MOLECULAR AND CELLULAR BIOLOGY, Nov. 1982, p. 1304–1319
0270-7306/82/111304-16$02.00/0

Vol. 2, No. 11

Construction of a Modular Dihydrofolate Reductase cDNA Gene: Analysis of Signals Utilized for Efficient Expression

RANDAL J. KAUFMAN[1] AND PHILLIP A. SHARP[2]*

Center for Cancer Research[1] and Department of Biology,[2] Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received 3 June 1982/Accepted 27 July 1982

Dihydrofolate reductase (DHFR) modular genes have been constructed with segments containing the adenovirus major late promoter, a 3′ splice site from a variable region immunoglobulin gene, a DHFR cDNA, and portions of the simian virus 40 (SV40) genome. DNA-mediated transfer of these genes transformed Chinese hamster ovary DHFR$^-$ cells to the DHFR$^+$ phenotype. Transformants contained one to several copies of the transfected DNA integrated into the host genome. Clones subjected to growth in increasing concentrations of methotrexate eventually gave rise to lines containing several hundred copies of the transforming DNA. Analysis of the DHFR mRNA produced in amplified lines indicated the following. (i) All clones utilize the adenovirus major late promoter for transcription initiation. (ii) A hybrid intron formed by the 5′ splice site of the adenovirus major late leader and a 3′ splice site from a variable-region immunoglobulin gene is properly excised. (iii) The mRNA is not efficiently polyadenylated at sequences in the 3′ end of the DHFR cDNA but rather uses polyadenylation signals downstream from the DHFR cDNA. Three independent clones produce a DHFR mRNA containing SV40 or pBR322 and SV40 sequences, and the RNA is polyadenylated at the SV40 late polyadenylation site. Another clone has recombined into cellular DNA and apparently uses a cellular sequence for polyadenylation. Introduction of a segment containing the SV40 early polyadenylation signal into the 3′ end of the DHFR cDNA gene generated a recombinant capable of transforming cells to the DHFR$^+$ phenotype with at least a 10-fold increase in efficiency, demonstrating the necessity for an efficient polyadenylation signal. Attachment of a DNA segment containing the transcription enhancer (72-base pair repeat) of SV40 further increased the biological activity of the modular DHFR gene 50- to 100-fold.

Sequence requirements for mRNA biogenesis in higher cells are quickly being defined with the advent of recombinant DNA technology and the availability of multiple systems to study RNA transcription, initiation, and termination, RNA processing, RNA polyadenylation, and further RNA transport, stabilization, and translation. It is now possible to modify DNA in vitro and examine the effect by utilizing in vitro reconstituted systems (38, 60) or by its introduction into cells by DNA transduction with viral DNA vectors (27, 28, 43), by its transient expression after DNA transfection (17, 42), by its stable introduction into cells after cotransfection with selectable markers (53, 61), or by its microinjection into oocytes, mammalian cells, or even embryos (11, 13, 25; M. P. Wickens and R. A. Laskey, *in* R. Williamson, ed., *Genetic Engineering*, vol. 1, in press). The dihydrofolate reductase (DHFR) gene and methotrexate (MTX) selection offer a unique opportunity for the introduction of an essential gene into a DHFR-deficient cell line (16), with the potential for further amplification of the transfected DNA by selection of cells in progressively increasing concentrations of MTX (3, 46). In addition, the level of expression from altered or chimeric genes may be easily monitored by MTX resistance, DHFR enzyme activity, or the use of a fluorescent derivative of MTX which quantitatively binds intracellular DHFR (30).

We describe the construction of a modular DHFR cDNA gene which is capable of efficient expression as monitored by its ability to transform cells deficient in DHFR to the DHFR$^+$ phenotype. Signals which increase DHFR expression are defined. This modular gene provides a sensitive, convenient assay for DNA

segments that affect gene expression, as well as a means of amplifying DNA segments in mammalian cells.

MATERIALS AND METHODS

Cell culture. The DHFR-deficient DUKX-CHO clone has been described previously (16). Clones transformed to the DHFR$^+$ phenotype were isolated and selected for stepwise MTX resistance. Unless otherwise noted, clones examined were resistant to 1 μM MTX. These clones contain 100 to 300 copies of the original plasmid DNA integrated into the Chinese hamster ovary (CHO) genome (details in R. J. Kaufman and P. A. Sharp, J. Mol. Biol., in press). The 3T3-R500 murine line has been described previously (12) and has been provided by P. Brown. Calcium phosphate-mediated DNA transfection (26) into the DHFR-deficient cells and selection for the DHFR$^+$ phenotype were described earlier (Kaufman and Sharp, in press).

Recombinant DNA. The construction of recombinant clones pDHFR26, pAdD26-1, pASD11, and pASD12 has been described (15; Kaufman and Sharp, in press). The region encoding the modular DHFR gene in recombinant pAdD26-1 has been transferred to plasmid pSV0d (42) to obtain pAdD26SV0d. pSV0d contains a 200-base pair (bp) segment of simian virus 40 (SV40) spanning the origin of replication (from the *Hin*dIII site [0.65 map unit] in the early region of SV40 to the *Eco*RII site [0.693 map unit] in the late region) between the *Eco*RI and *Hin*dIII sites of pBR322 and also has a 1.1-kilobase (kb) deletion in pBR322 which has removed sequences detrimental for propagation in mammalian cells (36). pAdD26-1 DNA was digested completely with *Eco*RI and partially with *Pst*I, and the 2,400-bp fragment containing the DHFR gene was isolated from an agarose gel by incubating the gel slice in sodium iodide (90%) at 37°C to dissolve the gel and then adding powdered silica dioxide and incubating for 1 h at 4°C. The silica dioxide was pelleted by centrifugation and washed sequentially in cold sodium iodide, cold 70% ethanol (twice), and 10 mM Tris (pH 7.5)–100 mM NaCl–1 mM EDTA. The DNA was eluted from the silica dioxide by incubation in 30 μl with 10 mM Tris (pH 7.5)–1 mM EDTA. DNA concentration was determined by microdot and used directly for ligation. The 2,400-bp fragment was ligated to a 2,700-bp fragment of pSV0d DNA which was similarly isolated after *Eco*RI and *Pst*I complete digestion.

Clone pAdD26SV(A) has been constructed from pAdD26SV0d by the insertion of a sequence encoding the early polyadenylation site of SV40. SV40 (strain 776) DNA was digested with *Bcl*I and *Pst*I, and the 900-bp fragment was isolated from an agarose gel as described above. pAdD26SV0d was digested partially with *Pst*I, and the singly cut linear fragment was isolated as described above. The eluted DNA was then digested partially with *Bgl*II, and the 4.5-kb fragment was isolated from an agarose gel. The SV40 fragment was ligated to this vector DNA, using T4 DNA ligase, and the recombinant was used to transform competent HB101. pAdD26SV(A) no. 3 was obtained after screening and contains the SV40 early polyadenylation site and an SV40 origin of replication in a pBR322 plasmid which has a 1.1-kb deletion.

The SV40 origin from pAdD26SV(A) no. 3 has been deleted by digestion of this plasmid with *Eco*RI and *Sal*I and isolation of the 4.4-kb fragment and ligation to pBR322 DNA, which had been digested with *Eco*RI and *Sal*I. The later DNA had been previously treated with calf intestine phosphatase. The ligase-treated DNA was used to transform HB101. pAdD26SV(A) no. 1 is identical to pAdD26SV(A) no. 3 except that it lacks the SV40 origin of replication.

The pBR322 sequences in recombinants pDHFR26, pAdD26-1, pASD11, and pASD12 are derived from the *Eco*RI to *Pst*I site of pBR322. The pBR322 sequences of recombinants pAdD26SV(A), pCVSVE, and pCVSVL are derived from a pBR322 derivative which contains a 1.1-kb deletion of sequences detrimental to replication in mammalian cells (36). However, recombinant pAdD26-1 was tested with both pBR322-derived plasmids (deleted and wild type), and transformation efficiencies were identical. In addition, recombinant pAdD26SV(A) was tested with the presence [pAdD26SV(A) (no. 3)] or absence [pAdD26SV(A) (no. 1)] of an origin of SV40 replication (Fig. 1), and results were identical.

pCVSVE and pCVSVL were cloned by C. Shoemaker (Genetics Institute) by applying *Xho*I linkers to the *Ava*II-*D* fragment of SV40 and inserting it into the *Xho*I site 250 bp upstream from the adenovirus major late promoter (MLP) cap site in pAdD26SV(A) (no. 1). pCVSVE contains the *Ava*II fragment in the orientation such that the early promoter of SV40 is in the same orientation as the adenovirus 2 (Ad2) MLP. pCVSVL contains the *Ava*II fragment in the opposite orientation.

Ligations were carried out with 5 μg of each vector and insert DNA per ml and 5 U of T4 DNA ligase per ml in 30 mM Tris-chloride (pH 7.8)–10 mM $MgCl_2$–5 mM dithiothreitol–1 mM ATP and incubating for 12 h at 14°C. Transformation of competent HB101 (18) was carried out, and recombinant clones were selected by growth on plates containing 15 μg of tetracycline per ml. Recombinant clones were screened and characterized by miniplasmid preparation and restriction endonuclease digestion. The appropriate plasmids were grown in large batches, and plasmid DNA was prepared by banding twice to equilibrium on CsCl gradients.

RNA isolation. Confluent cells were scraped and rinsed with cold phosphate-buffered saline, allowed to swell 5 min in 10 mM NaCl–10 mM Tris (pH 7.4)–3 mM $MgCl_2$ with 20 mM vanadyl adenosine complex, and lysed by adding 0.5% Nonidet P-40 and blending in a Vortex mixer. Nuclei were pelleted by centrifugation for 3 min at 2,500 rpm, and supernatants were subjected to phenol extraction (2×), chloroform extraction (2×), and ethanol precipitation. RNA was pelleted and passed twice over an oligodeoxythymidylate-cellulose column. The bound material {cytoplasmic polyadenylated [poly(A)$^+$]} was eluted for use.

