MOLECULAR AND CELLULAR BIOLOGY, Aug. 1985, p. 1847–1858
0270-7306/85/081847-12$02.00/0

Vol. 5, No. 8

Analysis of the Mouse *dhfr* Promoter Region: Existence of a Divergently Transcribed Gene

GRAY F. CROUSE,[1]† EUGENE J. LEYS,[2]‡ ROBERT N. McEWAN,[1]§ ELIZABETH G. FRAYNE,[2] AND RODNEY E. KELLEMS[2]

Laboratory of Molecular Biology, LBI—Basic Research Program, National Cancer Institute, Frederick Cancer Research Facility, Frederick, Maryland 21701,[1] and Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, Houston, Texas 77030[2]

Received 31 January 1985/Accepted 23 April 1985

The use of murine dihydrofolate reductase (*dhfr*) gene amplification mutants enabled us to identify important structural and functional features of the *dhfr* promoter region. We found another transcription unit, at least 14 kilobases in size, which initiates within 130 base pairs of the major *dhfr* transcript and is transcribed divergently. The 5′ ends of both transcripts were analyzed and found to have multiple initiation sites. The major *dhfr* transcript and the divergent transcript appear to share the same promoter region; the longer transcripts of the *dhfr* gene overlap with the divergent transcripts and use a different promoter region. The divergent transcript appears to code for a protein; an homologous sequence to its first exon is found in the corresponding location near the human *dhfr* gene.

Most cellular gene regulation studies have been concerned with genes that code for proteins produced in large quantities in specific types of terminally differentiated cells. A large body of evidence indicates that the expression of such genes is controlled transcriptionally (10). However, many cellular genes, perhaps most, code for enzymes which are present at low levels in virtually all types of cells. Studies concerning the expression of housekeeping enzymes or proteins such as dihydrofolate reductase (*dhfr*; 17, 20, 21), thymidine kinase (15), tubulin (7, 16), and glyceraldehyde-3-phosphate dehydrogenase (29) indicate that the differential expression of these genes seen as a function of cell growth or seen in different tissues is controlled posttranscriptionally. Thus, the factors controlling the expression of these genes differ significantly from those controlling genes whose transcriptional activity is restricted to a single type of terminally differentiated cell. In this regard, it is important to realize that our current understanding of RNA polymerase II promoters is based almost exclusively on the structural and functional analyses of a limited number of strong viral and cellular promoters (5, 26). This is a very restricted set of highly efficient and transcriptionally regulated promoters that may have certain features not associated with the promoters of cellular genes constitutively expressed at low levels in all cells. To obtain a more balanced view of promoter structure and function, it is necessary to study more promoters that fall within the latter category. For this reason, we focused our attention on the mouse *dhfr* gene, one of the best characterized genes of this category.

The coding sequence of the *dhfr* gene consists of 558 nucleotides, distributed among six exons spanning a distance of 35 kilobases (kb) (9, 27). The gene produces seven mRNAs ranging in size from 0.75 to 5.6 kb, which differ primarily in the length of 3′ untranslated region that they possess (34, 35). The 3′ ends of these mRNAs correspond to the utilization of seven polyadenylation sites distributed over a 5-kb region of genomic DNA at the 3′ end of the *dhfr* structural gene. The use of *dhfr* gene amplification mutants (2, 33) made it possible for us to detect labeled primary transcripts from whole cells and thus determine what regions of the *dhfr* gene are transcribed, even if those transcripts are not stable. It recently was shown that transcription proceeds at a constant level throughout the gene, including the seven polyadenylation sites, and terminates approximately 1 kb beyond the last polyadenylation site near a region of repeated DNA (12).

Although many features of *dhfr* gene structure and expression are well understood, the transcription initiation site has not been accurately identified or adequately characterized. Our studies of the *dhfr* promoter region revealed unexpected structural features and functional properties. In this paper we present evidence that the four highly conserved 48-base-pair (bp) repeats identified earlier to be at the 5′ end of the *dhfr* gene (9) function not only as a promoter for the primary *dhfr* transcript but also as a promoter for a divergently transcribed gene, despite the absence of any canonical TATA or CAAT sequence (5). In addition, there appear to be minor *dhfr* promoters further upstream which yield multiple RNA initiations and which overlap with the divergent transcript. This divergent gene appears to code for a protein, and the first exon has strong homology with a region present in the same position relative to the human *dhfr* gene. These results combined with previous results (20, 21) show that transcription of the mouse *dhfr* gene is strikingly different from most of the genes studied to date: the promoter region lacks consensus promoter sequences and functions bidirectionally, there are minor *dhfr* promoters which overlap with divergent transcripts, and variations in mRNA levels are not due to transcriptional regulation.

MATERIALS AND METHODS

DNA probes. The isolation and characterization of λCh4A phage containing portions of the mouse *dhfr* gene was described previously (9), as were some of the subclones used here (9, 12). To prepare the probes, fragments were cloned

* Corresponding author.
† Present address: Department of Biology, Emory University, Atlanta, GA 30322.
‡ Present address: Department of Embryology, Carnegie Institute of Washington, Baltimore, MD 21210.
§ Present address: Molecular Biology Research, The Upjohn Co., Kalamazoo, MI 49001.

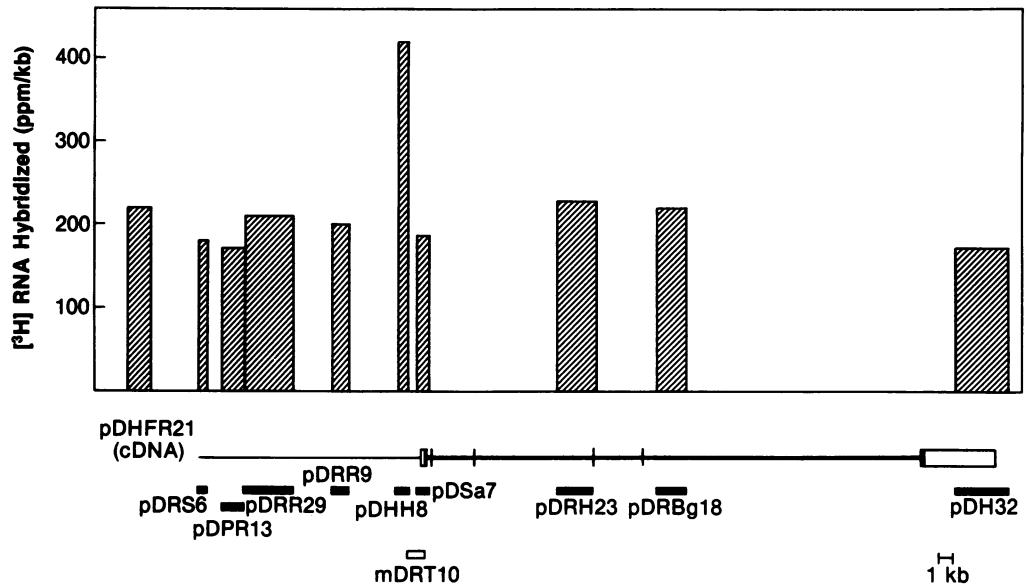

FIG. 1. Transcriptional activity measurements of sequences in the *dhfr* gene region. S180-M500 cells were pulse-labeled with [^{3}H]uridine for 5 min, and total RNA was isolated. The [^{3}H]RNA was hybridized to filters containing DNA corresponding to the regions of the *dhfr* gene as depicted at the bottom of the figure. The thin line represents genomic DNA, and the thick line represents the genomic DNA within the *dhfr* structural gene. The wide bars designate the six exons of the *dhfr* gene, and the open boxes on either end indicate the untranslated regions in the mRNA. The hatched bars indicate the amount of hybridization detected by each probe (shown in parts per million per kb). There was no detectable hybridization of pulse-labeled RNA from the parent (S180) cell line to any of these probes, indicating that the signal detected was coming from the amplified DNA.

directly from low-melting-point agarose gels into the indicated vector (8). pDRS6 is the 0.6-kb *Eco*RI-*Stu*I fragment of the 2.9-kb *Eco*RI fragment of λCh4ADHFR121 cloned into the *Eco*RI-*Hin*dII site of pEMBL8. pDPR13 is the 1.3-kb *Pvu*II-*Eco*RI fragment of the same 2.9-kb *Eco*RI fragment as that of pDRS6 cloned into the *Pvu*II-*Eco*RI sites of pEMBL9. pDRR29 is the 2.9-kb *Eco*RI-*Eco*RV fragment of the 6.4-kb fragment of λCh4ADHFR121 cloned into the *Eco*RI-*Eco*RV sites of pBR327. pDRR9 is the 0.9-kb *Eco*RV-*Eco*RI fragment of the same 6.4-kb *Eco*RI fragment cloned into the *Eco*RI-*Eco*RV sites of pBR327. pDHH8 is the 0.8-kb *Hin*dIII-*Hin*dII fragment of the 3.4-kb *Hin*dIII fragment of λCh4ADHFR121 cloned into the *Hin*dIII-*Hin*dII sites of pUC8. mDRR9 and mDHH8 are the cloned inserts of pDRR9 and pDHH8 inserted into the equivalent sites of both mp10 and mp11. mDRT10 is the 1.0-kb *Eco*RI-*Taq*I fragment of the 3.4-kb *Eco*RI fragment of λCh4ADHFR121 cloned into the *Eco*RI-*Acc*I sites of mp8 and mp11.

