Independent 5'- and 3'-End Determination of Multiple Dihydrofolate Reductase Transcripts

JONG-YOUNG J. YEN AND RODNEY E. KELLEMS*

Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, Houston, Texas 77030

Received 8 May 1987/Accepted 8 July 1987

Multiple dihydrofolate reductase (*dhfr*) mRNAs, differing substantially in abundance, are produced as a result of the utilization of multiple transcription initiation sites and multiple polyadenylation sites. We have shown that *dhfr* mRNAs initiating from an upstream promoter region utilize the same collection of six polyadenylation sites and generate multiple *dhfr* mRNAs at the same relative abundance as do the mRNAs initiating from the major transcription promoter region. These results indicate that the 5' and 3' ends of *dhfr* mRNAs are independently determined. We show that the relative abundance of steady-state *dhfr* mRNAs was the same in nuclear and cytoplasmic RNA fractions. This finding makes it unlikely that differences in mRNA stability account for differences in the relative abundance of the multiple *dhfr* mRNAs in the cytoplasm. Our analysis of the *dhfr* promoter region revealed the existence of stable cytoplasmic polyadenylated transcripts complementary to the first 300 nucleotides of the *dhfr* transcripts initiating from the upstream promoter region. Therefore, the *dhfr* locus hosts two divergent and partially overlapping genes which share the same promoter region.

Transcription promoter elements and polyadenylation sites are genetic signals which determine the 5' and 3' ends of stable mRNAs. Genes which encode multiple mRNA species contain redundant copies of either or both of these types of genetic signals. For some genes, the ability to choose among multiple transcription initiation sites and multiple polyadenylation sites makes possible the production of multiple mRNAs which encode functionally distinct polypeptides. Other genes produce multiple mature mRNA species with no apparent difference in physiological function (for a review, see reference 15).

The mouse dihydrofolate reductase (*dhfr*) gene is transcribed to produce multiple mature mRNAs, which apparently encode identical polypeptide products (23). The production of these multiple mRNAs results from the utilization of multiple transcription initiation sites and multiple polyadenylation sites. Thus, structural differences in the multiple mRNAs are believed to be restricted to their 5' and 3' nontranslated regions. The multiple *dhfr* mRNAs differ substantially in abundance, and the molecular mechanisms which determine the relative abundance of the multiple *dhfr* mRNAs are not known. Two transcription promoter domains have been functionally identified on the basis of gene transfer and expression in cultured mammalian cells (17). Nuclease protection analyses have identified two transcription start regions, each of which display microheterogeneity of initiation sites (5, 17, 21). Transcription initiation sites at positions -55 and -115 (relative to the ATG codon of *dhfr*) are governed by the major *dhfr* promoter. Transcripts initiating at positions -650, -600, and -550 result from the utilization of a minor upstream promoter. The 3' ends of stable *dhfr* transcripts are determined by the utilization of numerous polyadenylation sites distributed over a 5-kilobase (kb) region at the end of the structural gene (23; A. G. Hook and R. E. Kellems, submitted for publication).

McGrogan et al. (17) reported that a double-stranded probe corresponding to the region located between 405 and

543 base pairs (bp) upstream from the ATG codon is specific for transcripts originating from the minor upstream transcription initiation region and that this probe hybridizes to only some of the multiple *dhfr* mRNAs (17). When this type of analysis was performed on several independently isolated cell lines with amplified *dhfr* genes, it appeared that the upstream probe hybridized to different *dhfr* mRNAs in different cell lines. As a result of these observations, McGrogan and co-workers concluded that a portion of the 5'-ward region of the gene is represented differentially in some but not all *dhfr* mRNAs. This conclusion implies that the selection of a transcription initiation site is coordinated with the selection of a polyadenylation site. Their studies involved the use of a double-stranded probe that would not only hybridize to the *dhfr* transcripts initiating from the upstream promoter region but would also cross-hybridize with the transcription products of an overlapping gene transcribed from the opposite strand (5, 7).

We tested this hypothesis by using strand-specific probes that specifically hybridize to *dhfr* transcripts initiating from the upstream promoter region. Transcripts from a variety of cell lines, including those with and without amplified *dhfr* genes, were examined. Our results indicate that the 5' and 3' ends of stable *dhfr* transcripts are independently determined and that transcripts initiating from the upstream promoter region utilize the same collection of polyadenylation sites with the same relative efficiency as do transcripts initiating from the major *dhfr* promoter region. We also show that the *dhfr* transcripts which initiate from the upstream promoter region overlap by approximately 300 nucleotides (nt) with a cytoplasmic polyadenylated [poly(A)⁺] RNA that is transcribed from the opposite strand.