Northern blot analysis. Approximately 3 to 5 μg of poly(A)$^+$ RNA was treated with glyoxal and electrophoresed on 1.4% agarose gels as described by McMaster and Carmichael (41). RNA was directly transferred to nitrocellulose as described by Thomas (55). After transfer, the nitrocellulose was baked in vacuo at 80°C for 2 h and incubated at 42°C overnight in 50% formamide–5× SSC (SSC = 0.15 M NaCl plus 0.015 M sodium citrate)–25 mM sodium phosphate–5 mM EDTA–0.5% sodium dodecyl sulfate–0.05% each

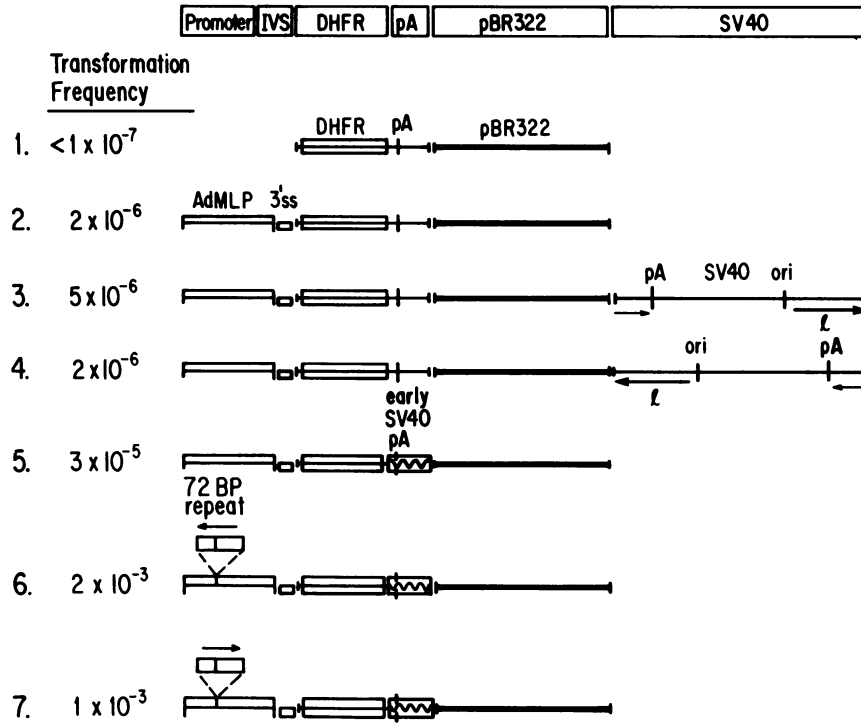

FIG. 1. DHFR⁻ to DHFR⁺ transformation frequency from various cDNA genes. Cells were transfected with plasmid DNA (1 μg of DNA per 10^6 cells) in the absence of carrier DNA. Transformation frequencies were determined from the number of DHFR⁺ transformants arising per number of cells plated into selective media. To ensure internal consistency, the results reported here are from a single experiment in which 3×10^6 cells were plated into selective media. All results have been duplicated in other experiments. Recombinants indicated are: (1) pDHFR26 (15); (2) pAdD26-1 (Kaufman and Sharp, in press) constructed from the DHFR cDNA clone pDHFR26, the Ad2 MLP (62), and a 3' SS isolated from a variable-region immunoglobulin gene (9); (3) pASD11 (Kaufman and Sharp, in press); (4) pASD12 (Kaufman and Sharp, in press); (5) pAdD26SV(A) no. 1 or 3, which contains an SV40 early polyadenylation (pA) site (see text); and (6) pCVSVE and (7) pCVSVL, which contain the 72-bp repeat of SV40 (see text). The arrows on the 72-bp repeat segment indicate the direction of late SV40 transcription, and those on the SV40 DNA segment indicate the late transcription unit.

bovine serum albumin, Ficoll, and polyvinylpyrrolidine (19). The filters were removed and hybridized for 8 to 13 h at 42°C in the same solution with the inclusion of 10% dextran sulfate (58), 10 μg of denatured salmon sperm DNA per ml, and approximately 10^7 cpm of ^{32}P radioactively labeled denatured probe (either pBR322 or pDHFR26 DNA) prepared to a specific activity of $>10^8$ cpm/μg by nick translation (45). After hybridization, filters were washed with 5× SSC with 0.5% sodium dodecyl sulfate at 68°C, 2× SSC with 0.5% sodium dodecyl sulfate at 68°C, and briefly with 2× SSC at room temperature. Filters were prepared for autoradiography, using preflashed (32) Kodak XR-5 film with intensifying screens at −70°C. Filters to be reused for hybridizations were rinsed two times for 30 min at 68°C in 50% formamide and then pretreated for prehybridization.

S1 nuclease mapping. A 3'-specific probe was prepared by digesting pDHFR26 DNA with *Taq*I and then using T4 DNA polymerase to fill in the 3' end with [α-^{32}P]dATP, [α-^{32}P]dCTP, and [α-^{32}P]dTTP (21). The DNA was incubated with dGTP and T4 DNA polymerase in 50 mM Tris (pH 8.0)–7 mM $MgCl_2$–7 mM β-mercaptoethanol for 20 min at 9°C to allow the 3' to 5' exonuclease activity to remove nucleotides up to the first guanosine and then adding α-^{32}P-radioactively labeled dATP, dCTP, and dTTP for 1 h at 9°C to allow the 5' to 3' polymerase activity to add labeled nucleotides. The reaction was phenol extracted and ethanol precipitated, and the pellet was suspended and digested with *Pst*I. The DNA was electrophoresed on a neutral agarose gel, and the band migrating at 950 bp was eluted, found to have a specific activity of 5×10^5 cpm/μg, and used for hybridization.

A 5'-specific probe was prepared by digesting plasmid pAdD26-1 with *Sac*I and then labeling the 5' recessed end with DNA polynucleotide kinase. After digestion, the DNA was treated with calf intestine phosphatase for 30 min at 37°C, phenol and ether extracted, and ethanol precipitated. After the DNA pellet was washed in 70% ethanol, it was suspended in 20 μl of 10 mM glycine–1 mM spermidine–1 mM EDTA, boiled for 2 min, and quick cooled. [γ-^{32}P]dATP was added and incubated with polynucleotide kinase in 0.1 M Tris (pH 8.0)–10 mM $MgCl_2$–5 mM dithiothreitol–0.1 mM spermidine for 2 h at 37°C. The

reaction was then put at 68°C for 20 min. The DNA was digested with *Xho*I, phenol extracted, ethanol precipitated, and electrophoresed on an agarose gel. The band migrating at approximately 800 bp was eluted and found to have a specific activity of $>10^6$ cpm/μg.

RNA-DNA hybridization was performed under R-loop conditions described by Casey and Davidson (14). Labeled DNA (50 ng) was combined with 5 μg of poly(A)$^+$ cytoplasmic RNA in 15 μl of 80% formamide–0.4 M NaCl–0.04 M PIPES [piperazine-*N*,*N*′-bis (2-ethanesulfonic acid)] (pH 6.5)–1 mM EDTA, and the reactions were denatured by incubation for 5 min at 68°C and then were allowed to anneal for 4 h at 50°C (for the 5′-labeled probe) or 55°C (for the 3′-labeled probe). These temperatures are the melting temperatures determined for these fragments under similar conditions. The hybridization was terminated by the addition of 200 μl of cold S1 nuclease buffer (0.25 M NaCl, 0.3 M $NaCH_3COO$ [pH 4.5], 1 mM $ZnSO_4$, 5% glycerol), and S1 nuclease digestion was carried out for 30 min at 45°C as described by Berk and Sharp (8). Samples were phenol extracted, ethanol precipitated with 10 μg of yeast tRNA carrier, and prepared for glyoxal gel electrophoresis (41) on 2% agarose gels. Gels were dried and exposed, using preflashed film and intensifying screens.

Primer extension using reverse transcriptase. A 5′-end-labeled primer was produced by treating pDHFR26 DNA with *Taq*I and calf intestine phosphatase. After phenol extraction and ethanol precipitation, the DNA was labeled with [γ-^{32}P]ATP and polynucleotide kinase. The reaction was heated to 68°C, ethanol precipitated two times, and suspended in buffer for *Pst*I digestion. Products were run on a 5% acrylamide gel (37), and the 100-bp band was eluted and run on a denaturing 8 M urea–8% acrylamide gel. The labeled strand was eluted and hybridized to RNA for reverse transcription.

Approximately 2×10^5 cpm of primer was ethanol precipitated with 3 to 5 μg of poly(A)$^+$ cytoplasmic RNA isolated from various cell lines. After washing the pellet with 70% ethanol, the pellet was dried and rehydrated in 100 μl of 50% deionized formamide–1 M NaCl–50 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid) (pH 6.95)–2.5 mM EDTA, heated for 10 min at 80°C, and transferred to a 45°C water bath for 4 h. The reaction was diluted twofold with water and ethanol precipitated. The pellet was suspended and ethanol precipitated twice again, rinsed in 70% ethanol, and dried. The pellet was rehydrated in 20 μl of water and used for reverse transcription in 100 μl of 20 mM KCl–50 mM Tris (pH 8.3)–10 mM $MgCl_2$–1.25 mM vanadyl adenosine–30 mM β-mercaptoethanol–1 mM each deoxynucleotide triphosphate–40 μg of actinomycin D per ml with avian myeloblastosis virus reverse transcriptase (provided by J. Beard) by incubation at 45°C for 1 h. The reaction was stopped by adding EDTA to 10 mM and NaOH to 0.3 N and heating at 65°C for 15 min. After neutralization with 1 M Tris-acetate (pH 5.8), products were phenol extracted, ether extracted, and ethanol precipitated twice with 5 μg of yeast tRNA carrier. The pellets were rinsed with 70% ethanol, redissolved in sequencing dye, heated to 90°C for 2 min, and applied to a 7 M urea–6% acrylamide gel. Gels were exposed wet, and extended products were isolated by electroelution for DNA sequencing as described by Maxam and Gilbert (40).

