Volume 16 Number 14 1988 Nucleic Acids Research

Analysis of gene expression using episomal mouse dihydrofolate reductase minigenes

Gray F.Crouse, Linda A.Stivaletta and Martin L.Smith

Department of Biology, Emory University, Atlanta, GA 30322, USA

Received January 20, 1988; Revised and Accepted June 4, 1988

ABSTRACT

We have constructed a plasmid encoding a mouse dihydrofolate reductase (*dhfr*) minigene which produces *dhfr* transcripts with all of the 5′ and 3′ ends observed from the chromosomal mouse *dhfr* gene. The minigene contains 5′ flanking regions, all *dhfr* coding sequences, one intervening sequence, 11.5 kb of 3′ flanking regions beyond the termination codon, an *E. coli* plasmid origin of replication and antibiotic resistance, and an SV40 minimal origin of replication; the total size is 17.2 kb. When transfected into cells constitutively producing a temperature sensitive SV40 T antigen, the plasmid minigene replicates at the permissive temperature, but fails to replicate at the nonpermissive temperature. Therefore, transcription can be observed in the presence or absence of minigene replication. In addition, a stable divergently transcribed RNA is produced from the *dhfr* minigene promoter region, with the same 5′ ends that are seen in the chromosomal divergently transcribed gene. We show that deletion of the sole remaining intron of the *dhfr* minigene significantly lowers the amount of *dhfr* transcript produced but does not affect the amount of divergent transcript. The promoter region for these transcripts contains four 48 bp repeats; reducing the number of these repeats lowers the amount of both *dhfr* and divergent transcripts produced from the minigene.

INTRODUCTION

In the past few years, transcription of the mouse dihydrofolate reductase (*dhfr*) gene has been extensively characterized (1-7). The *dhfr* coding sequence consists of 558 nucleotides, spanning 6 exons spread over a distance of 35 kilobases (kb) (8). The promoter region is quite complex, with major starts at -55 and -115 nucleotides from the ATG, and more minor starts in the vicinity of -300 and -500 nucleotides from the ATG (1,4,5). A distinctive feature of the promoter region is the presence of four 48 bp repeats, from -84 to -272 nucleotides from the ATG. These repeats are very GC-rich and similar sequences are present in 2 1/2 copies in CHO and 1 copy in human (8-10). In addition, from this same promoter region another gene is transcribed in the opposite orientation (1). Although there is a report to the contrary (2), the weight of evidence, both in mouse ((1), Crouse, Linton, and Selby, unpublished; Yen and Kellems, see (11)) and in CHO cells (10), suggests that the transcript codes for a cytoplasmic, polyadenylated mRNA, approximately 4 kb in size. The 3′ ends of *dhfr* mRNA are also complex, with at least 6 different mRNA sizes, due to 3′ untranslated regions ranging in size from 80 to nearly 5000 nucleotides (6,7,11). Transcription termination appears to occur approximately 1 kb beyond the last polyadenylation site (3).

Although the basic features of *dhfr* transcription are now clear, delineation of the transcripts cannot reveal the sequences governing transcription initiation and processing. In order to understand how *dhfr* transcription is regulated, one needs to be able to construct mutant genes and assay transcription of those mutants. Because we are interested both in transcription of *dhfr*

and the divergent gene, and in 3′ processing and termination, and because regulation of some other housekeeping genes has been found to involve coding sequences (12-14), a simple promoter assay using any sort of fusion gene is inappropriate. The large size of the *dhfr* gene poses severe technical obstacles to the use of a full length gene and its subsequent mutagenesis. We showed earlier that a minigene containing 5′ flanking sequences, all of the coding sequences, and either no, or one, intervening sequence, produced dihydrofolate reductase upon transfection into tissue culture cells and could be amplified by selection with a *dhfr* inhibitor, methotrexate (15). Several disadvantages of that system were the need to use stable transfection, the poor efficiency of transcription, the lack of 3′ flanking sequences in the minigene which are normally transcribed in the gene, and the inability of those constructs to express a stable transcript from the divergent promoter. In addition, as with most stable transfection systems which depend upon integration of the transfected DNA into the genome, the chromosomal location of the transfected DNA and its structure were uncontrollable. Another major difficulty with studying transcription of the *dhfr* gene is that the promoter appears to be very weak. One can estimate that the *dhfr* gene is normally transcribed approximately once every 50 min (16). That low level of transcription, less than 1/100 the level of the β-globin gene (16), is too low to make feasible the use of non-replicating transient assay systems. With any type of replicating system, the additional sequences which must be added to the plasmid containing the gene can potentially influence the transcription of the gene. For example, enhancer sequences and promoters are associated with the origin regions of SV40, bovine papilloma virus, and Epstein-Barr Virus (17-19).