Transcription level measurements. The level of transcription of the *dhfr* gene region was performed by labeling S180-M500 cells for 5 min with [^{3}H]uridine and hybridizing to DNA probes on filters as previously described (21).

UV transcription mapping. One day before irradiation, S180-M500 cells were split at a ratio of 1:2 into 100-mm tissue culture dishes. The next day, the medium was decanted and saved, and the dishes were rinsed twice with Dulbecco phosphate-buffered saline without calcium or magnesium (PBS) and exposed in 5 ml of PBS to measured dosages of UV light from a germicidal lamp. Three dishes were exposed at one time. The PBS was decanted, and the cells were incubated for 30 min in the dark in the reserved medium. The medium was then decanted, and the cells were rinsed with PBS and labeled for 6 min in 1.5 ml of Eagle minimal essential medium with 225 μCi of [^{3}H]uridine as previously described (21). Cells were harvested, and RNA was isolated as described previously (21).

S1 nuclease mapping. Uniformly labeled probes for S1 nuclease mapping were made from mDRT10 in a modification of the method described by Ley et al. (19). mDRT10 (4μg) with the insert in the desired orientation was mixed with 30 ng of M13 sequencing primer pentadecamer (New England BioLabs, Inc.) and 2 μl of *Pol*I buffer (70 mM Tris [pH 7.5], 120 mM $MgCl_2$, 500 mM NaCl) in a final volume of 25 μl and incubated in a boiling water bath for 5 min. Then it was allowed to anneal at room temperature for 45 min. After the annealing step, the following reagents were added: 20 μl of nucleotide mix (0.25-strength *Pol*I buffer, 125 μM dGTP and dATP, 62.5 μM dCTP and dTTP), 2 μl of 0.1 M dithiothreitol, 4 μl (40 μCi) of [^{32}P]dCTP, 4 μl of [^{32}P]dTTP, and 1 μl of Klenow fragment of DNA polymerase I (1 U). After incubation for 30 min at 37°C, the reaction was terminated, and the DNA was precipitated by the addition of a solution consisting of 5 μl of 0.1 M EDTA, 34 μl of 7.5 M ammonium acetate, and 200 μl of ethanol. After precipitation, the DNA was redissolved and restricted as desired. A full-length probe was generated by restriction with *Eco*RI and *Hin*dIII; smaller probes were made by further restriction digestion of this fragment. Probes were purified by preparative gel electrophoresis in agarose, followed by extraction with phenol (36), or electroluted from acrylamide urea sequencing gels.

A 5′-end-labeled probe corresponding to the above probe was made by labeling the *Taq*I sites of p*dhfr*2.9 with T4 polynucleotide kinase (22), followed by restriction with *Eco*RI and purification of the desired fragment.

For hybridization and S1 nuclease treatment (11), DNA and RNA were mixed together and ethanol precipitated. They were redissolved in 16 μl of deionized formamide, incubated for 12 min at 85°C, and quick-chilled on ice. A 4-μl volume of buffer was added to bring the solution to a final concentration of 80% formamide–40 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid; pH 6.4]–400 mM NaCl. Af-

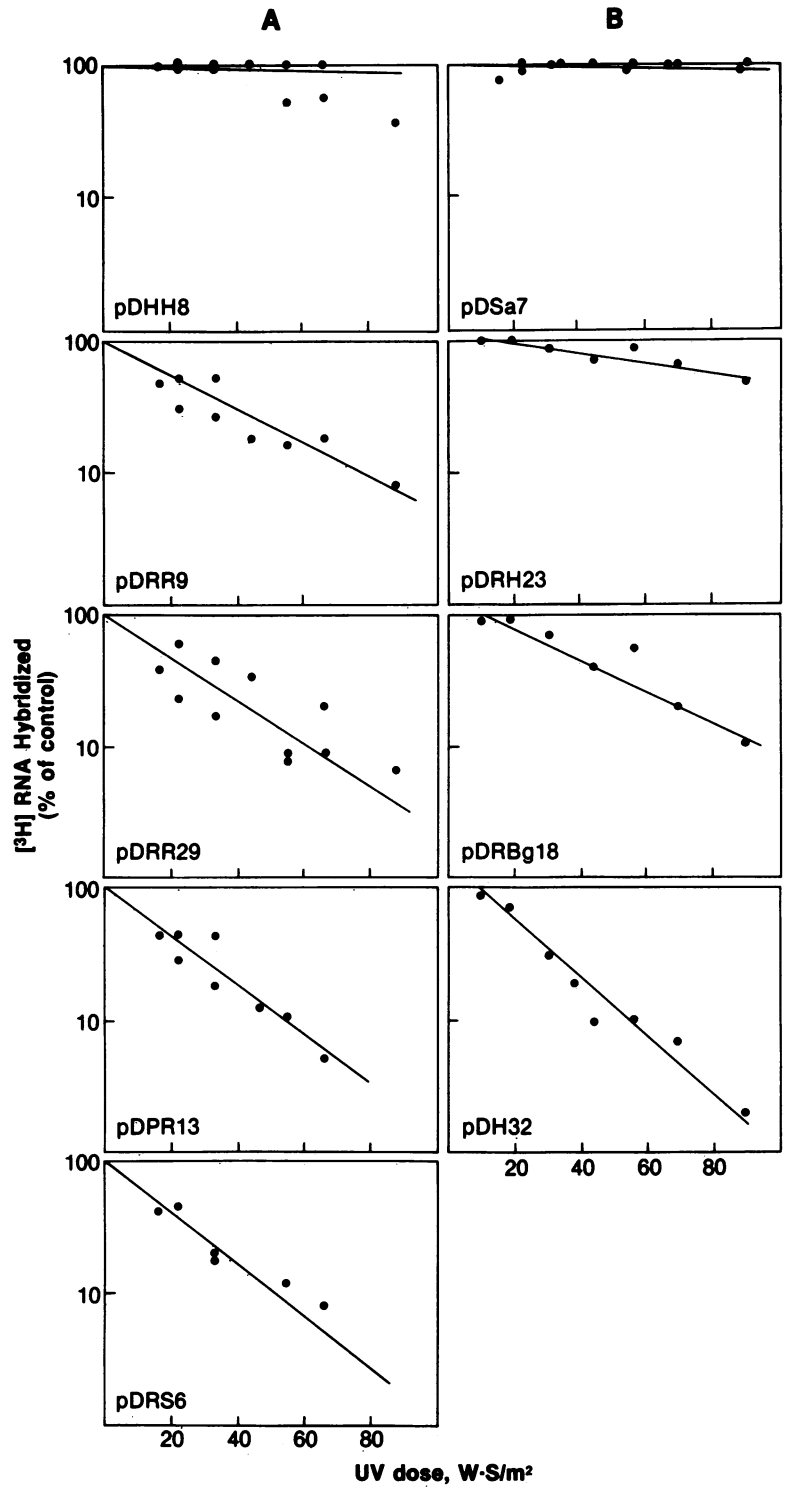

FIG. 2. Effects of UV irradiation on transcriptional activity in the *dhfr* gene region. S180-M500 cells were irradiated with UV at the doses indicated and 30 min later were pulse-labeled with [³H]uridine for 6 min. Total RNA was prepared and hybridized to filters containing probes 5′ of the *dhfr* gene (A) or in the *dhfr* gene (B). For each probe the amount of [³H]RNA hybridized is expressed as a percentage of control values (non-UV treated) and plotted as a function of increasing UV dosage. In each case the RNA was partially broken with alkali before hybridization.

ter incubation at 54°C overnight, 300 μl of cold S1 solution was added (50 mM sodium acetate [pH 4.6], 150 mM NaCl, 1 mM $ZnCl_2$, 300 U of nuclease S1 per ml), followed by incubation at 37°C for 30 min. The samples were precipitated by the addition of 85 μl of stop solution (2.4 M ammonium acetate, 60 mM EDTA, 120 μg of glycogen per ml) and 410 μl of isopropanol. Pellets were washed with 70% ethanol, dissolved in 8 M urea, and electrophoresed on 6% acryl-

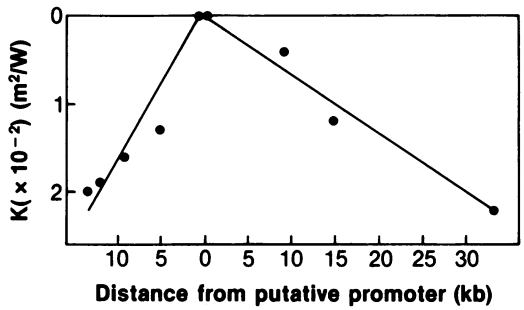

FIG. 3. Sensitivity to UV irradiation as a function of genomic distance. The slopes of the UV inactivation graphs of Fig. 2 are plotted as a function of the genomic map position of each probe used.

amide urea sequencing gels (22). Hybridizations were done in DNA excess; the incubation conditions effectively prevented probe renaturation (see Fig. 6b, lanes 2, 5, 8, and 11).