MATERIALS AND METHODS

Cell lines and RNA preparations. Cells were maintained in Dulbecco modified Eagle medium supplemented with 10% horse serum. S180M500 (23) and 3T3R500 (22) are two independently isolated methotrexate-resistant mouse cell lines grown in medium which contains 500 μ M methotrex-

* Corresponding author.

ate. S180P (11) is a mouse sarcoma cell line from which the S180M500 cells were derived. Confluent cultures of cells were harvested from roller bottles 1 day after being fed with fresh medium. Cytoplasm and nuclei were fractionated (25), and cytoplasmic RNA was prepared from the cytoplasmic fraction by phenol-chloroform extraction as previously described (25). Nuclear RNA was prepared from freshly isolated nuclei by the guanidine isothiocyanate-CsCl cushion method (4); 4 M guanidine isothiocyanate solution was added at a ratio of 9 ml of solution to 1 ml of nuclei. Poly(A)⁺ and nonpolyadenylated [poly(A)⁻] RNAs were fractionated by oligo(dT)-cellulose column chromatography (2). The quality of RNA preparations was assessed by examination of the 28S and 18S rRNA as visualized by ethidium bromide staining following agarose gel electrophoresis. The presence of a 5- and a 2-kb band at a 2:1 ratio of signal intensities was taken as an indication of a relatively undegraded RNA preparation.

Plasmid DNA construction. pSP64RT10 was prepared by insertion of the *EcoRI*-*TaqI* fragment (Fig. 3) into the *EcoRI* and *AccI* sites of pSP64 (Promega Biotec, Madison, Wis.). In vitro transcripts derived from pSP64RT10 with SP6 RNA polymerase (New England BioLabs, Inc., Beverly, Mass.) are complementary to the *dhfr* mRNA. pSP64A1 and pSP64A9 are recombinant plasmids constructed by the insertion of a 330-bp *AvaII* fragment (Fig. 3) into the *SmaI* site of the cloning cassette of pSP64 (Promega Biotec). The *AvaII* fragment was derived from the 1-kb *EcoRI*-*TaqI* fragment, blunt-ended with the Klenow fragment of *Escherichia coli* DNA polymerase I, and subcloned into the dephosphorylated *SmaI* site of the pSP64 vector in two different orientations. The in vitro transcripts with SP6 polymerase from clones A1 and A9 were complementary to the *dhfr* mRNAs and transcripts arising from the non-*dhfr* strand, respectively. pCVSVL2 is a minigene construct (13) that contains the entire coding region (approximately 800 bp) from the *dhfr* cDNA clone pDHFR26 (3). Digestion of pCVSVL2 with *Bam*HI and *Hind*III generated a 1.1-kb fragment containing cDNA sequences and simian virus 40 sequences, which were used as the cDNA probe in Northern (RNA blot) analysis (Fig. 1B).

Probe preparation. Double-stranded DNA probes were prepared by the random primer method as described by Feinberg and Vogelstein (8). The *dhfr* cDNA fragment and the *Ava*33 DNA fragment were isolated from low-gelling-temperature agarose gels and labeled as previously described (8). Single-stranded cRNA probes were prepared by using the protocol provided by Promega Biotec after linearization of pSP64A1, pSP64A9, and pSP64RT10 with *Pvu*II. The probe for S1 nuclease protection was prepared by the following procedure: linearization of pSP64RT10 at the unique *TaqI* site, removal of 5' phosphates and replacement with radiolabeled phosphates, digestion with *EcoRI*, and isolation of the resulting 1-kb *EcoRI*-*TaqI* fragment (the RT10 probe) following electrophoretic fractionation on 5% polyacrylamide gels.

RNAse H cleavage experiment. An oligonucleotide complementary to the nucleotide sequence from -75 to -89 (Fig. 5A) in the 5' nontranslated region of the mouse *dhfr* gene was prepared and hybridized to total cytoplasmic RNA isolated from S180M500 cells. The oligonucleotide was mixed with the RNA at a molar ratio of approximately 100:1 (15 µg of oligonucleotide, 100 µg of RNA) and incubated for 1 h at 30°C with *E. coli* RNAse H (Bethesda Research Laboratories, Gaithersburg, Md.) in a 50-µl reaction mixture which contained 20 mM Tris hydrochloride (pH 7.5), 10 mM

MgCl₂, 100 mM KCl, 0.1 mM dithiothreitol, and 5% sucrose. The reaction was stopped by the addition of 50 µl of 5 M ammonium acetate and extracted once each with phenol and CHCl₃-isoamyl alcohol (25:1). Cold ethanol was added to precipitate nucleic acids. The pellet was dissolved in sterile water, and a sample was used to check the efficiency of oligonucleotide-directed RNAse H cleavage. This was done by S1 nuclease protection experiments (see below) to determine the 5' end of the treated mRNA. The rest of the sample was run on 1.5% agarose gels containing formaldehyde for the Northern blot analysis as described below.