R-loop analysis. Approximately 0.1 μg of *Kpn*I-digested pASD11 or pASD12 was ethanol precipitated with approximately 5 μg of cytoplasmic poly(A)$^+$ RNA from either 1C cells or 2B cells grown in 50 μM MTX. The pellets were rinsed in 70% ethanol, dried, suspended in 3 μl of hybridization buffer (80% formamide, 0.4 M NaCl, 0.04 M PIPES [pH 6.5], 1 mM EDTA), and incubated for 3 h at 51°C. The reactions were stopped by diluting the DNA to 0.5 μg/ml in 50% formamide containing 100 mM Tris (pH 8.5)–10 mM EDTA–70 μg of cytochrome *c* per ml, spread on a hypophase of 20% (vol/vol) formamide–10 mM Tris (pH 8.5)–1 mM EDTA, picked up on Parlodion-coated grids, stained with uranyl acetate, and shadowed with platinum. Grids were examined in a Phillips 201 transmission electron microscope at 15,000-fold magnification.

RESULTS

A modular DHFR gene has been constructed from a DHFR cDNA clone by the addition of DNA segments to ensure proper transcription initiation, RNA splicing, and polyadenylation (Fig. 1). A 870-bp segment encoding the Ad2 MLP has been ligated to a 100-bp segment of a 3′ splice site (SS) junction obtained from an variable-region immunoglobulin gene and has been placed upstream from a DHFR cDNA segment (Kaufman and Sharp, in press). Transcription initiation at the Ad2 MLP should produce a 825-base mRNA with the adenovirus late leader spliced to the 3′ SS, and the mRNA should be polyadenylated at sequences in the 3′ end of the cDNA. The first AUG on this mRNA is that utilized for DHFR translation. However, when this recombinant was transfected into DHFR-CHO cells, DHFR$^+$ cells arose at a very low frequency (Fig. 1). Subsequently, the entire genome of SV40 was inserted at the *Eco*RI site, and these recombinants were able to transform DHFR$^-$ to DHFR$^+$ cells at a higher frequency (Fig. 1). All of the transformed clones isolated contained one to five copies of the plasmid DNA integrated into the CHO genome, and upon stepwise MTX selection to a concentration of 1 μM, these clones amplified various subsets of the transforming DNA several hundredfold (Kaufman and Sharp, in press). This amplification facilitated a detailed analysis of the DHFR mRNAs produced.

Northern blot analysis of poly(A)$^+$ RNA from MTX-resistant cells. To determine whether the various DNA segments utilized to construct the modular gene are actually functional, the DHFR mRNA produced in four amplified lines was analyzed in detail. Three clones, 1B, 1C, and 1D, were transformed with a plasmid containing SV40 in one orientation (pASD11) and another clone, 2B, was transformed with a plasmid containing SV40 in the opposite orientation

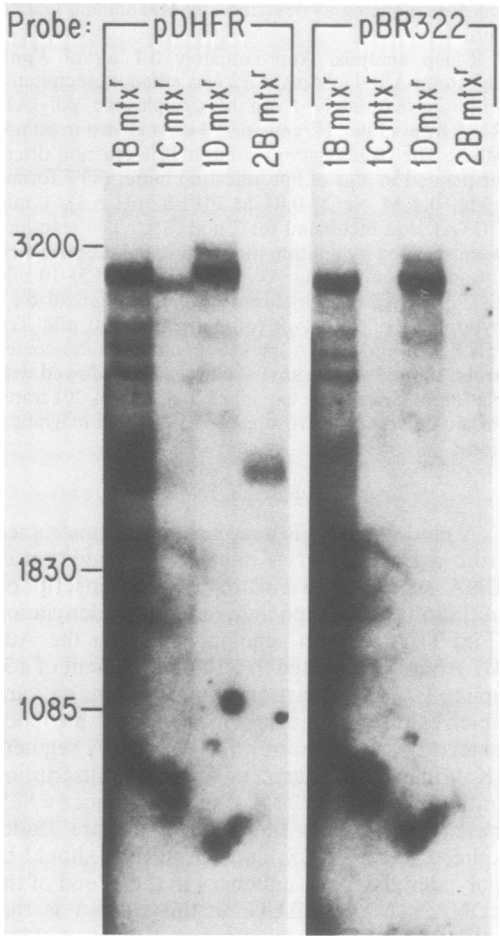

FIG. 2. Northern analysis of DHFR mRNA in amplified cell lines. Approximately 3 μg of cytoplasmic poly(A)$^+$ RNA isolated from four different transformants resistant to 1 μM MTX was electrophoresed on a glyoxal agarose gel, transferred to nitrocellulose, and hybridized to ^{32}P-labeled pDHFR26 (15) prepared by nick translation. Afterwards the [^{32}P]DNA was removed and the filter was rehybridized to ^{32}P-labeled pBR322 DNA (see text). The numbers on the left indicate sizes of *Hin*dIII SV40 DNA molecular weight markers. It should be noted that the probe pDHFR contains radioactive DNA complementary to DHFR and plasmid sequences.

(pASD12) (see Fig. 1). Poly(A)$^+$ RNA isolated from these four transformed cell lines resistant to 1 μM MTX was fractionated by agarose gel electrophoresis under denaturing conditions, transferred to nitrocellulose, and hybridized to ^{32}P-labeled DHFR plasmid DNA and to ^{32}P-labeled pBR322 DNA (Fig. 2).

If all components used to construct the modular gene were utilized, a 825-base mRNA would be expected. However, hybridization to ^{32}P-labeled pDHFR revealed predominately much larger mRNA species homologous to DHFR. Two clones (1B and 1D) show a major species of mRNA approximately 3 kb in length with numerous other species. In contrast, clones 1C and 2B produce a major mRNA species of approximately 3 and 2.2 kb in length, respectively. Hybridization of the same filter to ^{32}P-labeled pBR322 DNA indicated that clones 1B and 1D contain pBR322 sequences in the mRNA, whereas clones 1C and 2B do not. This is consistent with the analysis of the amplified DNA in these lines, where it has been demonstrated that clones 1C and 2B have specifically deleted pBR322 sequences (Kaufman and Sharp, in press).

5′-End analysis of mRNA. S1 nuclease mapping was carried out to determine whether the 3′ SS junction is utilized in the modular gene (8, 59). Poly(A)$^+$ mRNA from the four MTX-resistant clones described above was hybridized to a DNA fragment (uniquely labeled at one 5′ end, the *Sac*I site in the cDNA) under conditions favoring RNA-DNA hybridization (14). After S1 nuclease digestion, the products were analyzed by electrophoresis on denaturing agarose gels. If the 3′ SS was utilized, a band should appear at a size 25 bases larger than the DNA probe cut with *Pst*I since the 3′ SS is 25 bases upstream from the restriction endonuclease *Pst*I site bounding the cDNA segment and the 3′ SS segment. Results from all four clones show the same size band approximately 25 bases longer than the probe DNA cut with *Pst*I (Fig. 3). Hybridization to RNA isolated from mouse cells that are MTX resistant due to amplification of endogenous DHFR genes generates a band slightly smaller than the *Pst*I-cut probe. Eighteen extra bases from the homopolymeric guanine-cytosine tract present in the cDNA clone (15) are protected by mRNA transcribed from the modular gene. We conclude that the 3′ SS cloned into the modular gene is efficiently utilized during mRNA synthesis.

The 5′ end or initiation site of the mRNAs produced have been characterized by primer extension, using reverse transcriptase (24). A 108-bp *Taq-Pst*I fragment was uniquely labeled at one 5′ end (a *Taq*I site 10 bases 3′ of the AUG of DHFR), and the labeled strand was isolated and used as a primer for reverse transcription of poly(A)$^+$ RNA from three MTX-resistant clones (1C, 1D, and 2B). Extended products were analyzed by electrophoresis in urea-polyacrylamide gels (Fig. 4). All three clones produce identical extended products with a major product approximately 60 bases longer than the primer (arrow, Fig. 4). This is the size expected from an mRNA species derived from initiation of transcription at Ad2 MLP and excision of the hybrid intron by splicing. The band at 200 bases (asterisk) repre-

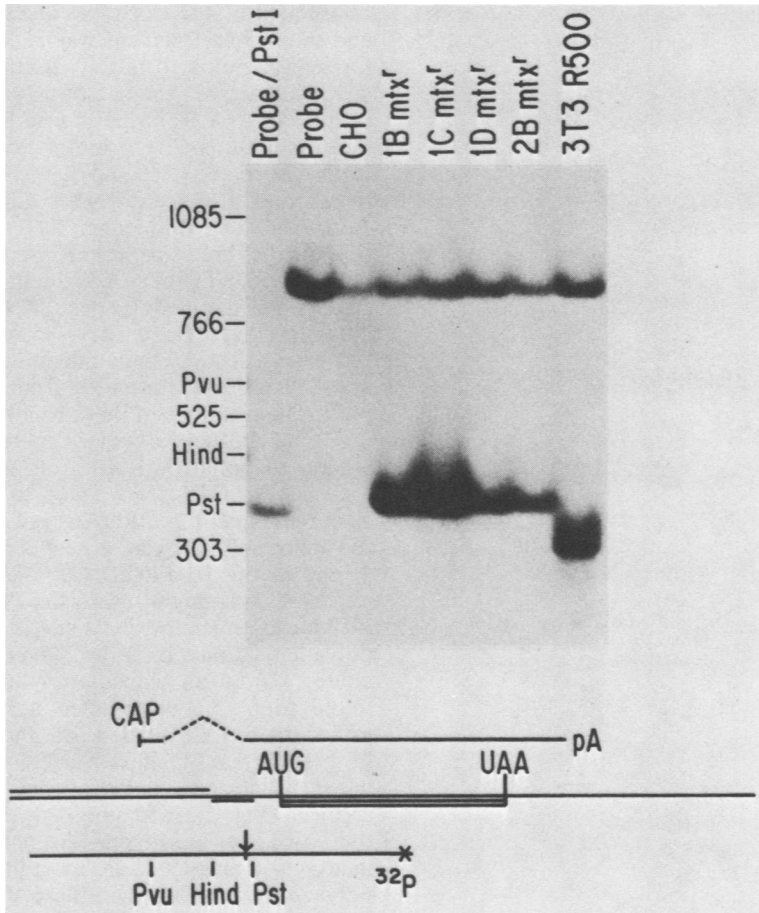

FIG. 3. S1 nuclease analysis of DHFR mRNA with 5'-labeled probe. DNA uniquely labeled at the 5' end of a *Sac*I site present in the DHFR coding region was prepared and hybridized to approximately 3 μg of cytoplasmic poly(A)$^+$ RNA isolated from the original DHFR-deficient CHO line, four CHO transformants resistant to 1 μM MTX, and a highly MTX-resistant murine 3T3 cell line with amplified endogenous DHFR genes. After S1 nuclease digestion the products were electrophoresed on a denaturing glyoxal agarose gel (see text). The lengths indicated on the left are molecular weight markers of *Hin*fI fragments of SV40 DNA. In addition are shown the points of migration of the *Sac*I end-labeled probe after digestion with *Pvu*II (Pvu), which cuts in the adenovirus leader sequence, with *Hin*dIII (Hind), which cuts at the junction of the adenovirus and immunoglobulin segment, and with *Pst*I (Pst), which cuts at the junction of the immunoglobulin segment and the homopolymeric guanine-cytosine-tailed DHFR cDNA. The diagram below depicts the 5'-end-labeled probe with the DHFR modular gene and the expected structure of DHFR mRNA formed by initiation of the Ad2 MLP, processing of the hybrid intron, and polyadenylation at sequences in the 3' end of the cDNA clone. The 3' SS is shown by an inverted arrow.