Primer extension. The desired labeled primer was hybridized to RNA as above and then ethanol precipitated. cDNA was synthesized in a reaction containing 100 mM Tris · hydrochloride (pH 8.3), 100 mM KCl, 8 mM $MgCl_2$, 0.5 mM dithiothreitol, 500 μM each of the four deoxynucleoside triphosphates, and 5 U of avian myeloblastosis virus reverse transcriptase (Life Sciences, Inc.) for 1 h at 42°C (3). The DNA was ethanol precipitated and then electrophoresed as described above.

RESULTS

Transcription occurs for at least 14 kb 5′ of the *dhfr* gene. The mouse S180-M500 cell line contains approximately 1,000 copies of the *dhfr* gene, and due to this gene amplification, it is possible to isolate enough pulse-labeled RNA to measure transcriptional activity throughout the *dhfr* gene region. We used this technique, which had been used previously to define the 3′ boundary of the *dhfr* transcription unit (12), in an effort to identify the 5′ end of the *dhfr* transcription unit. Pulse-labeled total RNA from the S180-M500 cells was hybridized to unique sequences throughout the *dhfr* structural gene and 5′-flanking region (Fig. 1). Contrary to our expectation, approximately equal levels of transcription were found up to at least 14 kb 5′ of the *dhfr* ATG.

The *dhfr* 5′-flanking region contains transcription initiation sites for two genes that are transcribed in opposite directions. UV transcription mapping (32) was performed to determine how many transcription units were present in this region. This technique relies on the apparent inability of RNA polymerase to transcribe through DNA damaged by UV irradiation. Therefore, the greater the distance of a region of DNA from the promoter, the greater the sensitivity to UV irradiation due to an increasing target size. The experiments were done by irradiating cells with various doses of short-wave UV light, incubating to allow all previously initiated polymerases to run off, and then pulse-labeling cells as above (Fig. 2).

The slopes of the semi-log plots in Fig. 2 are linearly related to the distance from the promoter of a given segment of DNA (32). A graph of the slope versus distance along the DNA is given in Fig. 3. From the graph, one can see that with increasing distance from pDSa7 through the *dhfr* gene there was increasing sensitivity to UV light, as expected for one transcription unit. The same was also true with increasing distance upstream of pDSa7, from pDHH8 to pDRS6. It is clear that these sequences are not part of the *dhfr* transcription unit; instead, they appear to form part of a separate transcription unit, transcribed in the opposite direction from *dhfr*.

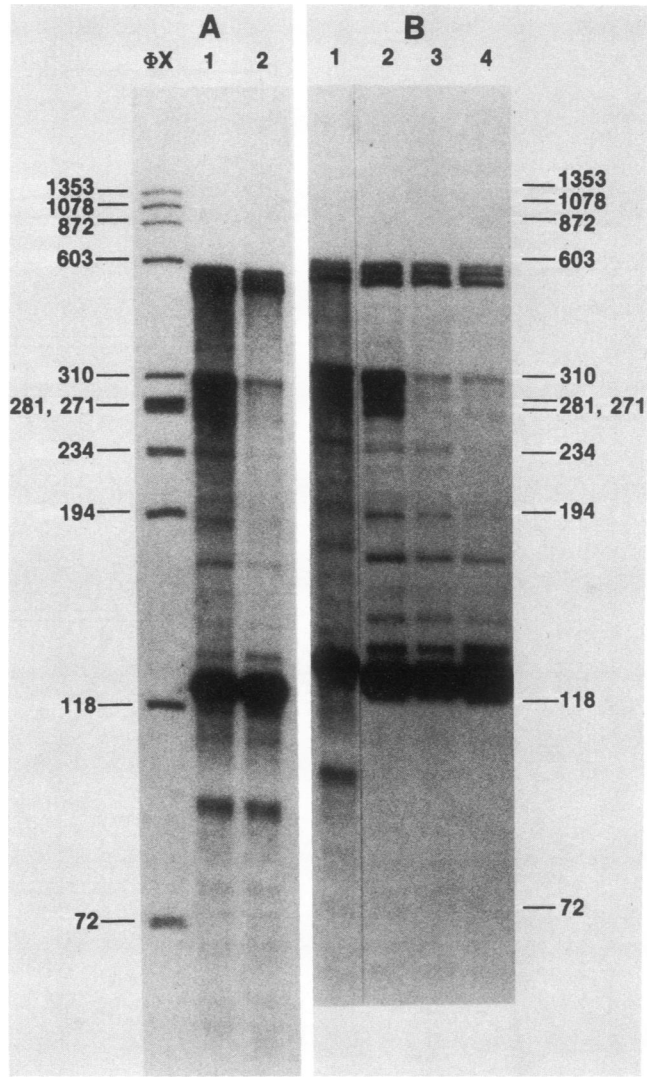

FIG. 4. S1 nuclease analysis of the *dhfr* 5′ end with uniformly labeled and 5′-end-labeled probes. (A) A uniformly labeled, 1-kb probe from mDRT10 was hybridized to S180-M500 RNA, treated with S1 nuclease, and electrophoresed as described in the text. The size markers are 5′-end-labeled ΦX-*Hae*III fragments. Lanes: 1, 9 μg of total RNA; 2, 10 μg of cytoplasmic RNA. (B) Lane 1, A uniformly labeled probe was hybridized to 9 μg of total S180-M500 RNA; lane 2, the same 1-kb fragment of DNA as in lane 1 was 5′ end labeled at the *Taq*I site (see Fig. 6B) and hybridized as above to 9 μg of S180-M500 total RNA; lane 3, the end-labeled probe was hybridized to 1.2 μg of S180-M500 polyadenylated RNA; lane 4, the end-labeled probe was hybridized to 10 μg of S180-M500 cytoplasmic RNA.

The two transcripts are transcribed from different strands of DNA. The UV transcription mapping results indicated that the separate transcription units were transcribed in opposite directions and thus must utilize different strands of DNA as templates for transcription. To confirm this result, various segments of the region were cloned into M13 phage in both orientations, and each strand was hybridized with pulse-labeled RNA as in Fig. 1. The results (Table 1) show that essentially all of the *dhfr* cDNA hybridization was to one

TABLE 1. Strand-specific measurement of transcriptional activity flanking the 5′ end of the *dhfr* gene[a]

Genomic DNA strand (relative to DHFR)	Activity in:			
	5′-flanking region			*dhfr* gene (mDHFR11)
	mDRR9	mDHH8	mDRT10	
+	0	10	44	200
−	230	350	100	7

[a] S180-M500 cells were pulse-labeled for 5 min with [^{3}H]uridine, and total RNA was extracted. The amount of radioactive RNA complementary to each single-stranded probe was determined and expressed as parts per million hybridized per kb of DNA.

strand. Two clones of segments which are further 5′ of the *dhfr* gene, mDHH8 and mDRR9 (see Fig. 1 for locations), showed hybridization only to the strand opposite that of the *dhfr* gene, confirming the UV transcription mapping results. There was significant hybridization to both strands of mDRT10, a probe with a 1-kb insert spanning the presumptive initiation sites of the two divergent transcription units.

Both transcription units are sensitive to α-amanitin. To determine whether the divergent transcripts were transcribed by RNA polymerase II, the α-amanitin sensitivity of pDHH8 and pDRR9 was determined. Transcription from each of these regions was as sensitive to α-amanitin as was transcription from the *dhfr* gene (Table 2). Note that all the transcription from pDHH8, which was anomalously high (Fig. 1), was inhibited by α-amanitin.

The *dhfr* transcripts have multiple 5′ ends. The experiments shown above indicated that there were no sequences 5′ of mDRT10 which were part of the *dhfr* gene and therefore implied that sequences from the 5′ end of the *dhfr* RNA must be in this region. Consequently, sequences from mDRT10 were used to map the 5′ end of the *dhfr* mRNA by hybridization and treatment with S1 nuclease. To detect any RNA in this region, whether or not it extended to the *dhfr* coding sequences, uniformly labeled DNA was used as a probe. These probes were made, as described in detail above, by cloning the desired sequence into an M13 vector, replicating across the cloned sequence, and excising the double-stranded labeled fragment. The isolated fragment, labeled in only one strand, was hybridized to RNA and then treated with S1 nuclease as described above. The clone used to map the *dhfr* DNA, mDRT10, contains an insert in M13mp8 covering the regions from a *Taq*I site just inside the *dhfr* coding sequence to an *Eco*RI site located 1 kb upstream. The result of hybridization of this probe to RNA from S180-M500 cells is shown in Fig. 4A. There was a major band of 122 nucleotides, a series of bands from 270 to 310 nucleotides, with the most prominent being at 310 nucleotides, and a triplet from 500 to 560 nucleotides. The middle series of bands is markedly reduced in cytoplasmic RNA. Because the probe was uniformly labeled, it was not possible, without additional information (see below), to assign the position of the nuclease-protected fragments within mDRT10.