We describe here enlarged *dhfr* minigenes which overcome the difficulties mentioned above. These minigenes have 106 bp of SV40 DNA containing a minimal origin of replication with no enhancer sequences and only minimal promoter activity. The host for these minigenes is a cell line producing a temperature sensitive T antigen so that the replication of the minigenes can be controlled. The resulting minigenes produce enough RNA to be assayed from one 100 cm dish of transfected cells, can be assayed in the absence of minigene replication by shifting transfected cells to the non-permissive temperature, produce correct 5′ ends of both *dhfr* and divergent transcripts, and produce all of the proper 3′ polyadenylated species assayed thus far. We show, in addition, the importance of the sole remaining intervening sequence on *dhfr* mRNA formation, and the quantitative effects of the 48 bp repeats on transcription in both directions from the promoter region.

MATERIALS AND METHODS

Cell culture and DNA transfection.

The S180-M500 cell line used as our standard for comparison in these studies is a mouse sarcoma cell line that contains approximately 1000 times more *dhfr* genes than the parent S180 cells (6). S180-M500 cells were grown in Eagles minimal essential medium containing Hank's balanced salt solution, 9% fetal calf serum, and 500 μM MTX. The other mouse cell line with amplified *dhfr* genes, used as a standard, 3T3-R500 (20) was grown in Dulbecco's modified Eagle medium (4.5g glucose/L) with 9% fetal calf serum and 500 μM MTX.

The cells used as recipients for most of the transfection experiments were COSts2 and were provided by R. Tjian. These cells constitutively produce a temperature sensitive SV40 T antigen (21). COS1 cells were used for some experiments (22). Both the COSts2 and COS1 cells were maintained in Dulbecco's modified Eagle medium (4.5g glucose/L) with 9% fetal calf serum.

Cells were transfected with calcium phosphate precipitates of cesium chloride-gradient purified DNA (23,24). One day prior to transfection, cells were seeded at 15-20% confluence in

100mm dishes. On the day of transfection, cells were treated with a calcium phosphate precipitate of 15-20 μg DNA for 4 hours, glycerol shocked at 37°C, and harvested 2-3 days later.

Construction of replicating minigenes.

The original *dhfr* minigene plasmid, p*dhfr*3.2, was constructed as previously (15); subsequently, 10.7 kb of 3′ flanking genomic sequences were incorporated by means of plasmid/phage recombination (Crouse, G.F. and McEwan, R.N., in preparation). In brief, this final recombination consisted of selecting for a recombination of p*dhfr*3.2 on a plasmid containing a *sup*F gene with the 3′ end of the *dhfr* gene contained in a Charon 4A lambdaphage. pBR327 (25) was used as the backbone of the plasmid vector in the minigene, as it is devoid of poison sequences, but was extensively modified in order to increase the number of cloning sites and to provide a screen for inserts. The resulting plasmid vector, pGC667, was 2.3 kb in size and contained a *lacZ*α region with a polylinker. The 5′ flanking region in the *dhfr* minigene was modified by adding sequences containing the SV40 polyadenylation signal and small t splice site at the *Hinc*II site downstream of the first divergent exon, at -750 bp relative to the *dhfr* ATG. These sequences were derived from the 1.6 kb *Eco*RI-*Bgl*II fragment of pRSVβ globin (26). The SV40 minimal replication origin sequences (SVori) were obtained from K. Subramanian and contain 106 bp of SV40, from sequence number 5171 to 34 (27). The SVori fragment was cloned into the construct in both orientations. Figure 1 illustrates the relative location of the parts of the minigene. In p*dhfr*14ori, the SV40 small t splice and polyadenylation sites are replaced by 500 bp of additional flanking genomic DNA. All cloning procedures were as described previously (28).

The first intron of the *dhfr* gene was removed by replacing the *Not*I-*Sst*I fragment of p*dhfr*15ori with the equivalent fragment of p*dhfr*2.9 (15) which lacks that intron. The variable numbers of copies of the 48 bp repeats in the promoter region were constructed by cloning the *Not*I-*Sst*I fragment of p*dhfr*15ori into πAN7 and then excising various restriction fragments from the promoter region. πAN7 was chosen due to its small size and consequent lack of restriction sites. Constructs with 3 repeats were made by excision of unwanted sequences with *Fok*I; constructs with 2 repeats by excision with *Eco*0109, and 1 repeat by excision with *Sty*I. The resulting promoter regions were then cloned back into p*dhfr*15ori.

RNA analyses: protection assays.

RNA and DNA were isolated by harvesting and homogenizing cells in guanidinium thiocyanate and purified by centrifugation on CsCl block gradients as previously described (15). RNA pellets were dissolved in 10 mM Tris pH 7.5, 1% SDS, 5 mM EDTA, extracted with one volume 4:1 chloroform: butanol, precipitated with ethanol, dissolved in water and measured by OD 260. DNA was quantitated by a fluorescence assay with a Hoeffer TKO-100 mini-fluorometer (29).