RNA blot preparation and hybridization. The RNA was pelleted following alcohol precipitation, dissolved in a denaturation buffer containing 20 mM sodium HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid) (pH 7.0), 5 mM sodium acetate, 1 mM disodium EDTA, 50% formamide (deionized and filtered), and 2.2 M formaldehyde (deionized) and fractionated in 1.5% formaldehyde-agarose gels (19) in the presence of 0.3 mg of ethidium bromide per liter. After electrophoresis, the RNA was blotted to a positively charged nylon membrane (Zetaprobe; BioRad Laboratories, Richmond, Calif.) by capillary transfer (16) and probed with either ³²P-labeled double-stranded DNA or SP6 polymerase-transcribed ³²P-labeled single-stranded RNA. Prehybridization and hybridization were performed as previously described (20), with the addition of a prewash step using 1% sodium dodecyl sulfate-0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) as recommended by AMF/Cuno, Meriden, Conn., before the prehybridization. After hybridization, blots were washed extensively by using washing buffers described previously (25) until a suitably low background was obtained.

When cRNA probes were used, incubation was performed at 80°C in a water bath throughout the prehybridization, hybridization, and washing steps. Filters were incubated with 0.1 mg of RNAse A per ml in 2× SSC at the end of the washing process to reduce nonspecific hybridization. The sizes of the mRNAs were determined by comparison with the migration of 28S and 18S rRNA or with an RNA size ladder from Bethesda Research Laboratories. The sizes of *dhfr* mRNAs averaged 200 nt longer than those reported previously (23). The previously reported sizes for the four smallest *dhfr* mRNAs are those of the corresponding cDNA sequence and do not reflect the presence of the poly(A) tail.

S1 nuclease protection. The S1 nuclease protection was conducted basically as described previously (1), except that the amount of the *TaqI* site end-labeled RT10 probe was approximately 2×10^5 cpm, and the total amount of nucleic acid was maintained at about 100 µg per reaction. The temperature for the hybridization was 46°C.

RESULTS

Model to account for the production of multiple stable transcripts from the mouse *dhfr* transcription unit. The mouse dihydrofolate reductase gene has been cloned (6) and the transcription initiation sites (5, 17, 21) and polyadenylation sites (24; Hook and Kellems, submitted) have been mapped (Fig. 1A). At the 5' end of the gene, transcription initiates from two regions, one near -55 and the other near -550 nt (numbered upstream from the ATG codon). Polyadenylation occurs at any one of six polyadenylation sites (not seven as previously believed; see below) which are distributed over a 5-kb region at the 3' end of the gene. As will be seen below, Northern blotting analysis can distinguish the multiple *dhfr*

A. Dihydrofolate Reductase Gene

B. Proposed *dhfr* Transcript Groups

FIG. 1. Model to account for the production of multiple stable transcription products of the mouse dihydrofolate reductase gene. (A) The structure of the *dhfr* gene is shown schematically. |, Six exons; □, 5' and 3' nontranslated regions. The bidirectional promoter region and the polyadenylation region are enlarged. →, Positions and orientations of transcription initiation sites; ↓, positions of polyadenylation sites (5, 17, 21, 23). The thickness of the arrows indicates the frequency with which the site is utilized (5). (B) The proposed structures for the 12 *dhfr* transcripts are based on the independent determination of 5' and 3' ends. The multiple *dhfr* transcripts are shown in two groups, based on the transcription initiation site from which the mRNA originated. The sizes of transcripts are indicated in kilobases, and the 5' cap site and 3' poly(A) tail are indicated as Cap and AAAn, respectively. □, Translated regions; ■, the two major mRNA species from each group. The location of the *dhfr* cDNA probe is indicated at the bottom.

transcripts which originate from two promoter regions and are cleaved and polyadenylated at six different polyadenylation sites. Microheterogeneity exists at each transcription initiation site and each polyadenylation site. In contrast to what has been suggested previously (17), we propose that the 5' and 3' ends are independently determined. According to our model, 12 distinct poly(A)⁺ mRNAs should be encoded by the mouse *dhfr* gene, and they can be grouped into two categories (Fig. 1B). The RNA species in group 1 are those which were transcribed from the major promoter region and initiated predominantly at position -55. The RNA species in group 2 were transcribed from a minor upstream promoter region and initiated from a position of approximately -550 nt. The members within each group differ predominantly at their 3' ends. Two RNA species in each group predominated, presumably due to bias in polyadenylation site utilization. The abundance of group 1 transcripts was about 30-fold higher than that of group 2 transcripts, presumably due to differences in the transcrip-

tional activity of the promoter from which they initiated. This difference of 30-fold was estimated from the relative abundance of the 5' ends of *dhfr* transcripts as determined by S1 nuclease protection experiments (21; unpublished observations) and from the ratio of signal intensities corresponding to the 1,800- and 2,300-nt RNA bands on Northern blots probed with *dhfr* cDNA. As explained below, these two RNAs were transcribed from the major and minor promoter regions, respectively, and correspond to *dhfr* mRNAs which utilize polyadenylation site 4.