To determine which of the RNA species continued through to the *Taq*I end of the probe, near the start of *dhfr* translation, the *dhfr* fragment present in mDRT10 was end labeled at the *Taq*I site with T4 polynucleotide kinase. This

TABLE 2. Transcript sensitivity to α-amanitin

α-Amanitin	[^{32}P]RNA hybridized (ppm/kb)		
	pDHFR21	pDHH8	pDRR9
−	230	440	270
+	10	5	0

[a] S180-M500 nuclei were incubated in vitro with [^{32}P]UTP in the presence or absence of 1 μg of α-amanitin per ml as described previously (20). [^{32}P]RNA was prepared and hybridized to the probes indicated. Hybridizations were done in triplicate with an increasing amount of input [^{32}P]RNA, ranging from 1×10^6 to 5×10^6 cpm. Fifty parts per million were bound to background filters. The results of one experiment are shown.

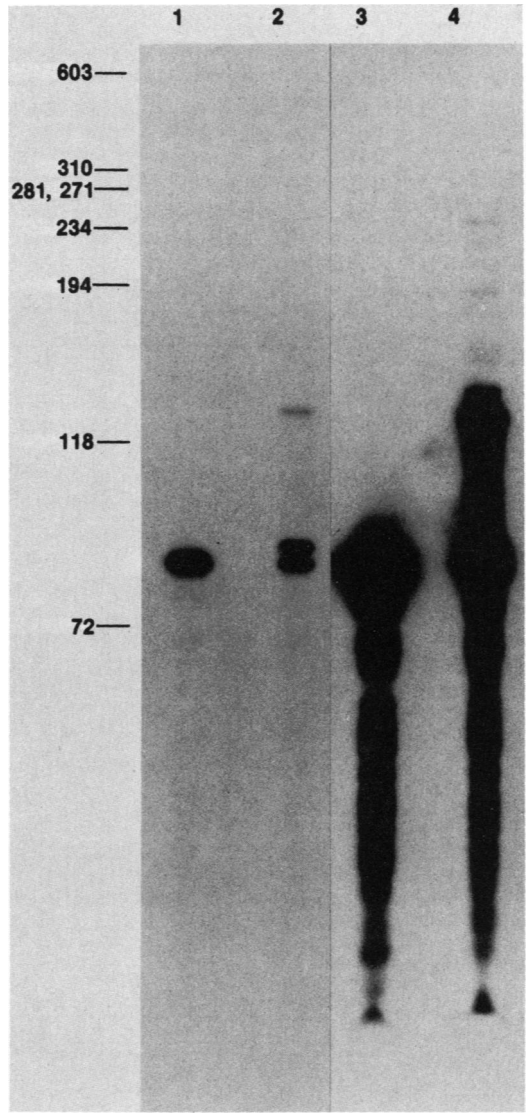

FIG. 5. Primer extension analysis of the *dhfr* 5′ end. A DNA fragment uniformly labeled by T4 DNA polymerase extending from a *Dde*I site 73 bp upstream of the *dhfr* AUG to the *Taq*I site just downstream of the AUG was hybridized to S180-M500 RNA as for S1 analysis. (Because of the particular clone used, the primer contained an additional 2 bp beyond the *Taq*I site.) After hybridization, the DNA was ethanol precipitated, redissolved, and extended with reverse transcriptase as described in the text. After ethanol precipitation, the products were electrophoresed on a urea acrylamide gel. Lane 1, Primer before reaction; lane 2, primer extension products. The band slightly above the primer is the primer with the *Dde*I site filled in by the reverse transcriptase. Both lanes are from a 1-h exposure. Lanes 3 and 4 are 16-h exposures of lanes 1 and 2. The bands below the primer are degradation products of the primer.

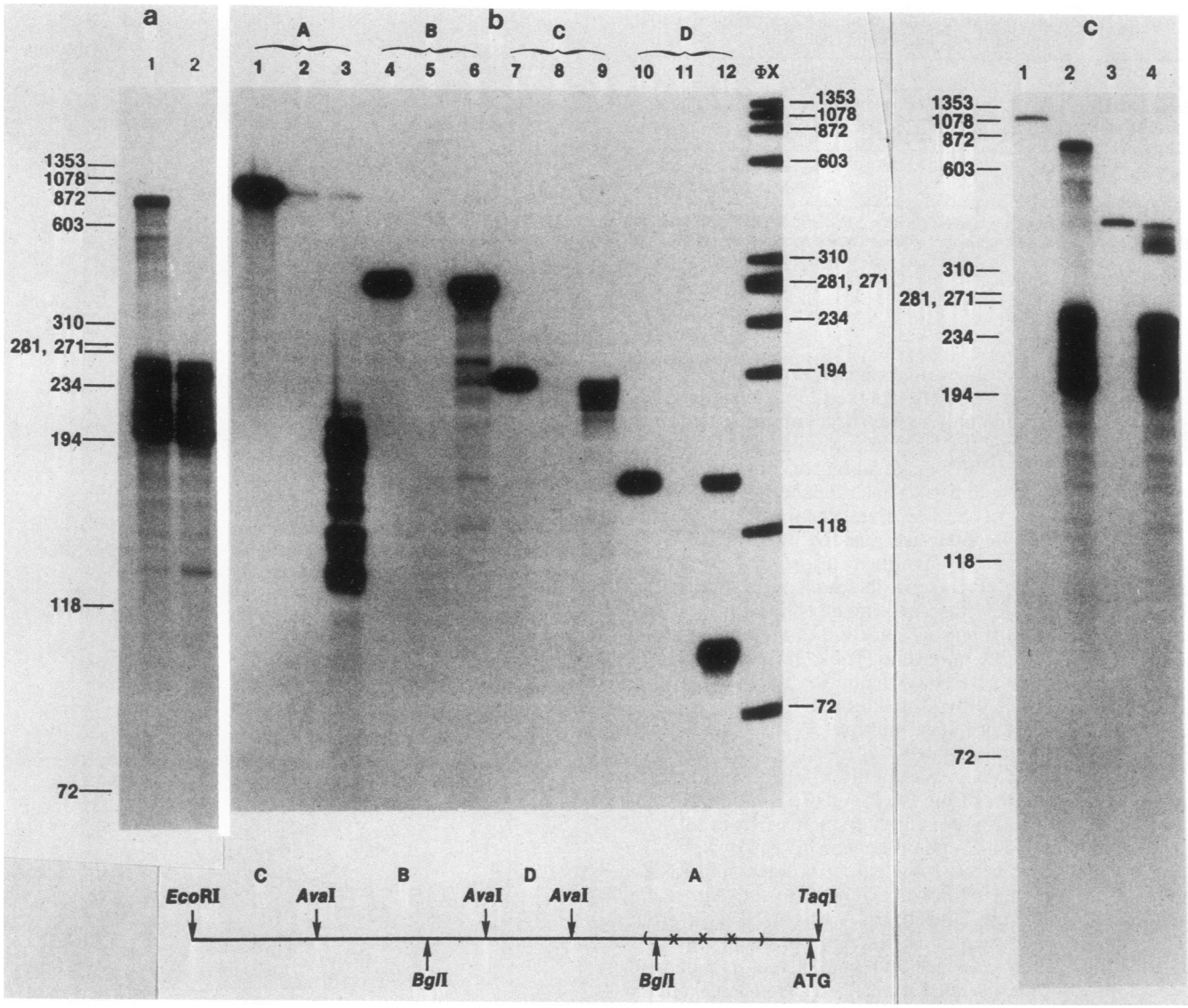

FIG. 6. S1 nuclease analysis of the divergent transcript with uniformly labeled probes. (a) A uniformly labeled, 1-kb probe from mDRT10 in the opposite orientation to that in Fig. 4 was hybridized to 9 μg of total RNA (lane 1) and to 10 μg of cytoplasmic RNA (lane 2) and treated with S1 nuclease as before. (b) The probe used in Fig. 6a was further restricted with *AvaI*, and the resulting four fragments were isolated and individually hybridized to total RNA. Lanes 1, 4, 7, and 10 are the isolated fragments before hybridization; lanes 2, 5, 8, and 11 are hybridized and treated with S1 nuclease without added RNA; and lanes 3, 6, 9, and 12 are hybridizations of the indicated fragment with 27 μg of S180-M500 total RNA. Below the lanes is a diagram of the mDRT10 probe. The parentheses indicate the location of the fourfold 48-bp repeat. (c) Lanes 1, Full-length mDRT10 probe; 2, hybridization of the probe in lane 1 to 9 μg of S180-M500 RNA; 3, 378-bp internal *BglI* fragment of the probe in lane 1; 4, *BglI* fragment hybridized to 9 μg of S180-M500 RNA.

fragment was then hybridized under the same conditions as above (Fig. 4B). The size distribution of nuclease-protected fragments was virtually the same for the end-labeled and uniformly labeled probes. The small difference in size between the lower bands in lanes 1 and 2, 3, and 4 was due to the additional nucleotides that the uniformly labeled probe had on the *TaqI* end which were complementary to the RNA and which were not present in the end-labeled probe. Because the same bands were present with the end-labeled probe as with the uniformly labeled probe, it can be concluded that all of the observed RNA species were colinear with the genomic DNA from the ATG upstream to the 5′ end of the RNA. There is one band of approximately 100 nucleotides seen with the uniformly labeled probe which is not observed with the end-labeled probe, and we are not sure whether this represents a breakdown product of the probe or a minor RNA species which does not extend to the ATG. Both cytoplasmic RNA and polyadenylic acid-containing RNA were greatly depleted in the RNA responsible for the 270- to 310-nucleotide fragments.