sents renaturation of the probe. The other bands have not been well characterized in detail; the band at 390 bases is generated by splicing of an SV40 late leader to the 3' SS. The 5' SS of this leader was identified from the sequence of the 390-base band and corresponds to the 5' SS sequence mapping at position 445 in the late leaders (33). The amount of this leader varies between the different cell lines and probably reflects transcription and processing from flanking integrated SV40 segments. The other bands are probably the product of incomplete extension by reverse transcriptase.

To characterize the 5' end in more detail, the 175-base extended product from 1D cell RNA was subject to DNA sequencing (40; Fig. 5). The sequence demonstrates correct initiation at the Ad2 MLP, and the hybrid intron is removed precisely as expected. The 3' end of the Ad2 first leader of the tripartite leader is correctly spliced to the 3' SS of the variable-region immunoglobulin gene.

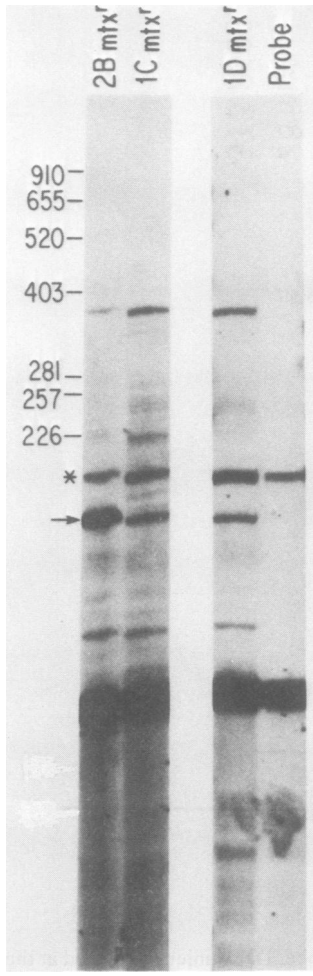

FIG. 4. Primer extension on DHFR mRNAs. A 108-base primer prepared by 5′ end labeling a *Taq*I site in DHFR sequences and digestion with *Pst*I was used to prime reverse transcription of cytoplasmic poly(A)$^+$ RNA isolated from three cell lines, 2B, 1C, and 1D, all resistant to 50 μM MTX. Products were run on a 7 M urea–6% acrylamide gel (see text). The molecular weight markers at the left are Alu-digested pBR322. The asterisk represents a renaturation product of the primer, and the arrow indicates the specific extended products from the Ad2 MLP.

3′-End analysis of DHFR mRNAs. Since different clones contain large different-sized DHFR mRNAs and all DHFR mRNAs in the different clones are identical at the 5′ end, the 3′ ends of the mRNAs were analyzed to determine the origin of the extra sequences. RNA-DNA hybridization and S1 nuclease digestion was carried out with the 3′-end-labeled DNA fragment shown in Fig. 6. The DNA fragment was labeled with T4 DNA polymerase by flushing the 3′ end at a *Taq*I site in the DHFR coding region. Subsequently, the DNA was digested with *Pst*I, and the 945-bp fragment was isolated and used for hybridization to mRNA from the cell lines described above. The site of polyadenylation of cellular mRNA from the mouse DHFR gene is heterogeneous. Two major sites generate mRNA species at 750 and 1,600 bases, and two minor sites produce mRNA species at 950 and 1,200 bases (47, 48). Sequences for polyadenylation for the three smaller mRNAs are contained within the DHFR cDNA segment used for construction of the modular gene. RNA polyadenylation at these three sites would produce S1 nuclease-resistant bands migrating as 450, 650, and 880 bases (determined from Setzer et al. [48]). The positions of these bands are indicated in Fig. 6. [In this particular experiment, hybridization to poly(A)$^+$ RNA from MTX-resistant mouse cells [3T3-R500] failed to generate distinct bands.] If the mRNA is polyadenylated at the major polyadenylation sequence (pA1) in the 3′ end of the DHFR cDNA, a band would be expected to comigrate with the *Bgl*II-cut probe (450 bases) since the *Bgl*II site is located at the polyadenylation site in the 3′ end of the cDNA clone (47). In all clones very little RNA was found to be polyadenylated at this site, but rather the RNA extended past the poly(A) site. Clones 1B and 1D show very similar patterns, which is consistent with their Northern blot analysis (Fig. 2), indicating that the mRNAs in these two lines are similar. The bands at approximately 650 and 900 bases most likely represent polyadenylation at the 950 (pA2)- and 1,200 (pA3)-base polyadenylation signals in the DHFR cDNA since these bands comigrate with protected bands generated by hybridization of the same probe to 3T3-R500 RNA (obvious from other experiments). A major portion of the protected probe is full length, detecting the presence of RNA continuous along the entire 3′ end of the cDNA. This RNA eventually enters pBR322 sequences. Such RNA species were detected in the Northern analysis of RNA from clones 1B and 1D (Fig. 2). Clone 1C shows two major protected fragments. The smaller species is representative of polyadenylation at the 950-base (pA2) site for cellular DHFR mRNA. In contrast, clone 2B has very little mRNA polyadenylation at DHFR sites but rather contains a single major protected species at 800 bases. The results from S1 mapping from 1C and 2B indicate a discontinuity between the mRNA transcript and the DNA at different sites past the major polyadenylation site in the 3′ end of the cDNA clone. Since the size of the mRNA determined by 5′ and 3′ S1 nuclease mapping are much less than that expected from the Northern blot hybridization experiments, the 3′ discontinuities in clones 1C and 2B must result from either RNA splicing

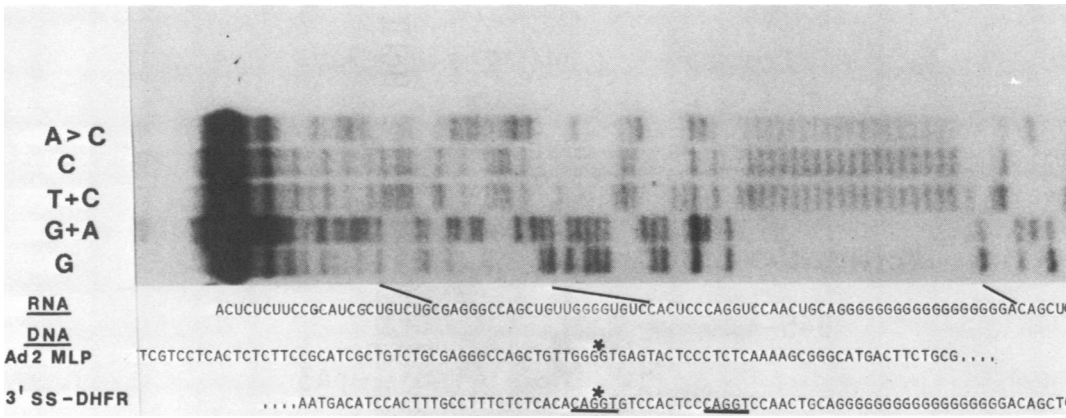

FIG. 5. Sequence of extended product from mRNA from 1D cell line. The 1D extended product (arrow) in Fig. 3 was eluted and subjected to Maxam and Gilbert (40) sequencing. The sequence of the extended product is shown with the sequence of the Ad2 MLP and the sequence of the 3' SS junction adjacent to the guanine-cytosine homopolymer tract of the DHFR cDNA. Nucleotides which match a consensus 3' SS junction (49) are indicated by underlining, and the 5' and 3' SS junctions utilized are indicated by asterisks. Although it is not possible to read the sequence up to the final 5' nucleotide, it is possible to estimate the most 5' nucleotide on the basis of the regular spacing between the bands.

to sequences downstream from the DHFR cDNA or rearrangements of DNA sequences of the 3' end of the modular DHFR gene in transformed cells.

To determine what sequences are represented in the 3' end of the mRNAs produced in the different clones, R-loop hybridization and electron microscopic examination of the products were performed (54). When pASD11 (the plasmid used to transform clones 1B, 1C, and 1D) was hybridized to 1C poly(A)$^+$ RNA, a small loop (sometimes attached to a collapsed RNA bush) was observed (Fig. 7). The size and location of this loop are consistent with RNA hybridization from the 3' SS of the immunoglobulin sequence to ca. 200 bases past the DHFR coding region (determined from measuring 20 molecules). All R-loop molecules observed had the same pattern. This suggests that the 3' portion of the mRNA in clone 1C is transcribed from nonplasmid DNA, probably from CHO DNA sequences joined to DHFR sequences.

In contrast, when 2B poly(A)$^+$ RNA was hybridized to pASD12 (the plasmid used to transform 2B cells), two RNA hybridization loops were observed. One loop corresponded to hybridization to the DHFR coding region. The other loop resulted from hybridization to sequences on the plasmid which were derived from SV40 (determined from measuring 15 molecules). The particular region of SV40 expressed as RNA corresponded to sequences in the late transcription unit of SV40 and extended to the late SV40 polyadenylation sequence. Since the SV40 RNA sequences were contiguous with the DHFR DNA sequences, the integrated DNA sequences in clone 2B probably have a rearrangement of plasmid DNA joining the late region of SV40 to the 3' end of the DHFR segments (see Fig. 6).