Overproduction of authentic *dhfr* mRNAs in methotrexate-resistant cell lines. When RNA blots were probed with a *dhfr* cDNA which has the same length of homology with all transcripts from both groups (Fig. 1B), a composite RNA pattern was expected. However, only seven species were easily identified, due largely to the very low abundance of most group 2 mRNA species and partly to the comigration of RNAs from groups 1 and 2. These seven *dhfr* mRNA species were present in the poly(A)⁺ fraction from both the parental

FIG. 2. Blot hybridization analysis of RNA from S180M500 and S180P cell lines. RNA was isolated from either S180M500 cells or S180P cells, electrophoretically fractionated on agarose gels, transferred to nylon membranes, and probed with radiolabeled *dhfr* cDNA as described in Materials and Methods. For lanes 1 to 6, the indicated amounts (in micrograms) of cytoplasmic RNA from S180M500 cells (lanes 1, 3, and 5) or S180P cells (lanes 2, 4, and 6) were analyzed. The autoradiographs represent overnight exposures. Lanes 7 and 8 show a 1-week exposure of lanes 5 and 6. Lanes 9 and 10 show an overnight exposure of an analysis using 10 µg of poly(A)⁺ and 50 µg of poly(A)⁻ cytoplasmic RNA from S180M500 cells. Lanes 11 to 14 show a 3-week exposure of an analysis using 20 µg of poly(A)⁺ and 50 µg of poly(A)⁻ nuclear RNA (Nuc) and 20 µg of poly(A)⁺ and 50 µg of poly(A)⁻ cytoplasmic RNA (cyt), respectively, from S180P cells. The exposure time required for lanes 11 to 14 was approximately 400 times that for lanes 1 to 6, and several contamination spots were observed in the 3-week exposure. The sizes (in nucleotides) of the RNA species are indicated on the left.

cells and methotrexate-resistant cells (Fig. 2), indicating that the multiple *dhfr* transcripts were naturally occurring and did not result from an artifact of gene amplification. Because the probe used in these experiments has the same length of homology for all *dhfr* transcripts, the relative abundance of the multiple *dhfr* mRNAs was reflected directly by the intensity of the signal obtained following hybridization and autoradiography. We did not observe enrichment for the higher-molecular-weight *dhfr* mRNAs (i.e., 2,300-, 4,100-, and 5,600-nt mRNAs) in the poly(A)⁻ fraction as described by Gasser and Schimke (9) or in the nuclear fraction as described by Setzer et al. (24). All *dhfr* transcripts, in amplified and nonamplified cell lines, were poly(A)⁺ and slightly enriched in the cytoplasmic RNA preparations relative to the nuclear RNA preparations (Fig. 2, lanes 11 and 13; unpublished data).

Six mRNAs originated from the upstream promoter region. To visualize the *dhfr* group 2 mRNAs transcribed from the upstream promoter exclusively, a strand-specific RNA probe corresponding to the *Ava*33 fragment (Fig. 3) was prepared and used to probe Northern blots. Overnight exposure revealed two dominant mRNAs species (Fig. 4, lane 1). Longer exposure showed six RNA species which correspond in size to an increase of approximately 500 nt over those of the predicted six group 1 transcripts (Fig. 4, lane 3). The relative abundance of these transcripts was similar to that of *dhfr* group 1 mRNAs as well, suggesting that they correspond to the hypothesized *dhfr* group 2 mRNAs. Six *dhfr* RNAs of the same size were also observed when RNA from another *dhfr*-overproducing cell line,

FIG. 3. 1,015-nt *Eco*RI-*Taq*I fragment spanning the mouse *dhfr* bidirectional promoter region. The 5' ends of *dhfr* transcripts and divergent transcripts are indicated above and below the line, respectively. The thickness of the arrow indicates the approximate abundance of each transcript. The ATG codon for *dhfr* mRNAs and the putative ATG codon and the first 5' splice site of divergent transcripts are shown. The broken lines represent the first intron of the divergent gene. The positions of several restriction endonuclease sites, four highly conserved 48-bp repeats, and two clones (RT10 and *Ava*33) used in subsequent experiments are also shown.

3T3R500, was examined (Fig. 4, lanes 5 and 7). When 3T3R500 RNA was probed with *dhfr* cDNA, the same seven RNA bands shown in Fig. 2 for S180 RNA were observed (data not shown). None of the group 2 mRNAs was observed to be enriched in the nuclear fraction or poly(A)⁻ fraction, indicating that they were not splicing precursors.

Six upstream promoter transcripts shared the same 5' nontranslational sequences and were colinear with *dhfr* group

FIG. 4. Identification of *dhfr* transcripts initiating from the upstream promoter region. RNA blots were prepared by using RNA from S180M500 and 3T3R500 cells and probed with a strand-specific RNA complementary to *dhfr* transcripts initiating from the upstream promoter region. The probe was prepared by in vitro transcription of pSP64A1, a recombinant plasmid containing the *Ava*33 fragment shown in Fig. 3. Poly(A)⁺ and poly(A)⁻ cytoplasmic (cyt) RNA from S180M500 cells was examined by blot hybridization analysis (lanes 1 and 2, overnight exposure; lanes 3 and 4, 1-week exposure). Lanes: 5 to 8, analysis of nuclear (Nuc) and cytoplasmic RNA from the 3T3R500 cell line; 1, 3, 5, and 7, poly(A)⁺ RNA (20 µg); 2, 4, 6, and 8, poly(A)⁻ RNA (50 µg). The sizes (in nucleotides) of RNA species are indicated on the left.