As another method of analyzing the 5′ end of the *dhfr* RNA, an 85-bp fragment of DNA from the 5′ end of exon I was labeled, hybridized to total RNA, and incubated with reverse transcriptase and deoxynucleoside triphosphates (see above) (Fig. 5). A short exposure of the primer-extended product (lane 2) revealed only a single band, other than that of the primer, at 122 nucleotides. Thus the reverse transcriptase extended the primer to the same position as that protected by S1 nuclease and corresponded to a major 5′ end at approximately −115 nucleotides from the ATG. A

longer exposure of the primer extension product (Fig. 5, lane 4) showed bands extending up to 250 nucleotides, and an even longer exposure (not shown) revealed products more than 500 nucleotides in length. The bands corresponding to products longer than 122 nucleotides were clearly present in much lower concentration than those observed in the S1 analysis. We attribute the weakness of the upper bands in the primer extension to the relatively long transcripts that must be made and in particular to the highly GC-rich region beyond the 122 nucleotide length (see Fig. 9), which could be difficult for the reverse transcriptase to transcribe.

The divergent transcript also has multiple 5′ ends. The results presented above (Fig. 1, 2, and 3 and Table 1) suggested that the promoter for the divergent transcript should be found within the region cloned in mDRT10. Although only a small part of this transcript might be stable, it was likely that part of the stable transcript should be at the very 5′ end of the transcription unit, as the first exon. Therefore, a uniformly labeled hybridization probe was used as before, with the same sequence as in mDRT10 cloned in the opposite orientation of M13 so that the opposite strand from above would be labeled. The result of hybridizing this probe to the same total RNA as before is shown in Fig. 6a. There were clearly substantial amounts of complementary RNA present. In fact, when probes of equal specific activity were hybridized to equivalent amounts of RNA, the sum of the intensity of the bands produced by the probe hybridizing to the upstream transcript was equal to, if not greater than, the sum of the intensities of the *dhfr* bands (results not shown). The pattern of bands was complex, with a minor band of more than 700 nucleotides and an array of major bands ranging in size from 190 to 250 nucleotides. All but the band at 700 nucleotides were found in cytoplasmic RNA (lane 2).

In an effort to determine which parts of the uniformly labeled probe were responsible for the observed hybridization, the double-stranded, labeled insert was further restricted with *Ava*I, to give the four fragments indicated at the bottom of Fig. 6b. Each of these fragments was isolated by gel electrophoresis and individually hybridized to S180-M500 total RNA. The results (Fig. 6b) indicated that none of the fragments showed any hybridization without RNA but that each of the four fragments hybridized to the RNA. Only part of fragment A hybridized to the RNA, and the RNA which hybridized gave multiple-sized bands, with a length distribution of approximately 100 to 155 bp, whose range of 55 bp is similar to that shown by the bands in the full-length probe. The next fragment 5′, D, showed two discrete bands of hybridization: a full-length band and a band approximately 84 nucleotides in length. Fragments B and C showed only full-length hybridization, although there is an indication of a band slightly smaller than full length in fragment C.

To check the positioning of the nuclease-protected fragments within mDRT10, S180-M500 RNA was hybridized with both a full-length probe and an internal 378-bp *Bgl*I fragment, whose location is indicated at the bottom of Fig. 6b (Fig. 6c). Both probes showed nearly identical multiple bands of 190 to 250 nucleotides in length. The only difference was that the top band of this series, at approximately 250 nucleotides, was missing in lane 4 and the next lower band was correspondingly more intense. Also, the band in lane 2 of approximately 700 nucleotides was replaced in lane 4 with a series of bands ranging in size from 330 to 380 bp.

The results of these experiments can be explained as follows. The 378-nucleotide *Bgl*I fragment when hybridized to total RNA yielded nearly the same pattern of bands of 190 to 250 nucleotides as that given by the full-length probe. The only exception is that the top band in this group was shortened. Therefore, most of the hybridizing RNA that yields this series of bands must have come from the *Bgl*I fragment. The 84-nucleotide band observed with fragment D when added to the sizes of the bands observed with fragment A gives the sizes of bands seen with the full-length probe. Therefore, the RNA of this series must have started from various points in fragment A and have a unique termination site in fragment D, 84 nucleotides from the *Ava*I site. On the basis of sequence analysis, we believe this apparent termination site is actually the result of a splice to downstream sequences (see below). The size of the top band of this series would extend just beyond the *Bgl*I site in fragment A, accounting for the shortened size observed when the internal *Bgl*I fragment was used for the hybridization. The bands corresponding to higher-molecular-weight species observed in Fig. 6a and c were due to another population of RNA which begins in the same region of fragment A as the shorter RNA molecules but extends throughout the rest of the probe to the *Eco*RI site. As these RNA molecules were not present in cytoplasmic RNA, we believe they represent unspliced transcripts (see below). This longer RNA appeared homogeneous in Fig. 6a, lane 1, and Fig. 6c, lane 2, but the use of the shorter *Bgl*I fragment in Fig. 6c, lane 4, revealed the heterogeneity in this longer RNA, with the bands ranging in size from 330 to 380 nucleotides.

A summary of the transcripts from both strands is given in Fig. 7. It can be seen that although the major *dhfr* transcript and the divergent transcripts are separated by more than 100 bp, the minor upstream *dhfr* transcripts actually overlap with the divergent transcript.

DISCUSSION

The *dhfr* transcription unit is approximately 36 kb. The work presented here combined with our previous results (12) serve to define the boundaries of the *dhfr* transcription unit. Hybridization of pulse-labeled RNA to filter-bound DNA showed that there was no transcription on the *dhfr* strand further than 1 kb 5′ of the *dhfr* ATG, and the level of hybridization to the probe mDRT10 suggested that transcription initiated within several hundred nucleotides of the ATG. The results of the S1 nuclease analysis were consistent with this result, as there were no detectable transcripts further than 600 bp from the ATG. Previous work has shown that transcription terminates approximately 6 kb from 3′ of the *dhfr* termination codon (12). The data from the UV mapping experiments suggested that all of the transcribed RNA was due to initiation in the same region and that there was no significant transcription initiation within the transcribed DNA. Therefore, the entire *dhfr* transcription unit, from initiation to termination, is approximately 36 kb.

It is clear from the results presented in Fig. 4 that *dhfr* mRNA has multiple 5′ ends. We do not believe they represent artifactual degradation products, as the same pattern has been observed from different RNA preparations, including total RNA preparations isolated by a guanidinium thiocyanate procedure. Several lines of evidence suggest that these multiple ends are the result of multiple initiations, rather than the result of posttranscriptional processing. The 5′ ends do not appear to be due to normal splicing events, as there is no evidence for sequences on the 5′ end of the RNA which are noncontiguous in the DNA; none of the 5′ ends are near sequences which resemble consensus splice acceptor sites; and all of the ends, except the major one at −115, are

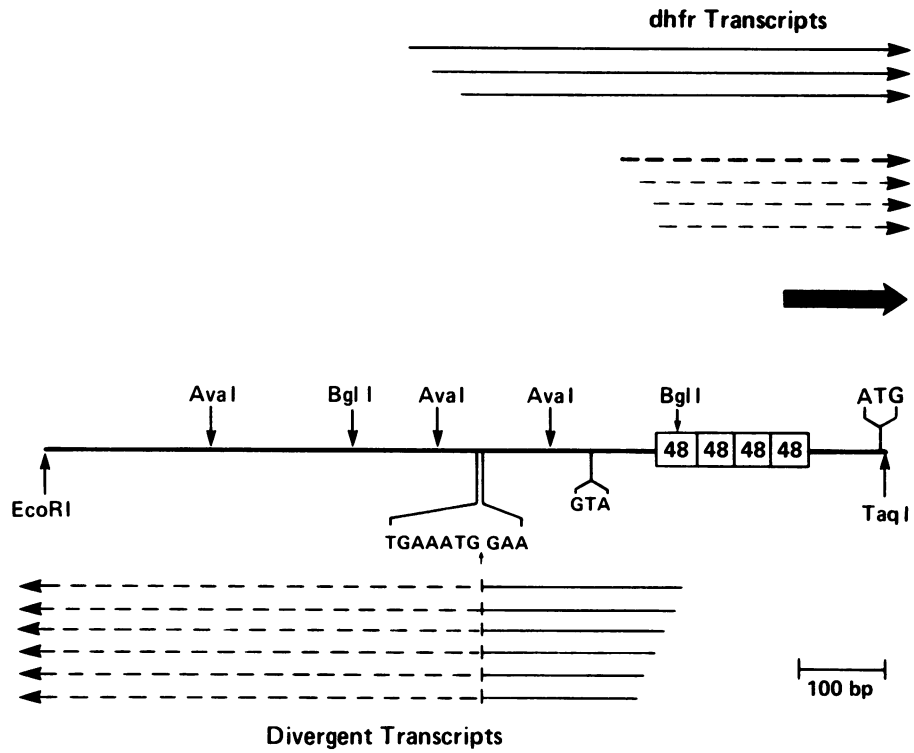

FIG. 7. Map of the RNA transcripts near the 5′ end of the *dhfr* gene. The central line represents the sequence contained in mDRT10, as shown in Fig. 6b. The *dhfr* transcripts are indicated above the lines, and the divergent transcripts are indicated below the line. The 48-bp repeats are indicated, as are the ATG for *dhfr* and the presumed ATG and splice junction for the divergent transcript. Portions of the transcripts not seen in cytoplasmic RNA are indicated by a dotted line.

part of a staggered series. Likewise, there is no evidence in the literature for processing steps which remove sequences from the 5′ end of RNA polymerase II transcripts. In addition, if all transcripts initiated at the site of the longest transcript, we would expect the hybridization observed with mDRT10 in Table 1 to be approximately 2.5-fold greater than that actually observed.