RNA probes used in protection assays were uniformly labeled with ^{32}P-α-CTP by *in vitro* transcription of linearized templates with the appropriate RNA polymerase (30). Reaction mixtures were treated with DNase, then phenol chloroform extracted, and the RNA was separated from unincorporated nucleotides by 2 successive precipitations in 2M ammonium acetate.

3′ end labeled DNA probes were prepared as follows (31): 1 μg DNA (restricted to excise the probe fragment) was incubated with 1 unit (New England Biolabs) T4 DNA polymerase for 2 min (37°C) in 33 mM Tris-acetate pH 7.9, 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM DTT and 100 μg/ml BSA in a total volume of 20 μl. The recessed 3′ strand was then resynthesized by adding 1 μl of 2 mM dATP, 2 mM dGTP, 2 mM dTTP and 50 μCi (5 μl) of ^{32}P-α-dCTP, and incubating for 15 min at 37°C; 1 μl of 2 mM dCTP was added and incubation continued for 10 min longer. Unincorporated nucleotides were separated on a Bio-Gel P60 (Bio-Rad) column. The eluate was precipitated with ethanol and run on a 0.8% Seaplaque

(FMC) agarose gel to separate the probe from the vector, and the probe was extracted from the gel by freezing in phenol and phenol-chloroform extracting (32).

Total cellular RNA was hybridized overnight, at the appropriate temperature, to either RNA (200,000 cpm/reaction) or DNA (50,000 cpm/reaction) probes, exactly as described earlier (1). 11 μg rat RNA was added to the hybridization reaction when less than 10 μg of total RNA were used. After hybridization, RNA probes were digested by adding 300 μl cold RNase buffer (10 mM Tris pH 7.5, 5 mM EDTA, 300 mM NaCl, 40 μg/ml RNase A, 600 units/ml RNase T1) and incubating at either 30°C or 37°C for 1 h, followed by treatment with 20 μl 10% SDS and 50 μg proteinase K for 15 min at 37°C. The samples were then phenol-chloroform extracted, 10 μg of glycogen added, and precipitated with ethanol. Pellets were dissolved in 8M urea, boiled 3 min, and electrophoresed on 6% acrylamide-urea gels. Hybridizations with DNA probes were treated with S1-nuclease as described previously (1). Quantitative estimates of differences between bands on the same gel were made by comparing exposures of different lengths, making the assumption that the length of time required to produce a given level of density on the film is linearly related to the relative amount of radioactivity in the band.

DNA analyses.

0.8% agarose gels for Southern analyses were depurinated, denatured and neutralized prior to horizontal electroblotting to a nylon filter (33,34). Filters were baked for 2 h in an 80°C vacuum oven, washed for 30 min at 65°C in 0.1xSSC, 0.5% SDS, and prehybridized in 10 ml hybridization buffer (50% formamide, 0.72 M NaCl, 0.04 M sodium phosphate pH 7.0, 0.04 M EDTA, 1% SDS, 0.5% Carnation non-fat powdered milk) at 42°C for several hours (35). DNA probes were uniformly labeled with ^{32}P-α-dCTP (36). $5x10^6$ cpm probe was then added to 10 ml fresh hybridization buffer and incubated at 42°C for 24 h. Hybridized filters were washed twice with 2xSSC, 0.1% SDS at room temperature, and once in 0.1xSSC, 0.1% SDS for 30 min at 50°C.

RESULTS

Construction of a replicating *dhfr* minigene.

The structure of p*dhfr*15ori is shown in Fig.1 and its construction outlined in Materials and Methods. There are a number of important features in its construction. Most significant is the presence of 10,700 bp of 3′ flanking sequences which were not present in earlier minigene constructions (15). This material was added using plasmid/phage homologous recombination (Crouse, G.F. and McEwan, R.N., in preparation) and was incorporated into the construct in order to have all the poly(A) sites and the previously mapped transcription termination site (3), as well as other potentially important 3′ flanking sequences.

Because of the existence of a divergently transcribed gene originating from the *dhfr* promoter region, the 5′ flanking sequences in the minigene were also modified. The 5′ flanking region containing the initiation points for all of the identified *dhfr* transcripts and the entire presumptive first exon of the divergent transcript was kept intact. The DNA 5′ (relative to *dhfr*) to a *Hinc*II site within the first intron of the divergent transcript was replaced with the SV40 small t splice site and polyadenylation site present in the pSV expression constructions (26). The design of the construction was such that if there was transcription through the first divergent exon, the product could be polyadenylated at the SV40 site and spliced using either the small t splice sites or the first exon donor and small t accepter splice sites.