These results indicate that sequences at the 3' end of the DHFR mRNA expressed in clone 1C come from approximately 2 kb of sequences not present on the original plasmid (Fig. 7 and 8). The sequences at the 3' end of the mRNA in 2B come from sequences in the late region of SV40. The major mRNA from 1B and 1D contain pBR322 sequences in the 3' untranslated portion of the mRNA. Although the discontinuity at the 3' end of some mRNA species in clones 1B and 1D comes from polyadenylation at the 950 (pA2)- and 1,200 (pA3)-base sites in the DHFR cDNA, the major DHFR mRNAs in these clones are transcribed through pBR322 and into the late SV40 polyadenylation sequence. In fact, S1 nuclease mapping has directly shown that mRNA is terminated at the late SV40 poly(A) sequence in clones 1D and 2B but not in clone 1C (not shown).

DNA analysis of modular DHFR genes in transformed cells. The results from analysis of the mRNAs produced predict certain structures for the DNA in the transformed cells. It has previously been shown that clones 1B and 1D contain intact pBR322 sequences in the transformed cells. In contrast, clone 1C has deleted pBR322 sequences and clone 2B has deleted pBR322 sequences and undergone extensive rearrangement in the process of amplification (Kaufman and Sharp, in press).

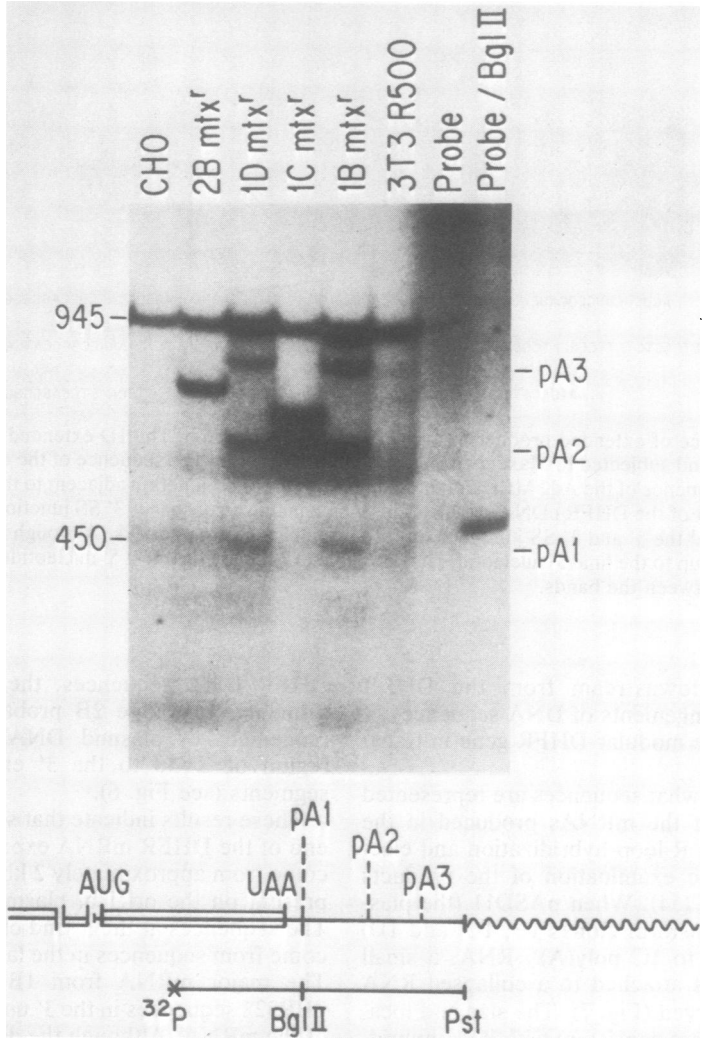

FIG. 6. 3′ S1 nuclease analysis of DHFR mRNA. A probe uniquely labeled in its 3′ end at a *Taq*I site in the DHFR coding region was prepared and hybridized to approximately 3 μg of cytoplasmic poly(A)$^+$ RNA isolated from 1B, 1C, 1D, and 2B cells resistant to 1 μM MTX and to original DHFR-deficient CHO cells and murine MTX-resistant 3T3 cells. After S1 nuclease digestion the products were analyzed by glyoxal gel electrophoresis (see text). The molecular weights indicate the 945-bp initial probe and the probe digested with *Bgl*II which cuts at sequences coding for the major polyadenylation site in cellular DHFR mRNA. The diagram below depicts the 3′ probe utilized and the region of the DHFR modular gene containing the DHFR coding region with the polyadenylation signals for the 750-, 950-, and 1,200-base mRNAs for cellular DHFR. These mRNAs should generate S1-resistant bands at 450, 650, and 880 bases, respectively. The wavy line depicts pBR322 sequence.

More defined Southern blot mapping is shown in Fig. 9, using restriction enzymes that cut within and around the DHFR coding region. *Sac*I cleaves the original plasmid DNA 400 bases upstream from the adenovirus MLP and within the DHFR coding region to generate a 1,050-bp fragment. Clones 1B, 1C, and 1D all have this *Sac*I fragment. *Bgl*II cleaves in the immunoglobulin 3′ SS sequence and at the poly(A) site in the 3′ end of the cDNA to generate a 800-bp fragment. All clones contain this restriction fragment. *Pst*I cleaves at each end of the cDNA clone to generate a 1,375-bp fragment. Only clones 1B and 1D contain this fragment. In addition, clones 1B and 1D contain an intact 3.8-kb *Pst*I band that contains all of pBR322 and at least 200 bp of SV40. *Taq*I cleaves in the 3′ end of the DHFR coding region and in pBR322 to generate a 2.1-kb fragment. Again, only clones 1B and 1D contain this

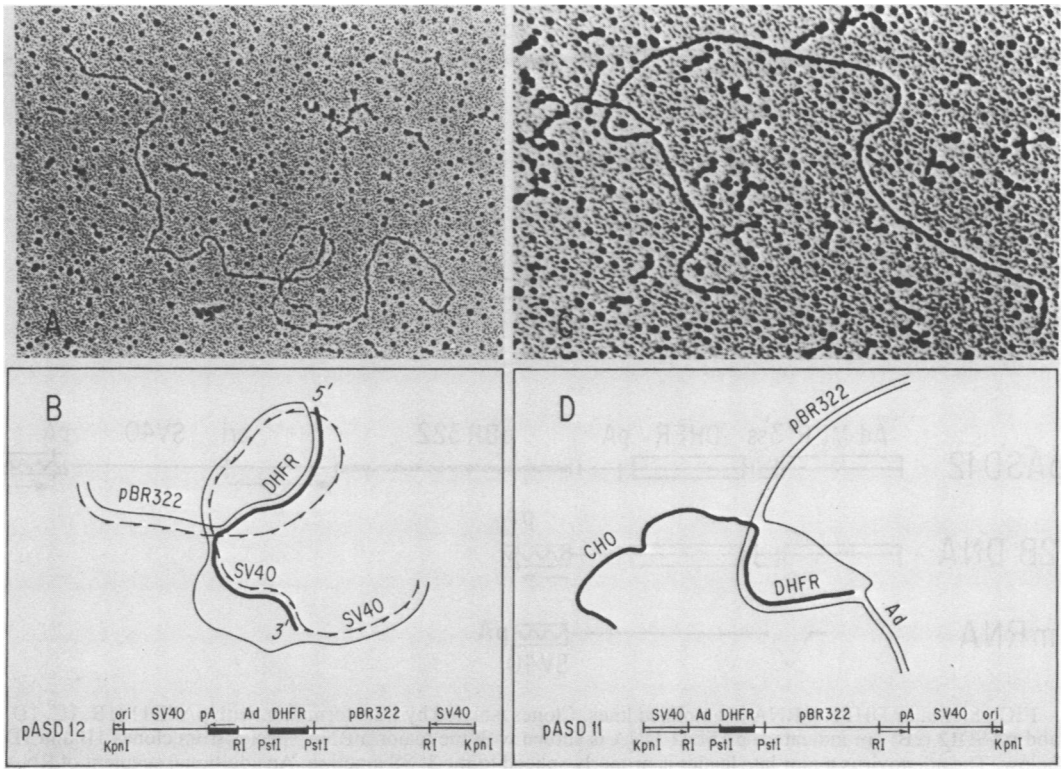

FIG. 7. Electron micrographs of RNA-DNA hybrids. RNA-DNA hybrids were prepared under R-loop hybridization conditions as described in the text. (A) Result after hybridization of 2B mRNA to clone pASD12 DNA digested with *Kpn*I (11.1 kb). (C) Result after hybridization of 1C mRNA to clone pASD11 DNA digested with *Kpn*I. (B and D) Representation of the duplex DNA (dashed and thin lines represent different DNA strands) and RNA hybrid (RNA is represented by the thick line). The 5′ and 3′ halves of the mRNA are also depicted. Below are schematics showing the *Kpn*I-digested DNA (pASD11 or pASD12) with indicated origins and restriction sites. Ori represents the SV40 origin of replication, and pA represents the late SV40 polyadenylation signal. The arrow indicates the late SV40 transcription unit. The regions showing homology to the mRNA are depicted as thick lines. The duplex DNA arms of each hybrid had average lengths of 2.80 ± 0.24 and 5.01 ± 0.39 kb for 2B mRNA and 2.38 ± 0.24 and 8.15 ± 0.26 kb for 1C mRNA. These lengths confirm the S1 nuclease and Northern blot mapping of the RNA sequences on the plasmid DNA.

fragment. The results of these experiments (summarized at the bottom of Fig. 9) indicate that clones 1B and 1D have conserved the modular gene sequences 5′ of the Ad2 MLP all the way through pBR322 past the SV40 late poly(A) site. Clone 1C has conserved the modular gene sequences 5′ of the Ad2 MLP past the second poly(A) site in the cDNA but beyond this point probably diverges into cellular CHO DNA. In clone 2B, the modular gene sequences have recombined close to the 5′ end of the Ad2 MLP and have also undergone rearrangement near the 3′ end of the DHFR cDNA segment. From the R-loop data of Fig. 7, this latter recombination must have involved inversion of sequences from the late region of SV40, including the late SV40 poly(A) site, and their insertion downstream from the DHFR cDNA segment.