Simonsen et al. (37) and McGrogan et al. (23), who first discovered the multiple 5′ ends of *dhfr* mRNA, also concluded that these ends were the result of multiple initiations, and McGrogan et al. (23) showed that at least two different regions could serve as promoters in transfection experiments. The transcripts with 5′ ends at −115 nucleotides and between −500 and −560 nucleotides have the same relative abundance in total, polyadenylated, and cytoplasmic RNA (Fig. 4) as would be expected for independent transcripts. In contrast, the transcripts with 5′ ends from −270 to −310 nucleotides are present at a much lower concentration in polyadenylated and cytoplasmic RNA than in total RNA (Fig. 4). Although it is possible that these transcripts do represent a precursor to the −115 nucleotide transcript, we favor the view that either they are unstable out of the nucleus or they fail to be transported from the nucleus. Either case would be interesting. One would expect, unless the site of initiation influences transcription termination and processing in some fashion, that the 3′ portions of the *dhfr* transcripts would be established independently of the 5′ end. This would suggest that the lack of mRNA sequences with 5′ ends from −270 to −310 nucleotides in the cytoplasm would be due to sequences in the 5′ untranslated region of the RNA.

Multiple initiation sites for transcription of a gene have been observed in other systems: a zein gene in maize (18), the ε-globin gene of humans (1), the ADH gene of *Drosophila melanogaster* (4), and the α-amylase gene of mice (38). In the last two cases, the 5′ ends are correlated with tissue specificity, but the reason for the existence of multiple 5′ ends is less clear in the first two examples. The situation is more complex in the *dhfr* gene, for there are multiple 5′ ends within each of the longer classes of transcripts. In addition, there are seven different size classes of *dhfr* mRNA, based on variations in the 3′ untranslated region content of the mRNA. We do not know how the 5′ and 3′ heterogeneity fit together. The simplest hypothesis would be that 5′ and 3′ lengths are independent, so that every transcript would have a certain probability of containing a given 5′ and 3′ end.

The divergent transcript appears to code for a spliced mRNA. We identified the 5′ ends and the first exon of the divergent transcript. The results of single-strand hybridization measurements (Table 1) showed that hybridization to the negative strand of mDRT10 (relative to *dhfr*) was lower than would be expected if the entire fragment were transcribed, suggesting that transcription initiated within this fragment. (In fact, the level of transcription was lower than expected for the observed transcript; the expected measurement would be about 1.5-fold higher, if we assume that the level of transcription in mDRT10 should be equal to that of the probes further 5′, such as pDRR9). Both the longer set of protected fragments, which gave a band of approximately 700 nucleotides, and the shorter set of protected fragments, which gave bands of 190 to 250 nucleotides (Fig. 6a), appeared to result from initiation over a 60-bp region of DNA from approximately −245 to −305 nucleotides relative

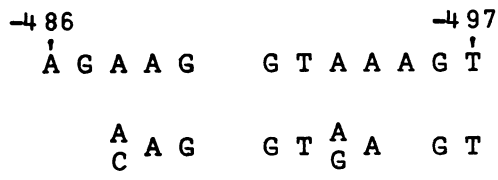

FIG. 8. Comparison of the region surrounding −490 and a consensus splice donor sequence. Shown in the top line is the sequence of the putative splice site in the divergent gene from nucleotides −486 to −497, numbered from the ATG of the *dhfr* gene (23). Below is given the consensus donor splice sequence from Breathnach and Chambon (5).

to the *dhfr* AUG. The longer set of transcripts were continuous throughout the rest of the probe to the *Eco*RI site, whereas the shorter transcripts terminated at −490 nucleotides. Inspection of the DNA sequence in this region (23) reveals an ATG at −357 nucleotides, followed by an in-frame open reading frame which extends to −528 nucleotides. Furthermore, the sequences around −490 nucleotides are an almost perfect consensus splice donor sequence (Fig. 8). Therefore, it appears that the shorter series of transcripts represent a spliced transcript and that the first exon of this RNA extends from the site of initiation to −490 nucleotides. Although this RNA is transcribed by RNA polymerase II, appears to be spliced, is polyadenylated and cytoplasmic, and has an open reading frame, we cannot be sure that it is translated until we find the corresponding protein product. This RNA appears to be amplified to the same extent as the *dhfr* mRNA. Various attempts have been made to look for amplified proteins other than *dhfr* in these and similar cells, without success. It may be that the protein made from this RNA is translated inefficiently, is unstable, or is obscured by more abundant proteins in various gel electrophoresis systems.

It is interesting to note that the minor transcripts yielding the longer set of fragments are unspliced and are found only in the nucleus and not in the cytoplasm (Fig. 6a). For a presumptive intermediate, their abundance in total RNA is relatively high (we estimate approximately 5%).

We cannot yet explain the high level of transcription observed with the pDHH8 probe. This probe, in either double-stranded form (Fig. 1) or single-stranded form (Table 1), gave up to twofold more hybridization than expected on the basis of other surrounding probes. pDHH8 has 260 bp of overlap with the mDRT10 probe and extends approximately 500 bp further 5′. This probe appears to be unique. It is possible that the high level of transcription was due to an undetected repeated sequence (although no hybridization was detected to RNA from the unamplified parental cell line) or to an additional transcript which does not extend as far as pDRR9, 3 kb from pDHH8. It is interesting to note that an in vitro promoter was mapped in the human *dhfr* gene region to a location analagous to the position of pDHH8 in the mouse (6; see below). Another puzzle is the result of the UV transcription mapping shown in Fig. 3. It is clear from this result that there are divergent transcripts in this region, but the different slopes for the two transcripts observed in Fig. 3 are unexpected. We are unsure whether the different slopes are the result of experimental variables or whether they are an indication of some biological differences in the transcripts, such as significant differences in the T content of the genes.

The two divergently transcribed genes share the same promoter region. As explained above, we believe that the 5′ ends of the transcripts that we identified correspond to sites of transcription initiation. Therefore, we would expect to find the promoters for transcription just upstream of these 5′ ends. The 5′ end of the most prominent *dhfr* transcript, at −115 nucleotides, is found in the fourth copy of a striking 48-bp repeat (Fig. 9). There is no classical TATA box upstream; there is a CACAATA sequence 20 bp upstream of the transcript start, but there is no evidence that this sequence has any effect on *dhfr* transcription.

The longest transcripts of the divergently transcribed gene initiate within the first copy of the 48-bp repeat (Fig. 7). Because the four copies of the repeats are so similar (Fig. 9), the same sequences are apparently able to serve as promoters for both transcripts. One question which remains unanswered is why only the copy of the repeated sequence closest to the gene is used to initiate transcription. The few differences between repeats may be significant, or an addition effect of the repeats may be necessary for initiation. There is no sequence in the negative strand of these repeats which has any semblance to a TATA box, so we hypothesize that: (i) a TATA box is not a necessary part of all eucaryotic promoters, and (ii) the same region of DNA can serve as a promoter for divergent transcripts.

Both of the above statements may be true for a number of eucaryotic genes and not just *dhfr*. For example, none of the multiple initiations 5′ of the major cap site of the human ε-globin gene have identifiable TATA boxes (1). We speculate that many genes which are expressed at relatively low levels will be found to have promoter regions different from those found on genes expressed at high levels, which includes most of the genes which have been analyzed closely. Another gene of the same type as *dhfr*, the hypoxanthine

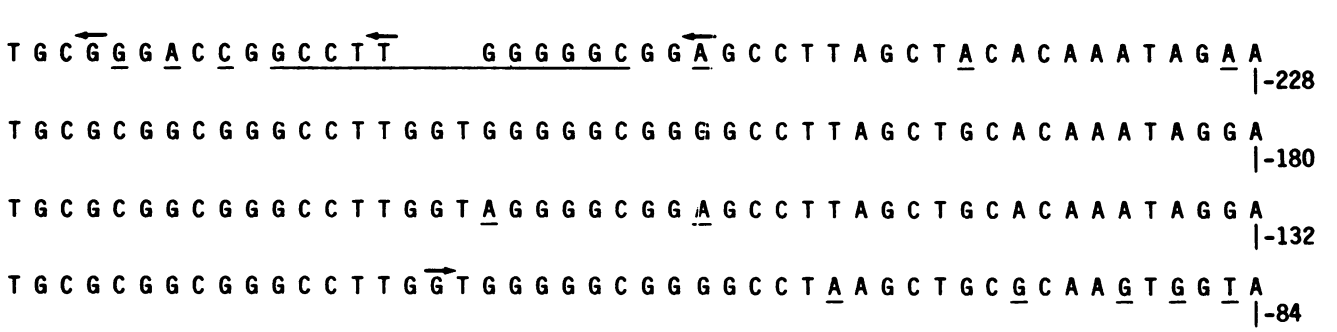

FIG. 9. Sequence of the major promoter region for the *dhfr* gene and the divergent transcript. Shown is the sequence from −84 to −272 nucleotides upstream of the *dhfr* ATG, with each line representing a copy of the 48-bp repeat. The bases underlined differ in sequence from the second copy of the repeat. A gap has been introduced into the first copy of the repeat to better align the sequence. The arrows denote the approximate locations of RNA 5′ ends. The *Bgl*I site is underlined.