In order that the plasmid could be replicated in mammalian cells, it was essential that the bacterial plasmid vector have no sequences poisonous to replication in mammalian cells (37). This was accomplished by using a derivative of pBR327, which lacks all known "poison" se-

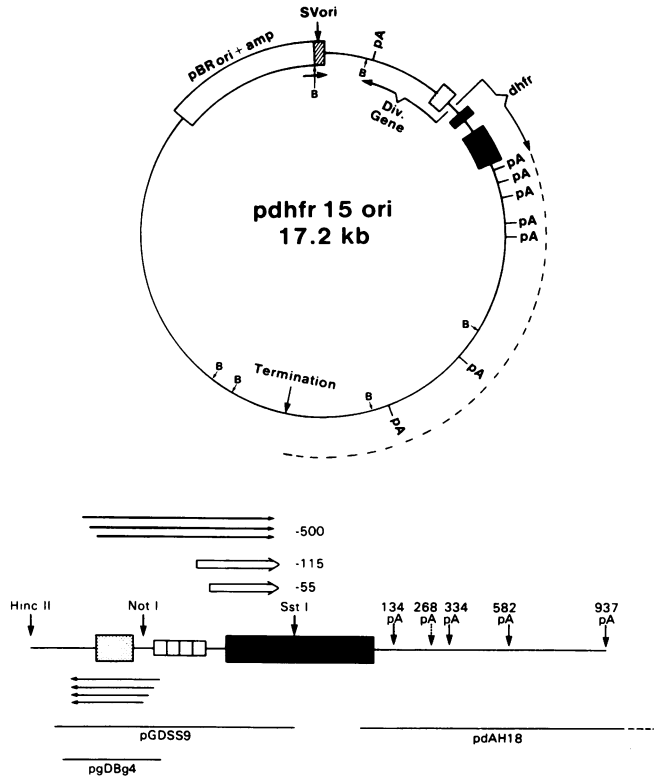

Figure 1. Structure of the *dhfr* minigene p*dhfr*15ori. *Dhfr* cDNA coding sequences are represented by black boxes separated by the first intervening sequence. The added *dhfr* 3′ flanking region extends to the *Xho*I site at the junction of the *E. coli* plasmid sequences, 10.7 kb past the *dhfr* translation termination site. The plasmid sequences, containing a ColE1 origin of replication and a β-lactamase gene, are indicated by an open box. The dotted box represents the first exon of the divergent gene, which is followed by a small t splice and polyadenylation sequence derived from SV40. The arrows indicate direction of transcription, and pA designates a polyadenylation site. The previously identified *dhfr* termination region (3) is designated with an arrow. The minimal SV40 origin is designated with the hatched box, and the orientation of the fragment is indicated by the arrow. *Bam*HI sites are denoted by a B. Details of the construction are given in Materials and Methods. Below p*dhfr*15ori is shown a portion of the minigene, lacking exon I of *dhfr*, in more detail. The four small open boxes between the *dhfr* and divergent exons represent the four 48 bp repeats. The location of various probe fragments are given at the bottom of the figure. The horizontal arrows represent transcript initiations. The numbers above the polyadenylation sites give the expected length of that mRNA protected by the pdAH18 probe.

quences. In addition, a mammalian origin of replication was needed and this was supplied by inserting an SV40 origin of replication. However, the complete SV40 origin region contains an enhancer region, promoters in both directions, and in COS cells can produce extremely high copy numbers of plasmid DNA. In order to reduce the copy number of the vector as well as