FIG. 5. Oligonucleotide-directed RNase H cleavage of group 2 *dhfr* mRNAs. (A) Schematic illustration and localization of an oligonucleotide and the complementary sequence in group 2 *dhfr* mRNAs. The numbers refer to the nucleotide positions upstream from the *dhfr* ATG initiation codon. □, Highly conserved 48-bp repeats; →, sites of transcription initiation. (B) Identification of the 5' ends of normal and cleaved *dhfr* mRNAs. RT10 DNA (Fig. 3) was labeled at the *TaqI* site (position +6), hybridized to normal and cleaved *dhfr* RNAs, and digested with S1 nuclease. The protected fragments were electrophoretically separated on polyacrylamide gels and visualized autoradiographically. On the basis of size determination relative to ϕ X174RF *HaeIII* markers, the 5' ends of the protected fragments were assigned positions (numbers on the right) upstream of the ATG codon of the *dhfr* gene. Lanes: 1 and 2, nuclease protection products obtained from 10 μ g of cytoplasmic RNA from S180P and S180M500, respectively; 3 and 4, nuclease protection products of truncated S180M500 RNA prepared by using 5 and 1 U of RNase H, respectively. (C) Blot hybridization analysis of RNA samples. The same RNAs referred to above were electrophoretically fractionated by formaldehyde-agarose gel electrophoresis, transferred to a nylon membrane, and hybridized with a strand-specific RNA probe complementary to *dhfr* transcripts initiating from the upstream promoter region (see legend to Fig. 4). Lanes 1 to 4 contain the same RNA samples shown in panel B. The sizes of the RNA species (in nucleotides) are shown on the left.

1 mRNAs. The fact that the *Ava33* probe (Fig. 3) hybridized to six transcripts which are approximately 500 nt longer than the predicted six group 1 transcripts suggests that virtually all of the additional sequences transcribed from the upstream promoter region are retained on mature mRNAs. It also suggests that the additional 500 nt present at the 5' ends of group 2 mRNAs are colinear with sequences found on group 1 mRNAs. An RNase H cleavage experiment was designed to test for the presence of the predicted 500-nt segment on group 2 mRNAs.

RNase H has an endonuclease activity which specifically digests the RNA strand in an RNA:DNA heteroduplex. An oligonucleotide was synthesized which should direct RNase H cleavage at approximately -80 nt in *dhfr* mRNA molecules initiating from the upstream promoter region (Fig. 5A). The efficiency of RNase H cleavage was judged by the use of S1 nuclease protection analysis to determine the 5' ends of

the RNA before and after oligonucleotide-directed RNase H cleavage. For this purpose, the RT10 fragment (Fig. 3) was labeled at the *TaqI* site (position +6) and used as a probe in an S1 nuclease protection experiment to determine the 5' ends of *dhfr* mRNAs. As shown previously by Sazer and Schimke (21), when RNA from *dhfr* amplified cell lines was used, 5' ends could be mapped to positions -55, -115, and -550, each position having some microheterogeneity (Fig. 5B, lane 2). No signal could be detected under these conditions when mRNA from parental cell lines was used (lane 1). When RNA was incubated with RNase H in the presence of the oligonucleotide, the original upstream 5' ends at positions -115 and -550 disappeared and multiple new 5' ends were generated near -80, the region complementary to the oligonucleotide (Fig. 5B, lanes 3 and 4). In the absence of the oligonucleotide, RNase H alone did not affect the 5' ends of *dhfr* mRNAs, suggesting that cleavage occurred in an oligo-

FIG. 6. Simultaneous detection of the two predominant *dhfr* mRNAs transcribed from each *dhfr* promoter region. Blot hybridization of S180M500 RNA was performed by using radiolabeled RNA prepared by in vitro transcription of pSP64RT10, a recombinant plasmid which contains an insert consisting of the RT10 fragment shown in Fig. 3. The single-stranded RNA probe is specific for *dhfr* mRNA. Sizes of the RNA species (in nucleotides) are shown on the left, and the structures of these transcripts are schematically illustrated on the right (refer to Fig. 1). Lanes 1 and 2 contain 20 µg of poly(A)⁺ and 50 µg of poly(A)⁻ cytoplasmic RNA, respectively, from S180M500 cells. AAAn, Poly(A) tail.

nucleotide-directed manner (data not shown). The RNA products of the cleavage reaction were electrophoretically separated on agarose gels, blotted to a nylon membrane, and probed with a strand-specific *Ava*33 probe from pSP64A1. A 480-nt signal was identified as the 5' cleavage product of group 2 mRNAs. Thus, after RNase H digestion, the *Ava*33 probe no longer detected the six group 2 mRNAs but instead detected a single RNA of approximately 500 nt. This result suggests that the *dhfr* group 2 mRNAs share the same 5' end and differ only at their 3' ends. Furthermore, these results indicate that the group 2 mRNAs are colinear with the group 1 mRNAs and differ only by the presence of an additional 500 nt at the 5' ends.