Necessity of functional polyadenylation signals to DHFR expression. Analysis of the mRNA has indicated that all clones utilize the Ad2 MLP and the hybrid intron is removed correctly, but the mRNA is not efficiently polyadenylated at sequences in the 3′ end of the cDNA. When SV40 is inserted, RNA transcription proceeds into SV40 and polyadenylation occurs at the late SV40 polyadenylation sequence. This suggests that one function SV40 DNA sequences provide for enhancing the efficiency of transformation of $DHFR^-$ cells to the $DHFR^+$ phenotype is polyadenylation. The original construct may have reduced efficiency of transformation due to inadequate polyadenylation signals. This has been tested directly by cloning the SV40 early polyadenylation site (*Bcl*I [0.19 map unit] to *Pst*I [0.04 map unit]) into the *Bgl*II and *Pst*I sites of the

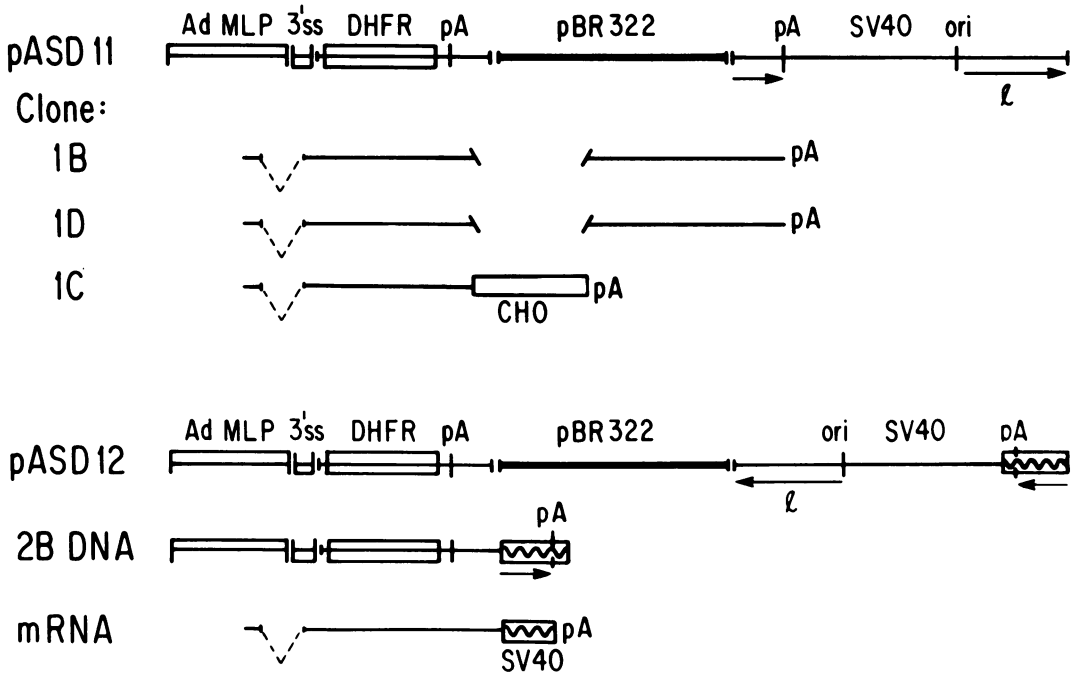

FIG. 8. Major DHFR mRNA in amplified lines. Clones isolated by transformation with pASD11 (1B, 1C, 1D) and pASD12 (2B) are indicated. pASD11 DNA is shown with the major mRNA species from clones 1B and 1D below. The adenovirus major late leader is properly spliced to the 3' SS junction. An additional segment of RNA (approximately 1 kb) is removed, presumably by splicing of sequences in pBR322. The mRNA is polyadenylated at the SV40 late polyadenylation signal. Clone 1C is identical to 1B and 1D at the 5' end, but its 3' end is derived from cellular DNA sequences that are joined to plasmid sequences (open box). pASD12 is depicted with the DNA in clone 2B, indicating a rearrangement of DNA sequences so that the late transcription unit of SV40 is positioned 3' to the DHFR segment. The mRNA from 2B cells does not contain pBR322 sequences but is polyadenylated at the SV40 late polyadenylation signal.

original construct. This recombinant transforms DHFR$^-$ cells to the DHFR$^+$ phenotype with an efficiency 10-fold greater than recombinants containing the entire SV40 genomes (Fig. 1).

Enhancing effect of 72-bp repeat from SV40. The role of the 72-bp repeat from SV40 for enhancing expression from the Ad2 MLP in the modular DHFR cDNA gene has been examined by the insertion of the *Ava*II-D fragment of SV40, which contains the 72-bp repeat, into an *Xho*I site 250 bp upstream from the Ad2 MLP cap site. With its insertion in either orientation, the transformation frequency has increased 50- to 100-fold to 10^{-3} of the transfected cells (Fig. 1). Although the 5' end of the mRNA has not been mapped in these transformants, a direct role of the 72-bp repeat is suggested since it increases the transformation efficiency by its insertion in either orientation.

DISCUSSION

A CHO cell line deficient in DHFR (DHFR$^-$; 16) has been utilized to test the biological activity of recombinants constructed with a mouse DHFR cDNA clone. When a plasmid containing the entire coding region for DHFR and a polyadenylation signal in its 3' end was transfected into CHO DHFR$^-$ cells, no DHFR$^+$ transformants were observed. Subsequently, this DHFR cDNA clone has been converted into an active gene by the addition of proper transcription initiation, RNA processing, and polyadenylation signals. A modular recombinant containing the DHFR cDNA segment was able to transform DHFR$^-$ to DHFR$^+$ cells at a very low frequency. This modular transcription unit contains, in a 5'-to-3' polarity, the Ad2 MLP (encoding the first leader and 5' SS of the adenovirus tripartite mRNA leader sequence) adjoined to a 3' SS sequence isolated from an immunoglobulin variable-region gene and the DHFR coding region. However, DHFR$^+$ cells arose at a higher frequency when the entire genome of SV40 was cloned into this recombinant. Four cloned transformants were isolated and subjected to stepwise increments of MTX in the culture medium and eventually generated four lines which had amplified the transforming DNA over several

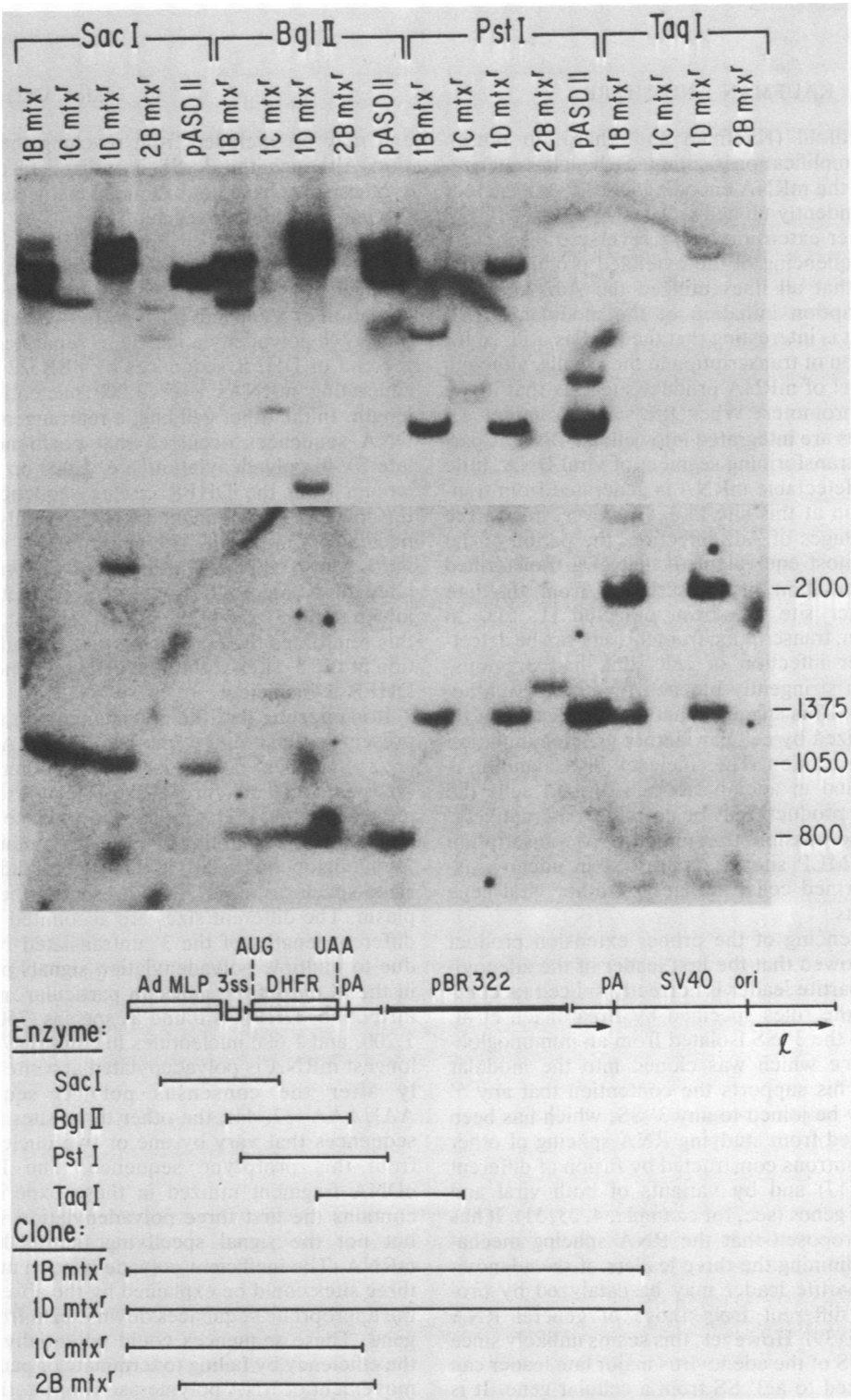

FIG. 9. Southern blot analysis of transforming DNA. A 10-μg portion of genomic DNA isolated from clones 1B, 1C, 1D, and 2B resistant to 1 μM MTX was isolated, treated with the restriction endonucleases indicated, and utilized for agarose gel electrophoresis and Southern blot transfer and hybridization as according to Southern (52), with modifications (Kaufman and Sharp, in press). The hybridization probe was prepared by nick translation of pDHFR26. pASD11 represents a reconstruction of CHO DNA with various amounts of pASD11 DNA. The upper and lower halves of this autoradiogram are shown as different exposure times. Below is a representation of the *Sac*I, *Bgl*II, *Pst*I, and *Taq*I restriction sites around the DHFR coding region in pASD11 which generate fragments approximately 1,050, 800, 1,375, and 2,100 bp long, respectively. The lengths of the amplified transforming DNA in the different clones which are colinear with the plasmid DNA around the DHFR coding region are depicted at the bottom. For simplified representation, the pBR322 and SV40 segments are depicted at 1/2 scale compared with the other DNA segments.

hundredfold (Kaufman and Sharp, in press). This amplification facilitated the characterization of the mRNA encoding DHFR in these four independently cloned isolates.