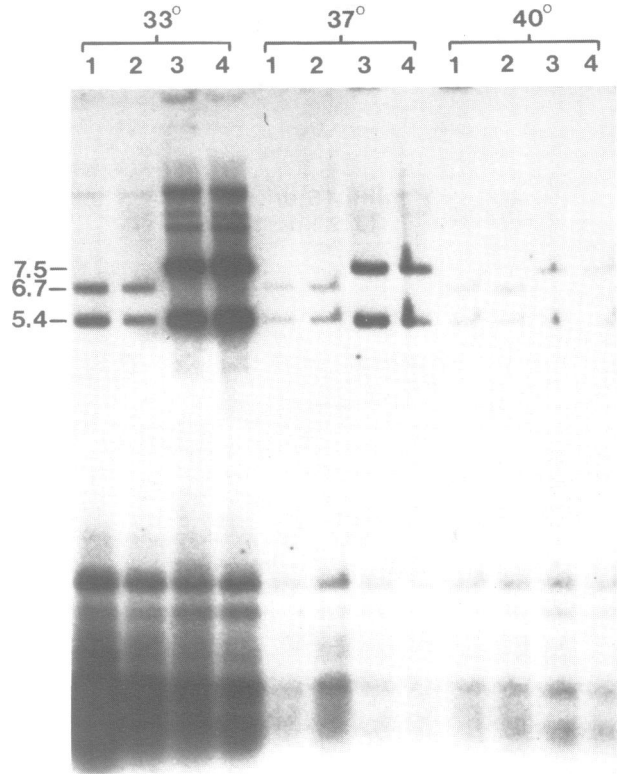

Figure 2. Southern analysis of minigene replication in COSts cells. COSts cells were transfected as described with 20 μg p*dhfr*15ori DNA per dish and maintained at either 33°C, 37°C, or 40°C for 2 days. For each sample, 2 dishes were pooled to harvest cells and isolate DNA. 8 μg of DNA were restricted with *Dpn*I and *Bam*HI, after which the DNA concentration was measured, and 5 μg of digested DNA run per lane on a 0.8% agarose gel. The gel was treated and electroblotted as described in Materials and Methods, and hybridized to a uniformly labeled DNA plasmid containing *dhfr* coding sequences. Lanes 1 and 2 represent DNA from cells transfected with p*dhfr*15ori and lanes 3 and 4 transfected with p*dhfr*15oric, at each temperature. The bands of 5.4 kb and either 6.7 or 7.5 kb were the only *Bam*HI fragments of the plasmids with homology to the probe. The small bands at the bottom of the gel are due to *Dpn*I cleavage of non-replicated DNA.

eliminate the enhancer region and at least most of the promoter region, a minimal origin of replication, containing approximately 100 bp, obtained from K. Subramanian was used (27).

Replication of the minigene in COS cells.

COS cells constitutively produce T antigen, which drives the replication of a plasmid containing an SV40 origin of replication. COSts cells (21) produce a temperature-sensitive T antigen which is active at 33°C and inactive at 40°C. To measure the replication of p*dhfr*15ori and its temperature dependence in COSts cells, we transfected p*dhfr*15ori into COSts cells at 33°C, 37°C, and 40°C and allowed the transfected cells to grow for 2 days, after which the DNA