Major groups 1 and 2 mRNAs could be visualized simultaneously and with similar intensity by using an appropriate hybridization probe. Previous studies on the mapping of the 5' ends of *dhfr* mRNA revealed that a 1-kb *Eco*RI-*Taq*I fragment (RT10; Fig. 3) does not hybridize well under normal conditions to transcripts which initiate from position -55 (21). This is believed to be due to the presence of an A+T-rich sequence within the first 60 nt. At 65°C, only transcripts initiating at positions -115 and -550 can be simultaneously mapped under standard aqueous hybridization conditions. Thus, if the independent-end-determination hypothesis for the formation of multiple *dhfr* transcripts is correct, the use of the RT10 fragment as a hybridization probe in RNA blot analysis should permit simultaneous visualization of transcripts initiating at -115 and -550 and should allow these transcription products to be visualized at comparable intensities. The predicted pattern for a short exposure time should consist of four RNA species which represent the two dominant transcripts from each group.

Hybridization with a strand-specific RNA probe prepared by in vitro transcription of pSP64RT10 to a blot of S180M500 RNA revealed the predicted pattern of four species of RNAs with sizes of 950, 1,450, 1,800 and 2,300 nt (Fig. 6). The postulated structures of the observed transcripts are shown schematically beside the autoradiograph in Fig. 6 and repre-

sent the two predominant *dhfr* mRNAs transcribed from each promoter region (Fig. 1B). The sizes and relative abundances of these RNAs fit well with the model of independent 5'- and 3'-end determination.

***dhfr* transcripts initiating from the upstream promoter overlapped by approximately 300 nt with cytoplasmic poly (A)⁺ RNAs transcribed from the opposite strand.** McGrogan et al. (17) have previously used a double-stranded probe corresponding to the region -405 to -543 to detect transcripts of approximately 920, 1,800, and 2,300 nt. They postulated that these represent *dhfr* transcripts originating from the upstream promoter region and that the transcripts utilize only three of the available six polyadenylation sites. This interpretation is inconsistent with the results shown above (Fig. 4), for which we used a *dhfr* strand-specific probe to detect six RNA species which appeared in the same relative abundance but which were approximately 500 nt longer than the six group 1 *dhfr* transcripts. We hypothesize that the transcripts detected by McGrogan et al. (17) are derived from an overlapping gene which is transcribed in the opposite direction from *dhfr* and that transcripts derived from this gene are much more abundant than are *dhfr* transcripts originating from the upstream *dhfr* promoter region. To test this hypothesis, an RNA blot hybridization analysis was performed with the RNA of S180M500 cells by using a strand-specific RNA probe complementary to the divergent transcripts. The results of this experiment (Fig. 7, lanes 3 and 4) reveal the existence of at least five divergent transcripts with sizes of 920, 1,800, 2,300, 4,000, and 6,000 nt. A double-stranded probe corresponding to this region was also shown to hybridize to these same RNA species and an extra 1,450-nt species (Fig. 7, lane 1). These data indicate that the 1,450-nt signal and part of the 2,300-nt signal in Fig.

FIG. 7. Complementary comigrating RNAs derived from the *dhfr* bidirectional promoter region. RNA blots were hybridized with either a uniformly labeled, double-stranded *Ava*33 DNA probe (lanes 1 and 2) or a uniformly labeled strand-specific riboprobe prepared by in vitro transcription of the *Ava*33 fragment in pSP64A9 (lanes 3 and 4). The location of the *Ava*33 fragment is shown in Fig. 3. The in vitro transcript of pSP64A9 specifically hybridizes to transcripts complementary to *dhfr* transcripts initiating from the upstream promoter region. Lanes: 1 and 3, 20 µg of poly(A)⁺ RNA; 2 and 4, 50 µg of poly(A)⁻ RNA from S180M500 cells. The sizes (in nucleotides) of the two major *dhfr* transcripts transcribed from the upstream promoter region are shown on the left. The sizes (in nucleotides) on the right indicate the locations of transcripts complementary to *dhfr* transcripts derived from the upstream promoter region.

7, lane 1, were *dhfr* upstream transcripts and that they were much less abundant than were divergent transcripts. Therefore, three transcripts reported previously by McGrogan et al. (17) are not *dhfr* group mRNAs but instead are most likely the three dominant species of divergent transcripts. All transcripts detected were cytoplasmic poly(A)⁺ RNAs (Fig. 7). By using a different methotrexate-resistant cell line, somewhat different results have been observed by Farnham et al. (7), who reported that the divergent transcript is a small nuclear, poly(A)⁻ RNA. In this regard, we have recently examined a number of methotrexate-resistant cell lines, as well as parental cell lines, and have shown that in each case the authentic divergent mRNA is a poly(A)⁺ cytoplasmic RNA (J.-Y. J. Yen and R. E. Kellems, manuscript in preparation).