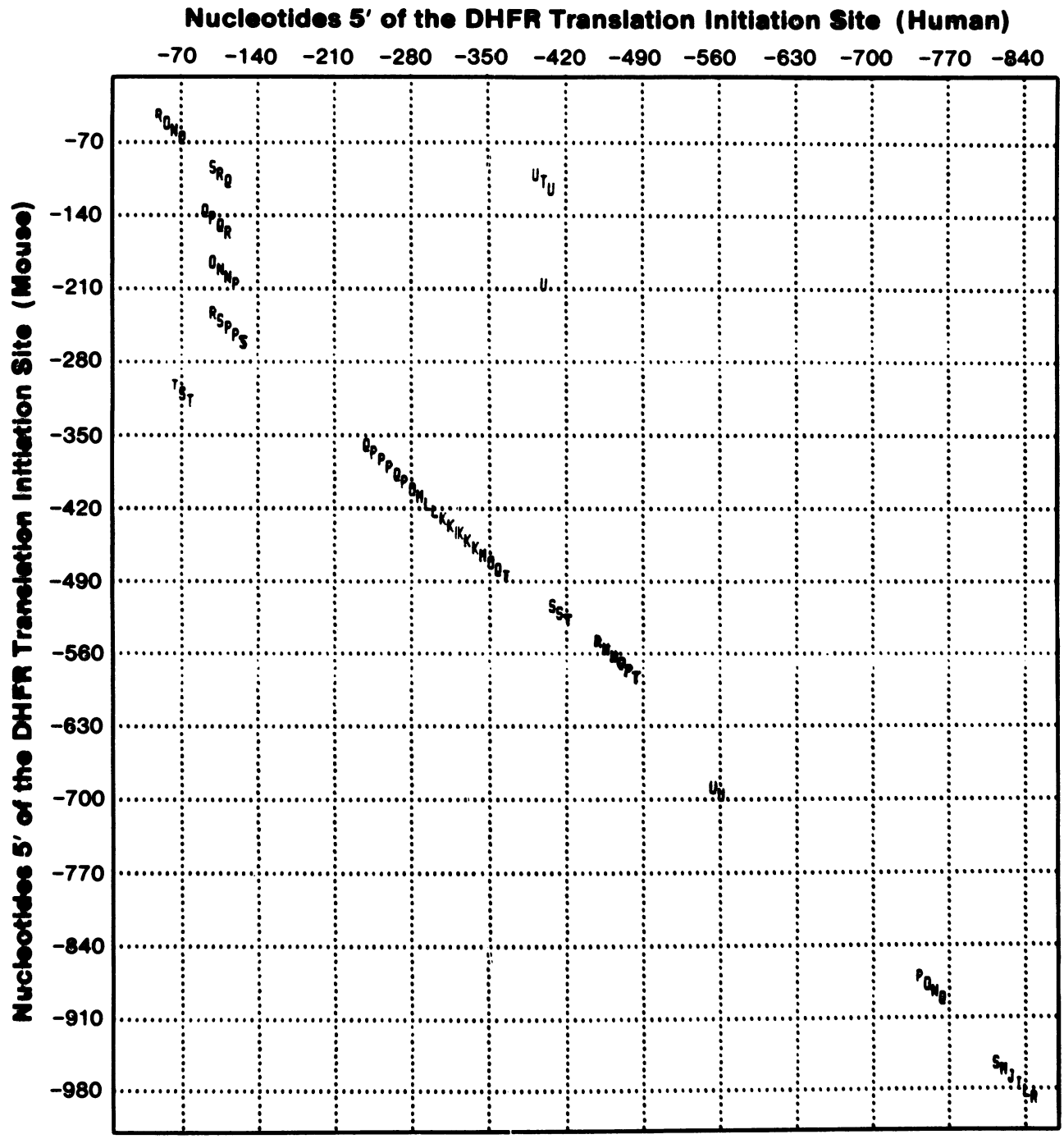

FIG. 10. Matrix homology analysis of the mouse and human *dhfr* genes. The 5'-flanking sequences of the human *dhfr* gene (6) and mouse *dhfr* gene (23) were compared by using the program of Pustell and Kafatos (30) as supplied by International Biotechnologies Inc. Some minor changes in the mouse sequence, detected in a partial resequencing of the region, were incorporated into the comparison. The ATG of *dhfr* is assigned base no. 1. The parameters used for this plot were a range of 20, a scale factor of 0.95, and a minimum value plotted of 60%. The letters represent the minimum value of the match, with A = 100% to Z = 50 to 51% (see reference 30 for a more complete description).

phosphoribosyltransferase gene, also is lacking a TATA box and has a 5' region which is highly rich in GC (25). We also note that there is a homology between the distal signals of the herpesvirus thymidine kinase gene (24) and sequences within each copy of the 48-bp repeat. The 5'-GGGCGG-3' sequence of the first distal signal is found in each copy of the repeat, and three of the four copies of the repeat have a longer homology with the inverted version of this sequence found in the second distal signal, 5'-CCCCGCCC-3'. It is interesting that the second distal signal has been found to function in both orientations. This region also shares homology with corresponding sequences in the 21-bp repeats of the simian virus 40 promoter, which are important for both early and late transcription (13, 14). We also find five copies of the sequence CCGCCC within 50 bp of the *dhfr* 5' ends at −500 to −560 nucleotides.

One of the most interesting comparisons of this region is with the putative promoter region of 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMG CoA reductase) (31). In HMG CoA reductase, there are three blocks of CCGCCC imbedded in a GC-rich region and there is no identifiable TATA box. The HMG CoA reductase mRNA also has multiple 5′ ends, although these are present within a region of only 90 nucleotides. In this respect, it is important to note that the major *dhfr* transcript at −115 nucleotides appears to have a precise start. It may be that there are elements within the 48-bp repeat, missing in the HMG CoA reductase gene, which serve to position accurately the start of this transcript. This most logically would be the CACAAATA sequence in the 48-bp repeats which bears some resemblance to the TATA box, which appears to position the start site of transcription (5, 26). The CACAAATA sequence is not present in the complementary strand of DNA, and so the divergent transcript has heterogeneous starts. Other than transcripts within certain gene families, we are unaware of any other chromosomal genes in higher eucaryotes that have divergent mRNA transcripts. If such divergent transcripts were to be found chiefly among housekeeping genes, the failure to find divergently transcribed genes could be explained by the relatively few genes examined to date and by the failure to look for such transcripts. Padgett et al. (28) found a presumptive mRNA transcript in a restriction fragment immediately 5′ of a fragment hybridizing to the CAD gene. This transcript was not further characterized, and it is not clear from which strand it came, but it may be another example of a divergent transcript.

The mouse and human *dhfr* promoter regions share many features. The sequences of the 5′ region of the mouse *dhfr* gene (23) were compared by matrix homology analysis (30) with the equivalent region of the human *dhfr* gene (6), the only other eucaryotic *dhfr* gene sequenced in this region. A similar comparison was made by Chen et al. (6), but the use of the Pustell program (Fig. 10) gave a lower background than that of Chen et al. There are several regions of homology throughout the 1 kb of mouse *dhfr* plotted, which corresponds to the sequences in mDRT10. As one proceeds 5′ from the ATG in the mouse *dhfr* sequence, the first striking homology is a match of 21 of 22 bases in the 5′ untranslated region of the mRNA. We do not know what the significance of the sequence might be. The next homology is with the fourfold 48-bp repeat in mice. There is only one copy of this repeat in humans, resulting in the four parallel lines in the matrix analysis. There is then only scattered homology up to approximately −360 from the mouse ATG. There is quite significant homology between −360 and approximately −490. These endpoints correspond to the ATG at −357 and the splice junction at −490 of the upstream transcript. Therefore, it looks as though the same gene that is just upstream of the mouse *dhfr* gene is in the equivalent position in the human genome and might also share the same promoter region. We do not know whether this gene is transcribed in humans.