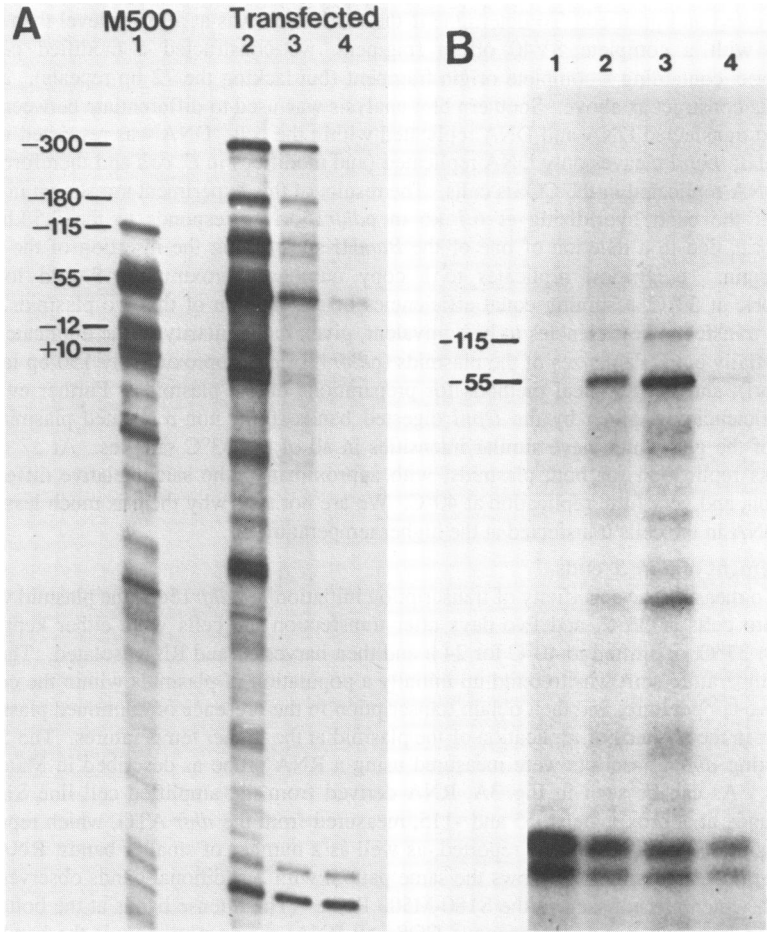

Figure 3. Protection analysis of *dhfr* 5′ ends. A. Protection of *dhfr* RNA from transfected COSts cells. COSts cells were transfected at 33°C with 15 μg p*dhfr*15ori per dish and maintained at 33°C for 2 days after transfection, then either kept at 33°C or shifted to 37°C or 40°C for 24 hrs. and harvested. RNA from approximately 30 dishes at each temperature was isolated and measured by absorbance at 260nm. Total RNA was hybridized at 54°C to a uniformly labeled RNA probe of the pGDSS9 *Stu*I-*Sst*I 0.9kb fragment. This fragment, from an intronless minigene, extends from 657 bp 5′ to the *dhfr* ATG to 261 bp 3′ of the ATG. Samples were then treated with RNase and run on a urea polyacrylamide gel as described, and sizes indicate bp position relative to *dhfr* ATG. These, and all other protection experiments, were measured relative to a 123 bp ladder (BRL). Lane 1 represents 0.1 μg of S180-M500 total RNA hybridized, lanes 2, 3, and 4, 80 μg of RNA from cells transfected and shifted to either 33°C, 37°C, or 40°C respectively. Lanes (not shown) with RNA from nontransfected COSts cells or rat RNA showed no significant hybridization. B. Protection of *dhfr* RNA from transfected COSI cells. COSI cells were transfected with 20 μg of DNA per dish at 33°. After 3 days, RNA was harvested and 25 μg hybridized as above. Lane 1 is RNA from untransfected COSI cells; Lane 2 is RNA from cells transfected with p*dhfr*15ori; Lane 3 is RNA from cells tranfected with p*dhfr*15ori and incubated for 2 days at 33° and for the final day at 40°; Lane 4 is RNA from cells transfected with p*dhfr*15, which lacks the SV40 ori.

was harvested and purified. To determine if the replication was at a lower level than would be obtained with a complete SV40 origin fragment, we constructed a modified p*dhfr*15ori, p*dhfr*15oric, containing a complete origin fragment (but lacking the 72 bp repeats), and transfected this construct as above. Southern blot analysis was used to differentiate between the non-replicated transfected DNA and DNA replicated within the cell. DNA was restricted with *Dpn*I and *Bam*HI; *Dpn*I cleaves only DNA replicated (and modified) in *E. coli* and therefore does not cleave DNA replicated in the COSts cells. The results of this experiment are shown in Figure 2. Note that the band hybridizing at 6.7 kb in p*dhfr*15ori corresponds to the 7.5 kb band in p*dhfr*15oric, due to a deletion of one of the *Bam*HI sites during the insertion of the complete SV40 origin. p*dhfr*15ori replicates to a copy number approximately 8 fold lower than p*dhfr*15oric at 33°C, assuming equal efficiencies of transfection of the two plasmids. We believe the transfection efficiencies to be equivalent, given the similarity of the duplicate samples, the essentially equivalent sizes of the plasmids (p*dhfr*15oric is approximately 150 bp larger than p*dhfr*15ori), and the identical methods for preparations of the plasmids. Further evidence of equal efficiencies is given by the *Dpn*I digested bands (from non-replicated plasmids) at the bottom of the gel, which have similar intensities in all of the 33°C samples. At 37°C there is much less replication for both plasmids, with approximately the same relative differences in replication, and very little replication at 40°C. We are not sure why there is much less *Dpn*I digested DNA in the cells transfected at the higher temperatures.

Measurement of *dhfr* 5′ ends.

To measure the specificity of transcription initiation in p*dhfr*15ori, the plasmid was transfected into cells at 33°C, and two days after transfection the cells were either kept at 33°C, shifted to 37°C, or shifted to 40°C for 24 h and then harvested and RNA isolated. The purpose of the temperature shift was to build up initially a population of plasmids within the cells at the permissive temperature, and then obtain transcription in the presence of continued plasmid replication, or in the absence of replication of the plasmid at the higher temperatures. The 5′ ends of the resulting *dhfr* transcripts were measured using a RNA probe as described in Materials and Methods. As can be seen in Fig. 3A, RNA derived from the amplified cell line S180-M500 shows bands at approximately -55 and -115, measured from the *dhfr* ATG, which represent the major *dhfr* transcripts previously reported, as well as a number of smaller bands. RNA derived from the transfected minigene shows the same pattern with 2 additional bands observed, at -180 and -300, which are not seen in the S180-M500 RNA. (The intense bands at the bottom of the transfected lanes are due to endogenous COS cell RNA; we do not know if the bands smaller than -55 represent real initiations, or artifactual cleavage in the probe). In addition, there is much less *dhfr* RNA observed at the higher temperatures. In order to determine if there might be differential *dhfr* RNA stability at the two temperatures, COS1 cells, which constitutively produce a normal T antigen in the same cellular background as COSts cells, were transfected with p*dhfr*15ori. As shown in Fig. 3B, there is no less *dhfr* RNA at 40°C than at 33°C; in fact there appears to be somewhat more *dhfr* RNA at 40°C. Also shown in Fig. 3B is the fact that if the SV40 origin sequence is removed from p*dhfr*15ori, there is very little *dhfr* RNA produced in COS cells.

To determine whether the addition of the small t splice site and SV40 polyadenylation site upstream of the *dhfr* promoter region had contributed to the presence of the extra bands at -180 and -300, a plasmid lacking this region, p*dhfr*14ori, was transfected into COSts cells and the resulting 5′ ends compared to those from p*dhfr*15ori. The pattern of the 5′ ends produced was the same from both minigenes (results not shown).

Effect of temperature shifts on plasmid content and transcription.