DISCUSSION

We demonstrated that the *dhfr* mRNAs initiating from an upstream promoter region utilize the same collection of polyadenylation sites and generate multiple *dhfr* mRNAs at the same relative abundance as do the mRNAs initiating from the major transcription promoter region. These results indicate that the utilization of multiple cleavage and polyadenylation sites located at the extreme 3' end of the *dhfr* transcription unit occurs with an efficiency that is independent of the promoter region from which the transcript initiates. A contrasting conclusion was reached by McGrogan et al. (17), who interpreted their results to suggest that *dhfr* transcripts originating from the upstream promoter region utilize only a limited portion of the polyadenylation sites available to transcripts initiating from the major promoter region. Their studies involved the use of a double-stranded probe that would not only hybridize to the *dhfr* transcripts initiating from the upstream promoter region but would also cross-hybridize with the transcription products of an overlapping gene transcribed from the opposite strand (5). We showed that the signals with S180M500 cells detected by McGrogan et al. are the transcription products of an overlapping gene. We have shown elsewhere (Yen and Kellems, in preparation) that the variable sizes of these transcripts observed by McGrogan et al. (17) in different methotrexate-resistant cell lines result from variable coamplification of the neighboring gene in its intact form. Since the *dhfr* gene encodes such a large number of stable transcripts, it is not surprising that different-sized divergent transcripts fortuitously comigrate with one or more of the *dhfr* mRNAs.

The multiplicity of 5' and 3' ends of *dhfr* mRNAs has caused some confusion in determining the structural relationship among the different *dhfr* mRNAs. It was initially suggested that the seven *dhfr* RNAs observed by RNA blot analysis with *dhfr* cDNA represent RNAs which are colinear at their 5' ends and which differ exclusively at their 3' ends as a result of the utilization of seven polyadenylation sites (23, 24). The polyadenylation sites for the 950-, 1,200-, 1,400-, 1,800-, 4,100-, and 5,600-nt *dhfr* mRNAs have been mapped (23, 24; Hook and Kellems, submitted), but no polyadenylation site corresponding to 2,300-nt mRNA was observed. When using a single-stranded probe specific for upstream *dhfr* transcripts, we have repeatedly detected only six *dhfr* mRNA species. The most abundant of these is a 2,300-nt RNA, which we postulate corresponds to a transcript which initiates from the upstream promoter region and is cleaved and polyadenylated at polyadenylation site 4. This is presumably the same 2,300-nt RNA which is detected by hybridization with *dhfr* cDNA. The simplest interpretation

of these data is that the *dhfr* gene contains six polyadenylation sites instead of seven and that the 2,300-nt RNA corresponds to a transcript which originates from the upstream promoter region and is cleaved and polyadenylated at the same polyadenylation site as the 1,800-nt mRNA of group 1 messengers.

The molecular mechanisms which govern the relative abundance of the multiple *dhfr* mRNAs are not understood. The first three of the six polyadenylation sites which are utilized do not have the highly conserved hexanucleotide signals generally associated with polyadenylation sites (24; Hook and Kellems, submitted). The relative abundance of the multiple mRNAs is presumably governed by the efficiency with which each site is utilized. We have shown here that the relative abundance of the multiple mature *dhfr* mRNAs is the same in both nuclear and cytoplasmic RNA fractions, suggesting that the abundance of individual mRNAs is determined before these messengers are transported out of the nucleus. Thus, it is unlikely that differences in mRNA stability account for differences in the relative abundance of the multiple *dhfr* mRNAs in the cytoplasm. Furthermore, the splicing or maturation of *dhfr* mRNA must be a rapid process, because we did not observe any precursor RNAs in the nuclear fraction. Changes in the level of *dhfr* mRNAs that accompany growth stimulation, viral infection, or cell cycle regulation are accompanied by coordinate changes in all of the multiple *dhfr* mRNAs (10, 12, 14, 26). Thus, physiological parameters affecting the production of *dhfr* mRNA levels do not preferentially affect the accumulation of any one of the multiple mRNAs.

Our analysis of the amplified *dhfr* genomic locus revealed the existence of stable transcription products that are complementary to *dhfr* transcripts initiating from the upstream promoter region. The region of complementarity is approximately 300 nt. In contrast to what has been reported by Farnham et al. (7), we have shown here and elsewhere (Yen and Kellems, in preparation) that the complementary transcription products are poly(A)⁺ cytoplasmic RNAs. Mitchell et al. (18) have also observed an opposite-strand transcript derived from the 5'-flanking region of the hamster *dhfr* gene. The small nuclear poly(A)⁻ transcript observed by Farnham et al. (7) is presumably the result of a gene amplification anomaly which occurs in the 5'-flanking region of the *dhfr* gene in the methotrexate-resistant cell lines they used (Yen and Kellems, in preparation). The *dhfr* promoter region is G+C rich, contains four highly conserved 48-bp repeats and multiple SP1-binding sites, and serves as the transcription initiation site for multiple *dhfr* mRNAs and a collection of divergent transcripts. The divergent transcripts are complementary to approximately the first 300 nt of *dhfr* transcripts initiating from the upstream promoter region. Thus, the *dhfr* locus hosts two divergent and partially overlapping genes which share the same promoter region. The functional basis for this organizational arrangement of transcription units is not yet understood.

