. **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.
18. **Parsons, S. J., D. J. McCarley, C. Ely, D. C. Benjamin, and J. T. Parsons.** 1984. Monoclonal antibodies to the Rous sarcoma virus pp60^{src} react with enzymatically active pp60^{src} of avian and mammalian origin. J. Virol. **51**:272–282.
19. **Patschinsky, T., T. Hunter, and B. M. Sefton.** 1986. Phosphorylation of the transforming protein of Rous sarcoma virus: direct demonstration of phosphorylation of serine-17 and identification of an additional site of tyrosine phosphorylation in pp60$^{v\text{-}src}$ of Prague Rous sarcoma virus. J. Virol. **59**:73–81.
20. **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.
21. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
22. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5466.
23. **Shalloway, D., P. M. Coussens, and P. Yacuik.** 1984. Overexpression of the c-*src* protein does not induce transformation of NIH 3T3 cells. Proc. Natl. Acad. Sci. USA **81**:7071–7075.
24. **Shalloway, D., A. D. Zelenetz, and G. M. Cooper.** 1981. Molecular cloning and characterization of the chicken gene homologous to the transforming gene of RSV. Cell **24**:531–541.
25. **Shimotohno, K., and H. M. Temin.** 1982. Loss of intervening sequences in genomic mouse α-globin DNA in an infectious retrovirus vector. Nature (London) **299**:265–268.
26. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293**:543–548.
27. **Sorge, J., and S. Hughes.** 1982. Splicing of intervening sequences introduced into an infectious retroviral vector. J. Mol. Appl. Genet. **1**:547–559.
28. **Southern, P. J., and P. Berg.** 1982. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. **1**:327–341.
29. **Takeya, T., and H. Hanafusa.** 1983. Structure and sequence of the cellular gene homologous to the Rous sarcoma virus *src* gene and the mechanism of generating the transforming virus. Cell **32**:881–890.
30. **Weiss, R., N. Teich, H. Varmus, and J. Coffin.** 1982. RNA tumor viruses, p. 369–512. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.
31. **Wigler, M., A. Pellicer, S. Silverstein, R. Axel, G. Urlaub, and L. Chasin.** 1979. DNA mediated transfer of the adenine phosphoribosyltransferase locus into mammalian cells. Proc. Natl. Acad. Sci. USA **76**:1373–1376.