The results of the experiment shown in Fig. 3 were somewhat surprising, as they showed

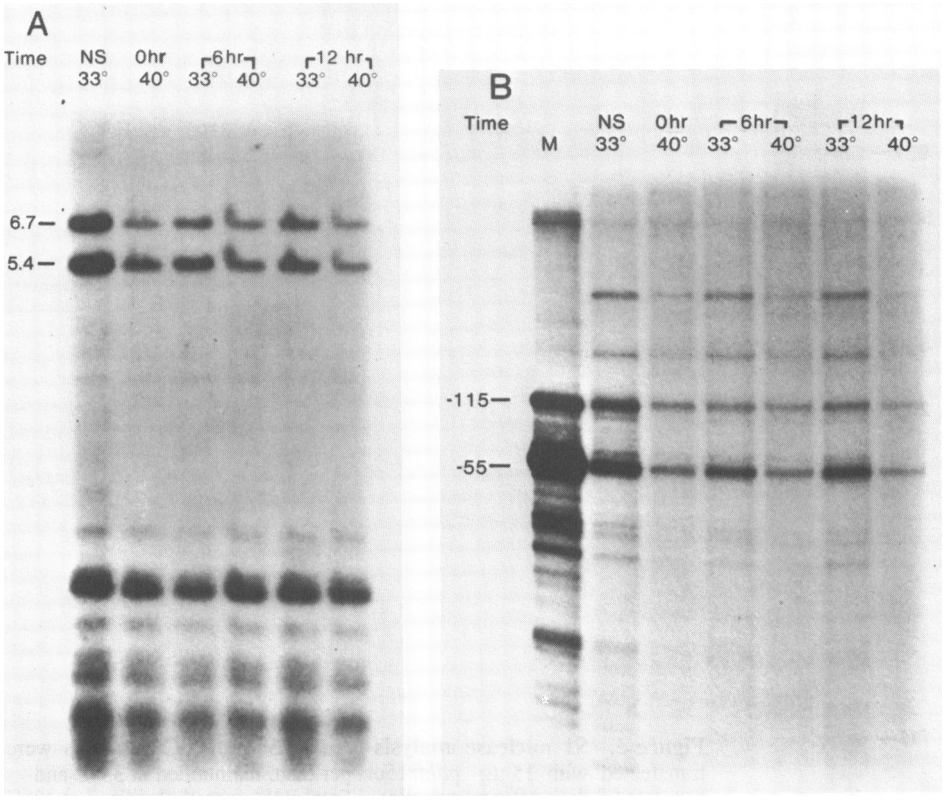

Figure 4. Effect of temperature shift on plasmid content and transcription. COSts cells were transfected with 15 μg p*dhfr*15ori per dish and maintained at 33°C. After 2 days, one set of dishes was kept at 33°C (not shifted [NS]) while the remaining dishes were shifted to 40°C for 15 hrs. At this point (Ohr., 40°C), 4 dishes were harvested, and half of the remaining dishes of cells were shifted back to 33°C. 4 dishes of cells from each temperature (33°C and 40°C) were harvested 6 hrs. and 12 hrs. later. A. DNA analysis. Southern analysis of the DNA was performed exactly as described in Fig. 2. B. RNA analysis. Protection analysis of transfected RNA hybridized to the 5′ RNA probe was as described in Fig. 3. Lane M represents 0.1 μg S180-M500 RNA; 80 μg of RNA from transfected cells was hybridized for each time point. Sizes indicate bp from the *dhfr* ATG.

a large difference in *dhfr* transcripts between cells held at 33°C for the final 24 h compared to cells shifted to 40°C for the final 24 h of transfection. Those results suggested that there was either a large increase in plasmid copy number during the third day of transfection, that there was a large loss of plasmids upon a shift to 40°C, or that there was a correlation between replication and transcription of the *dhfr* gene. We tested these possibilities by performing the experiment shown in Fig. 4. Cells were transfected and grown at 33°C (the permissive temperature) for 48 h and then shifted to 40°C for 12 h. Some cells were then shifted back to the permissive temperature and harvested either 6 h or 12 h later. DNA and RNA were prepared from each set of cells and analyzed as in Fig. 2 and Fig. 3. In Fig. 4A is shown the result of the DNA anal-

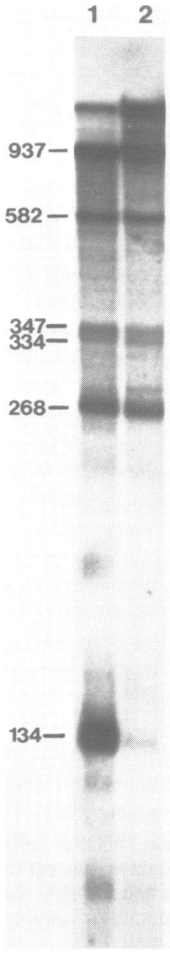

Figure 5. S1 nuclease analysis of *dhfr* 3′ ends. COSts cells were transfected with 15 μg p*dhfr*15ori per dish, maintained at 33°C and harvested 3 days after transfection. Total RNA was hybridized at 42°C to a 3′ end labeled DNA probe of the 1.8 kb insert in pdAH18. This fragment extends from 52 bp 5′ to approximately 1.75 kb 3′ of the *dhfr* stop codon. Samples were then treated with S1 nuclease and run on a urea polyacrylamide gel; sizes indicate the length of fragments in nucleotides. Lane 1 represents 0.1 μg of S180-M500 total RNA hybridized, lane 2 is 80 μg of RNA from transfected cells.