ACKNOWLEDGMENTS

The assistance of Diane M. Gentry in mammalian cell culture is gratefully acknowledged. We thank A. Craig Chinault, Diane E. Ingolia, and Ming-Jer Tsai for their critical comments on this work and Debra P. Jankel for her help in preparing the manuscript.

This research was supported by Public Health Service grant GM30204 and Research Career Development Award CA00828. Methotrexate was supplied by the Drug Synthesis and Chemistry Branch, Division of Cancer Treatment, National Cancer Institute.

LITERATURE CITED

1. 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.
2. Bullock, D. W., S. L. C. Woo, and B. W. O'Malley. 1976. Uteroglobin messenger RNA: translation in vitro. *Biol. Reprod.* 15:435-443.
3. 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.
4. Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18:5294-5299.
5. Crouse, G. F., E. J. Leys, R. N. McEwan, E. G. Frayne, and R. E. Kellems. 1985. Analysis of the mouse *dhfr* promoter region: existence of a divergently transcribed gene. *Mol. Cell. Biol.* 5:1847-1858.
6. 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.
7. Farnham, P. J., J. M. Abrams, and R. T. Schimke. 1985. Opposite-strand RNAs from the 5' flanking region of the mouse dihydrofolate reductase gene. *Proc. Natl. Acad. Sci. USA* 82:3978-3982.
8. Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6-13.
9. Gasser, C. S., and R. T. Schimke. 1986. Cell cycle regulation of transfected murine dihydrofolate reductase genes. *J. Biol. Chem.* 261:6938-6946.
10. Gudewicz, T. M., V. B. Morhenn, and R. E. Kellems. 1981. The effect of polyoma virus, serum factors, and dibutyryl cyclic AMP on dihydrofolate reductase synthesis, and the entry of quiescent cells into S phase. *J. Cell. Physiol.* 108:1-8.
11. Hakala, M. T., S. F. Zakrzewski, and C. A. Nichol. 1961. Relation of folic acid reductase to amethopterin resistance in cultured mammalian cells. *J. Biol. Chem.* 236:952-958.
12. Johnson, L. F., C. L. Fuhrman, and L. M. Weidemann. 1978. Regulation of dihydrofolate reductase gene expression in mouse fibroblasts during the transition from resting to growing state. *J. Cell. Physiol.* 97:397-406.
13. Kaufman, R. J., and P. A. Sharp. 1982. Construction of a modular dihydrofolate reductase cDNA gene: analysis of signals utilized for efficient expression. *Mol. Cell. Biol.* 2:1304-1319.
14. Kellems, R. E., V. B. Morhenn, E. A. Pfendt, F. W. Alt, and R. T. Schimke. 1979. Polyoma virus and cyclic AMP-mediated control of dihydrofolate reductase mRNA abundance in methotrexate-resistant mouse fibroblasts. *J. Biol. Chem.* 255:309-318.
15. Leff, S. E., and M. G. Rosenfeld. 1986. Complex transcriptional units: diversity in gene expression by alternative RNA processing. *Annu. Rev. Biochem.* 55:1091-1117.
16. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual, p. 383-386. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. 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.
18. Mitchell, P. J., A. M. Carothers, J. H. Han, J. D. Harding, E. Kas, L. Venolia, and L. A. Chasin. 1986. Multiple transcription start sites, DNase I-hypersensitive sites, and an opposite-strand exon in the 5' region of the CHO *dhfr* gene. *Mol. Cell. Biol.* 6:425-440.
19. Rave, N., R. Crkvenjakov, and H. Boedtker. 1979. Identification of procollagen mRNAs transferred to diazobenzyloxymethyl paper from formaldehyde agarose gels. *Nucleic Acids Res.* 6:3559-3567.
20. Reed, K. C., and D. A. Mann. 1985. Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res.* 13:7207-7221.
21. Sazer, S., and R. T. Schimke. 1986. A re-examination of the 5' termini of mouse dihydrofolate reductase RNA. *J. Biol. Chem.* 261:4685-4690.
22. 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.
23. 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.
24. 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.
25. Yeung, C.-Y., E. G. Frayne, M. R. Al-Ubaidi, A. G. Hook, D. E. Ingolia, C. Bobonis, D. A. Wright, and R. E. Kellems. 1983. Amplification and molecular cloning of murine adenosine deaminase gene sequences. *J. Biol. Chem.* 258:15179-15185.
26. Yoder, S. S., B. L. Robberson, E. J. Leys, A. G. Hook, M. R. Al-Ubaidi, C.-Y. Yeung, R. E. Kellems, and S. M. Berget. 1983. Control of cellular gene expression during adenovirus infection: induction and shut-off of dihydrofolate reductase gene expression by adenovirus type 2. *Mol. Cell. Biol.* 3:819-828.