Primer extension, using reverse transcriptase and sequencing of the extended products, indicated that all lines utilized the Ad2 MLP for transcription initiation of the modular DHFR gene. It is interesting that the MLP is utilized for initiation of transcription in these cells, although the level of mRNA product suggests that it is a weak promoter. When the same promoter sequences are integrated into cellular DNA as part of the transforming segment of viral DNA, little or no detectable mRNA is generated from transcription at this site (50). However, during the early stages of Ad2 infection, the period of the cycle most equivalent to that of a transformed cell, initiation of transcription from the late promoter site has been detected (1, 51). In addition, transcription from this site can be detected after infection of cells that had previously been stringently blocked for protein synthesis (35). This suggests that the MLP site can be recognized by cellular factors yielding initiation of transcription. The efficiency of recognition is poor, and in adenovirus-transformed cells the mRNA products may be unstable. Alternatively, it is also possible that initiation of transcription at the MLP site is suppressed in adenovirus-transformed cells, perhaps by other viral gene products.

Sequencing of the primer extension product also showed that the first leader of the adenovirus tripartite leader is correctly spliced (according to the rules specified by Breathnach et al. [10]) to the 3′ SS isolated from an immunoglobulin gene which was cloned into the modular gene. This supports the contention that any 5′ SS may be joined to any 3′ SS, which has been suggested from studying RNA splicing of other hybrid introns constructed by fusion of different genes (17) and by variants of both viral and cellular genes (see, for example, 4, 23, 31). It has been proposed that the RNA splicing mechanism adjoining the three leaders of the adenovirus tripartite leader may be catalyzed by processes different from those of general RNA splicing (39). However, this seems unlikely since the 5′ SS of the adenovirus major late leader can be spliced to a 3′ SS from a cellular gene. It is also of interest that the inserted immunoglobulin segment has two possible 3′ SS's that are similar to the consensus 3′ SS sequence (see Fig. 5). However, the only 3′ SS joined to the Ad2 leader is that utilized in the formation of immunoglobulin mRNA. Thus, there must be some differences in these two potential 3′ SS's (49) or, as proposed by Chu and Sharp (17), RNA splicing may occur in a processive manner and may be excluded from reaching the other site. Although the 3′ SS is utilized in mRNA synthesis, we have not examined the necessity of the immunoglobulin segment.

The major mRNA encoding DHFR in each of three independently isolated cell lines was polyadenylated at sequences that specify the polyadenylation of SV40 late mRNA. In two cell lines, this SV40 polyadenylation site is separated from the end of DHFR sequences by pBR322 DNA, generating mRNAs over 3,000 nucleotides in length. In the other cell line, a rearrangement of DNA sequences occurred that positioned the late SV40 polyadenylation site, 1,000 bp downstream from the DHFR coding sequences. In this line, the predominant mRNA was only 2,200 nucleotides in length. The major DHFR mRNA in the fourth cell line was apparently polyadenylated at a site in CHO sequences that were joined to the 3′ end of the cDNA segment. All of this reinforces the conclusion that polyadenylation at the 3′ sites within the cDNA segment for DHFR is inefficient.

It is puzzling that the polyadenylation signals present in the 3′ end of the DHFR cDNA clone are so inefficient. DHFR is similar to other genes (yeast alcohol dehydrogenase [6], mouse, α_{2u}-microglobulin [57], mouse α-amylase [56], chicken ovalbumin-like X gene [29], mouse μ-immunoglobulin [2, 20]) in that several different-sized species of mRNA are present in the cytoplasm. The different sizes are accounted for by different lengths of the 3′ untranslated regions due to multiple polyadenylation signals present in the 3′ end of the gene. In particular, murine mRNA for DHFR is found as species (750, 950, 1,200, and 1,600 nucleotides in length; 47). The longest mRNA is polyadenylated at a site shortly after the consensus poly(A) sequence AAUAAA (22, 44); the other three sites follow sequences that vary by one or two nucleotides from this prototype sequence. The DHFR cDNA fragment utilized in these experiments contains the first three polyadenylation signals but not the signal specifying the 1,600-base mRNA. The inefficient polyadenylation at these three sites could be explained by the absence of the appropriate sequences downstream from the gene. These sequences could potentially affect the efficiency by failing to terminate or pause the movement of RNA polymerase II or altering the secondary structure of the nascent transcript. An alternative is that this process is under physiological control in cells and the cellular state of the MTX-selected CHO cells is unfavorable for polyadenylation at these sites.

It is likely that insertion of a more efficient polyadenylation site 3′ to the DHFR cDNA segment is required for transformation of the CHO DHFR$^-$ cells with the modular gene. As a

direct test of this possibility the effect of insertion of a segment encoding the early SV40 polyadenylation site immediately 3′ to the DHFR cDNA segment was investigated. This construct enhanced the efficiency of transformation by the modular gene by 10-fold. If the enhanced transformation frequency reflects an increased level of DHFR mRNA synthesis, it would suggest that a large fraction of the primary transcripts from the original DHFR cDNA recombinant failed to be processed to mature mRNAs. A direct experiment to establish the inefficiency of processing of the primary transcript would be difficult with the available cell lines. It should be possible to directly study the efficiency of polyadenylation and subsequent mRNA production by cloning of the DHFR segment into an autonomously replicating vector.

The biologically active DNA segment containing the modular gene and the early SV40 poly(A) site did not contain the SV40 enhancer 72-bp repeat segment. Others have shown that insertion of this DNA segment enhances expression from many promoter sites (5, 7, 34; M. Fromm and P. Berg, submitted for publication). To test whether a similar enhancement could be observed with the Ad2 MLP, the 72-bp repeat was inserted into plasmids containing the modular gene. These plasmids transformed CHO cells with a 50- to 100-fold greater efficiency than the equivalent construct without the 72-bp repeat. One cell in a thousand was transformed after transfection. Although the 5′ end of the mRNA has not been characterized in these transformants and the inserted sequence contains both SV40 early and late transcriptional promoters, the increased efficiency is probably due to an enhancer effect since it is observed with the enhancer sequence in either orientation. This suggests that the Ad2 MLP lacks such an enhancer sequence and the SV40 enhancer can augment expression from the Ad2 MLP.

The availability of a small modular DHFR gene offers a unique tool for studying sequences necessary for gene expression and for regulation of gene expression. The major advantage of the use of the system to examine gene expression is that minimal expression is required for transformation to the DHFR$^+$ phenotype. Thus, this assay is exquisitely sensitive to gene expression. Any alterations severely affecting DHFR expression will result in decreased transformation efficiencies. In addition, quantitation of DHFR expression from different plasmids can be monitored by the ability of transformed cells to grow in various concentrations of MTX, since MTX resistance is roughly linear with DHFR expression. Thus, segments in the modular gene can be mutagenized or exchanged for other segments, and alterations can be conveniently assayed by complementation of the CHO DHFR-deficient cells. A similar approach has been taken to study hormonal induction of gene expression (34). Certainly, the ability to amplify the recombinant DHFR modular genes introduced into cells facilitates a detailed analysis of events involved in their expression and regulation. Also, a fluorescent derivative of MTX may be used in conjunction with the fluorescence-activated cell sorter to identify and isolate variants of DHFR expression.

ACKNOWLEDGMENTS

We are grateful to Ihor Lemischka for helpful discussions and for sequencing the 5′ extended fragment, to Chris Riser for assistance in electron microscopy, to Chuck Shoemaker for providing plasmids pCVSVE and pCVSVL, and to Margarita Siafaca for assistance in preparation of this manuscript.

R.J.K. is a Helen Hay Whitney postdoctoral fellow. This work was supported by National Science Foundation grant PCM78-23230, Public Health Service program project grant CA26717 to P.A.S. from the National Cancer Institute, and partially by the National Institutes of Health Center for Cancer Biology at MIT (grant CA14051).