There is another region of homology just beyond the putative first exon of the upstream transcript in mice, at −550 to −580. We do not know the significance of this homology but note that it is the probable promoter region for the longest *dhfr* 5′ ends. A final intriguing region of homology is that at −860 to −900 nucleotides and at −950 to −1,000 nucleotides in the mouse sequence. An examination of the mouse sequence in this region reveals an ATG at −868, followed by an open reading frame extending to the end of the sequenced fragment. No transcript has been identified as starting in this region, although it is interesting to note the increased transcriptional activity observed with this region in Fig. 1 (see pDHH8) and to note that Chen et al. (6) mapped an in vitro promoter to the equivalent region in human DNA, transcribed from the strand opposite *dhfr*. It is possible that this region could represent the second exon of the upstream transcript, although the S1 analysis gave no evidence of such a product, and such an explanation would not account for the higher transcriptional activity in this region. It is also possible, as we suggested above, that there is an additional transcript from this region which extends less than 3 kb to the region cloned in pDRR9. If such a transcript existed, it would be within the first intervening sequence of the divergently transcribed gene.

ACKNOWLEDGMENTS

We thank Kathleen Dixon and Cheeptip Benyajati for helpful discussions during the course of this work and Robert Devlin and Victoria Finnerty for comments on the manuscript. We are grateful to Chris Simonsen for providing the sequence of the mouse 5′-flanking region before publication and to J. Michael Clark for assistance in the preparation of the manuscript.

This research was supported by the National Cancer Institute under contract NO1-CO-23909 with Litton Bionetics, Inc. (G.F.C. and R.N.M.) and by Public Health Service grant GM 30204 from the National Institutes of Health and grant PCM 8104533 from the National Science Foundation to R.E.K. R.E.K. was supported by a Research Career Development Award (1 KO4 CA00828). Methotrexate was supplied by the Drug Synthesis and Chemistry Branch, Div. of Cancer Treatment, National Cancer Institute.

LITERATURE CITED

1. **Allan, M., W. G. Lanyon, and J. Paul.** 1983. Multiple origins of transcription in the 4.5 kb upstream of the ε-globin gene. Cell **35**:187–197.
2. **Alt, F. W., R. E. Kellems, J. R. Bertino, and R. T. Schimke.** 1978. Selective multiplication of dihydrofolate reductase genes in methotrexate-resistant variants of cultured murine cells. J. Biol. Chem. **253**:1357–1370.
3. **Benyajati, C., and J. F. Dray.** 1984. Cloned *Drosophila* alcohol dehydrogenase genes are correctly expressed after transfection into *Drosophila* cells in culture. Proc. Natl. Acad. Sci. U.S.A. **81**:1701–1705.
4. **Benyajati, C., N. Spoerel, H. Haymerle, and M. Ashburner.** 1983. The messenger RNA for alcohol dehydrogenase in Drosophila melanogaster differs in its 5′ end in different developmental stages. Cell **33**:125–133.
5. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
6. **Chen, M.-J., T. Shimada, A. D. Moulton, A. Cline, R. K. Humphries, J. Maizel, and A. W. Nienhuis.** 1984. The functional human dihydrofolate reductase gene. J. Biol. Chem. **259**:3933–3943.
7. **Cleveland, D. W., and J. C. Havercroft.** 1983. Is apparent autoregulatory control of tubulin synthesis nontranscriptionally regulated? J. Cell Biol. **97**:919–924.
8. **Crouse, G. F., A. Frischauf, and H. Lehrach.** 1983. An integrated and simplified approach to cloning into plasmids and single-stranded phages. Methods Enzymol. **101**:78–89.
9. **Crouse, G. F., C. C. Simonsen, R. N. McEwan, and R. T. Schimke.** 1982. Structure of amplified normal and variant dihydrofolate reductase genes in mouse sarcoma S180 cells. J. Biol. Chem. **257**:7887–7897.
10. **Darnell, J. E.** 1982. Variety in the level of gene control in eukaryotic cells. Nature (London) **297**:365–371.
11. **Favaloro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: analysis by two-

dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
12. **Frayne, E. G., E. J. Leys, G. F. Crouse, A. G. Hook, and R. E. Kellems.** 1984. Transcription of the mouse dihydrofolate reductase gene proceeds unabated through seven polyadenylation sites and terminates near a region of repeated DNA. Mol. Cell. Biol. **4**:2921–2924.
13. **Fromm, M., and P. Berg.** 1982. Deletion mapping of DNA regions required for SV40 early region promoter function in vivo. J. Mol. Appl. Genet. **1**:457–481.
14. **Fromm, M., and P. Berg.** 1983. Transcription in vivo from SV40 early promoter deletion mutants without repression by large T antigen. J. Mol. Appl. Genet. **2**:127–135.
15. **Groudine, M., and C. Casimir.** 1984. Post-transcriptional regulation of the chicken thymidine kinase gene. Nucleic Acids Res. **12**:1427–1446.
16. **Howe, C. C., D. K. Lugg, and G. C. Overton.** 1984. Post-transcriptional regulation of the abundance of mRNAs encoding α-tubulin and a 94,000-dalton protein in teratocarcinoma-derived stem cells versus differentiated cells. Mol. Cell. Biol. **4**:2428–2436.
17. **Kaufman, R. J., and P. A. Sharp.** 1983. Growth-dependent expression of dihydrofolate reductase mRNA from modular cDNA genes. Mol. Cell. Biol. **3**:1598–1608.
18. **Langridge, P., and G. Feix.** 1983. A zein gene of maize is transcribed from two widely separated promoter regions. Cell **34**:1015–1022.
19. **Ley, T. J., N. P. Anagnov, G. Pepe, and A. W. Nienhuis.** 1982. RNA processing errors in patients with β-thalassemia. Proc. Natl. Acad. Sci. U.S.A. **79**:4775–4779.
20. **Leys, E. J., G. F. Crouse, and R. E. Kellems.** 1984. Dihydrofolate reductase gene expression in cultured mouse cells is regulated by transcript stabilization in the nucleus. J. Cell Biol. **99**:180–187.
21. **Leys, E. J., and R. E. Kellems.** 1981. Control of dihydrofolate reductase messenger ribonucleic acid production. Mol. Cell. Biol. **1**:961–971.
22. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
23. **McGrogan, M., C. C. Simonsen, D. T. Smouse, P. J. Farnham, and R. T. Schimke.** 1985. Heterogeneity at the 5′ termini of mouse dihydrofolate reductase mRNAs. J. Biol. Chem. **260**:2307–2314.
24. **McKnight, S. L., R. C. Kingsbury, A. Spence, and M. Smith.** 1984. The distal transcription signals of the herpesvirus *tk* gene share a common hexanucleotide control sequence. Cell **37**:253–262.
25. **Melton, D. W., D. S. Konecki, J. Brennard, and C. T. Caskey.** 1984. Structure, expression, and mutation of the hypoxanthine phosphoribosyl-transferase gene. Proc. Natl. Acad. Sci. U.S.A. **81**:2147–2151.
26. **Nevins, J. R.** 1983. The pathway of eukaryotic mRNA formation. Annu. Rev. Biochem. **52**:441–466.
27. **Nunberg, J. H., R. J. Kaufman, A. C. Y. Chang, S. N. Cohen, and R. T. Schimke.** 1980. Structure and genomic organization of the mouse dihydrofolate reductase gene. Cell **19**:355–364.
28. **Padgett, R. A., G. M. Wahl, and G. R. Stark.** 1982. Structure of the gene for CAD, the multifunctional protein that initiates UMP synthesis in Syrian hamster cells. Mol. Cell. Biol. **2**:293–301.
29. **Piechaczyk, M., J. M. Blanchard, L. Marty, C. Dani, F. Panabieres, S. El Sabouty, P. Fort, and P. Jeanteur.** 1984. Post-transcriptional regulation of glyceraldehyde-3-phosphate-dehydrogenase gene expression in rat tissues. Nucleic Acids Res. **12**:6951–6963.
30. **Pustell, J., and F. C. Kafatos.** 1984. A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determination. Nucleic Acids Res. **12**:643–655.
31. **Reynolds, G. A., S. K. Basu, T. F. Osborne, D. J. Chin, G. Gil, M. S. Brown, J. L. Goldstein, and K. L. Luskey.** 1984. HMG CoA reductase: a negatively regulated gene with unusual promoter and 5′ untranslated regions. Cell **38**:275–285.
32. **Sauerbier, W., and K. Hercules.** 1978. Gene and transcription unit mapping by radiation effects. Annu. Rev. Genet. **12**:329–363.
33. **Schmike, R. T., R. J. Kaufman, F. W. Alt, and R. E. Kellems.** 1978. Gene amplification and drug resistance in cultured murine cells. Science **202**:1051–1055.
34. **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.
35. **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.
36. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
37. **Simonsen, C. C., P. C. Brown, G. F. Crouse, M. McGrogan, D. Setzer, D. Sweetser, R. J. Kaufman, and R. T. Schimke.** 1981. Gene amplification as a mechanism for drug resistance in cultured animal cells, p. 343–359. *In* T. P. Singer and R. N. ONdarza (ed.), Proceedings of the International Symposium on the Molecular Basis of Drug Action. Elsevier/North-Holland Publishing Co., New York.
38. **Young, R. A., O. Hagenbuchle, and U. Schibler.** 1981. A single mouse α-amylase gene specifies two different tissue-specific mRNAs. Cell **23**:451–458.