LITERATURE CITED

1. **Akusjarvi, G., and H. Persson.** 1980. Controls of RNA splicing and termination in the major late adenovirus transcription unit. Nature **292**:420–426.
2. **Alt, F. W., V. Enea, A. L. M. Bothwell, and D. Baltimore.** 1980. Activity of multiple light chain genes in murine myeloma cells producing a single, functional light chain. Cell **21**:1–12.
3. **Alt, F. W., R. E. Kellems, J. R. Bertino, and R. T. Schimke.** 1978. Selective amplification of dihydrofolate reductase genes in methotrexate-resistant variants of cultured murine cells. J. Biol. Chem. **253**:1357–1370.
4. **Baird, M., C. Driscoll, H. Schreiner, G. V. Sciarratta, B. Sansone, G. Niazi, F. Ramirez, and A. Bank.** 1981. A nucleotide change at a splice junction in the human β-globin gene is associated with β°-thalassemia. Proc. Natl. Acad. Sci. U.S.A. **78**:4218–4221.
5. **Banerji, J., S. Rusconi, and W. Schaffner.** 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell **27**:299–308.
6. **Bennetzen, J. L., and B. D. Hall.** 1981. The primary structure of the *Saccharomyces cerevisiae* gene for alcohol dehydrogenase. J. Biol. Chem. **257**:3018–3025.
7. **Benoist, C., and P. Chambon.** 1981. *In vivo* sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–310.
8. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**:721–732.
9. **Bothwell, A. L. M., M. Paskind, M. Reth, T. Imanishi-Kari, K. Rajewsky, and D. Baltimore.** 1981. Heavy chain variable region contribution to the NP^b family of antibodies: somatic mutation evident in a γ2a variable region. Cell **24**:625–637.
10. **Breathnach, R., C. Benoist, K. O'Hare, F. Gannon, and P. Chambon.** 1978. Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc. Natl. Acad. Sci. U.S.A. **75**:4853–4857.
11. **Brinster, R. L., H. Y. Chen, M. Trumbauer, A. W. Senear, R. Warren, and R. D. Palmiter.** 1981. Somatic expression of herpes thymidine kinase in mice following infection of a fusion gene into eggs. Cell **27**:223–231.
12. **Brown, P. C., S. M. Beverly, and R. T. Schimke.** 1981.

Relationship of amplified dihydrofolate reductase genes to double minute chromosomes in unstably resistant mouse fibroblast cell lines. Mol. Cell. Biol. 1:1077–1083.
13. **Capecchi, M. R.** 1980. High efficiency transformation by direct microinjection of DNA into mammalian cells. Cell 22:479–488.
14. **Casey, J., and N. Davidson.** 1977. Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. Nucleic Acids Res. 4:1539–1552.
15. **Chang, A. C. Y., J. H. Nunberg, R. J. Kaufman, H. A. Erlich, R. T. Schimke, and S. N. Cohen.** 1978. Phenotypic expression in E. coli of a DNA sequence coding for mouse dihydrofolate reductase. Nature (London) 275:617–624.
16. **Chasin, L. A., and G. Urlaub.** 1980. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. Proc. Natl. Acad. Sci. U.S.A. 77:4216–4220.
17. **Chu, G., and P. A. Sharp.** 1981. A gene chimaera of SV40 and mouse α-globin is transcribed and properly spliced. Nature (London) 289:378–382.
18. **Dagert, M., and S. D. Ehrlich.** 1979. Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells. Gene 6:23–28.
19. **Denhardt, D. T.** 1966. A membrane-filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Commun. 23:641–652.
20. **Early, P., J. Rogers, M. Davis, K. Calame, M. Bond, R. Wall, and L. Hood.** 1980. Two mRNAs with different 3′ ends encode membrane-bound and secreted forms of immunoglobulin μ chains. Cell 20:313–319.
21. **Englund, P. T.** 1971. Analysis of nucleotide sequences at 3′ termini of duplex deoxyribonucleic acid with the use of the T4 deoxyribonucleic acid polymerase. J. Biol. Chem. 246:3269–3276.
22. **Fitzgerald, M., and T. Shenk.** 1981. The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell 24:251–260.
23. **Fukumaki, Y., P. K. Ghosh, E. J. Benz, Jr., V. B. Reddy, P. Lebowitz, B. G. Forget, and S. M. Weissman.** 1981. Abnormally spliced messenger RNA in erythroid cells from patients with β^+ thalassemia and monkey cells expressing a cloned β^+-thalassemic gene. Cell 28:585–593.
24. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1978. Heterogeneity and 5′ terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. 126:813–846.
25. **Gordon, J. W., G. A. Scangos, D. J. Plokkin, J. A. Barbosa, and F. H. Ruddle.** 1980. Genetic transformation of mouse embryos by microinjection of purified DNA. Proc. Natl. Acad. Sci. U.S.A. 77:7380–7384.
26. **Graham, F., and A. van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 52:456–467.
27. **Gruss, P., C.-J. Lai, R. Dhar, and G. Khoury.** 1979. Splicing as a requirement for biogenesis of functional 16S mRNA of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. 76:4317–4321.
28. **Hamer, D. H., and P. Leder.** 1979. Splicing and the formation of stable RNA. Cell 18:1299–1302.
29. **Hellig, R., F. Perrin, F. Gannon, J. L. Mandel, and P. Chambon.** 1980. The ovalbumin gene family: structure of the X gene and evolution of duplicated split genes. Cell 20:625–637.
30. **Kaufman, R. J., J. R. Bertino, and R. T. Schimke.** 1978. Quantitation of dihydrofolate reductase in individual parental and methotrexate-resistant murine cells. J. Biol. Chem. 253:5852–5860.
31. **Khoury, G., J. Alwine, N. Goldman, P. Gruss, and G. Jay.** 1980. New chimeric splice junction in adenovirus type 2-simian virus 40 hybrid viral mRNA. J. Virol. 36:143–151.
32. **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.
33. **Lebowitz, P., and S. M. Weissman.** 1979. Organization and transcription of the SV40 genome. Curr. Top. Microbiol. Immunol. 87:144–151.
34. **Lee, F., R. Mulligan, P. Berg, and G. Ringold.** 1981. Glucocorticoids regulate expression of dihydrofolate reductase cDNA in mouse mammary tumour virus chimaeric plasmids. Nature (London) 294:228–232.
35. **Lewis, J. B., and M. B. Mathews.** 1980. Control of adenovirus early gene expression: a class of immediate early products. Cell 21:303–313.
36. **Lusky, M., and M. Botchan.** 1981. Inhibition of SV40 replication in simian cells by specific pBR322 DNA sequences. Nature (London) 293:79–81.
37. **Maniatis, T., A. Jeffrey, and H. Van de Sande.** 1975. Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis. Biochemistry 14:3787–3794.
38. **Manley, J., A. Fire, A. Cano, P. A. Sharp, and M. Gefter.** 1980. DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. U.S.A. 77:3855–3859.
39. **Mathews, M.** 1980. Binding of adenovirus VA RNA to mRNA: a possible role in splicing? Nature (London) 285:575–577.
40. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. 74:560–564.
41. **McMaster, G. K., and G. G. Carmichael.** 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Natl. Acad. Sci. U.S.A. 74:4835–4838.
42. **Mellon, P., V. Parker, Y. Gluzman, and T. Maniatis.** 1981. Identification of DNA sequences required for transcription of the human α-globin gene in a new SV40-host vector system. Cell 27:279–288.
43. **Mulligan, R. C., B. H. Howard, and P. Berg.** 1979. Synthesis of rabbit beta globin in cultured monkey cells following infection with a SV40 beta globin recombinant genome. Nature (London) 277:108–114.
44. **Proudfoot, N. J., and G. G. Brownlee.** 1976. Sequence at the 3′ end of globin mRNA shows homology with immunoglobulin light chain mRNA. Nature (London) 252:359–362.
45. **Rigby, P. W. J., M. Diekmann, 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.
46. **Schimke, R. T., P. C. Brown, R. J. Kaufman, M. McGrogan, and D. L. Slate.** 1981. Chromosomal and extrachromosomal localization of amplified dihydrofolate reductase genes in cultured mammalian cells. Cold Spring Harbor Symp. Quant. Biol. 45:785–797.
47. **Setzer, D. R., M. McGrogan, J. H. Nunberg, and R. T. Schimke.** 1980. Size heterogeneity in the 3′ end of dihydrofolate reductase messenger RNAs in mouse cells. Cell 22:361–370.
48. **Setzer, D. R., M. McGrogan, and R. T. Schimke.** 1982. Nucleotide sequence surrounding multiple polyadenylation sites in the mouse dihydrofolate reductase gene. J. Biol. Chem. 257:5143–5147.
49. **Sharp, P. A.** 1981. Speculations on RNA splicing. Cell 23:643–646.
50. **Sharp, P. A., P. H. Gallimore, and S. J. Flint.** 1975. Mapping of adenovirus 2 RNA sequences in lytically infected cells and transformed cell lines. Cold Spring Harbor Symp. Quant. Biol. 39:457–474.
51. **Shaw, A., and E. Ziff.** 1980. Transcripts from the adenovirus-2 major late promoter yield a single family of 3′ coterminal mRNAs during early infection and five families at late times. Cell 22:905–916.
52. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503–518.
53. **Szybalski, E. H., and W. Szybalski.** 1962. Genetics of human cell lines. IV. DNA mediated heritable transformation of a biochemical trait. Proc. Natl. Acad. Sci. U.S.A.

48:2028–2034.

54. **Thomas. M., R. L. White, and R. W. Davis.** 1976. Hybridization of RNA to double-stranded DNA: formation R-loops. Proc. Natl. Acad. Sci. U.S.A. 73:2294–2298.
55. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. U.S.A. 77:5201–5205.
56. **Tosi, M., R. A. Young, O. Hagenbuchle, and U. Schibler.** 1981. Multiple polyadenylation sites in a mouse α-amylase gene. Nucleic Acids Res. 9:2313–2323.
57. **Unterman, R. D., K. R. Lynch, H. L. Nakhas, K. P. Dolan, J. W. Hamilton, D. V. Cohn, and P. Feigelson.** 1981. Cloning and sequence of several α_{2u}-globin cDNAs. Proc. Natl. Acad. Sci. U.S.A. 78:3478–3482.
58. **Wahl, G. M., M. Stern, and G. R. Stark.** 1979. Efficient transfer of large DNA fragments from agarose gels to diazobensyloxy methyl paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. U.S.A. 76:3683–3687.
59. **Weaver, R. F., and C. Weissman.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5′ termini of 15S α-globin mRNA have identical map coordinates. Nucleic Acids Res. 7:1175–1193.
60. **Weil, P. A., D. Luse, J. Segall, and R. Roeder.** 1979. Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell 18:469–484.
61. **Wigler, M., S. Silverstein, L. Lee, A. Pellicer, Y. Cheng, and R. Axel.** 1977. Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells. Cell 11:223–232.
62. **Ziff, E., and R. Evans.** 1978. Coincidence of the promoter and capped 5′ terminus of RNA from the adenovirus 2 major late transcription unit. Cell 15:1463–1475.