ysis. It is clear that there is a large increase in plasmid copy number during the last 24 h of transfection, as the copy number of p*dhfr*15ori is much greater in the cells kept at 33°C (NS 33°) compared to the cells kept the last 24 h at 40°C (12 h 40°). The copy number is essentially unchanged in the cells kept at 40°C, whereas there is a noticeable increase in copy number in cells shifted back to 33°C for even 6 h. The *dhfr* transcripts are related to the amount of p*dhfr*15ori present, as shown in Fig. 4B. At least most of the difference in amount of *dhfr* 5′ ends appears to be due to changes in copy number of the p*dhfr*15ori plasmid.

Measurement of *dhfr* 3′ ends.

The utilization of the first four polyadenylation sites was determined by S1 nuclease protection analysis using a 3′ end labelled DNA probe as described in Materials and Methods. As can be seen in Fig. 5, all the species observed in S180-M500 RNA are observed in the RNA from transfected cells, and with one exception, at approximately the same ratios. The major ex-

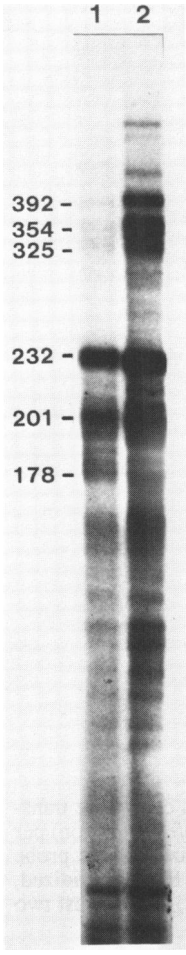

Figure 6. Protection analysis of divergent transcripts. COSts cells were transfected with 20 µg p*dhfr*15ori per dish at 33°C and harvested 2 days after transfection. Total RNA was hybridized at 54°C to a uniformly labelled RNA probe of the pgDBg4 insert, which contains sequences from -255 to -637 bp, relative to the *dhfr* ATG, treated with RNase and run on a urea polyacrylamide gel. Sizes indicate length in nucleotides. Lane 1 represents 1.0 µg of S180- M500 total RNA hybridized, lane 2 is the RNA isolated from 2 dishes of transfected cells. Rat RNA (not shown) displays no hybridization to the probe.

ception is the band at 134 nt, which is barely present in RNA from transfected cells. Possible reasons for this discrepancy are discussed below. The bands at 134, 334 & 347, 582, and 937 nt correspond to mRNAs previously mapped (6). The band at 268 nt does not correspond to a previously mapped mRNA. We are not sure whether this band represents an additional mRNA, or is the product of an artifactual cleavage, as there is a stretch of 20 bp near the cleavage site which is 90% AT. The other point of note is that there appears to be somewhat more RNA in the transfected sample extending up to, or beyond, the end of the probe compared to the S180-M500 RNA, as evidenced by a greater amount of fully protected probe.

Transcription of the divergent exon.

In order to determine if a stable transcription product were being transcribed divergently from the *dhfr* promoter region, a RNA probe spanning the region of the putative first divergent exon was used in a nuclease protection experiment. The results are shown in Fig. 6. The bands from 170 to 230 nt represent RNA transcripts which have different 5′ ends, but an identical

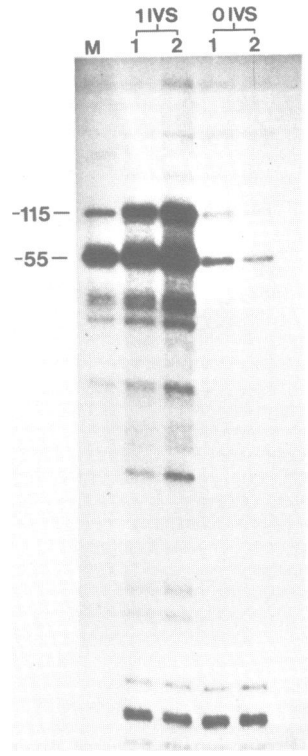

Figure 7. Quantitation of p*dhfr*15ori and p*dhfr*14.7ori transcripts. COSts cells were transfected with 15 μg of either p*dhfr*15ori or p*dhfr*14.7ori (lacking precisely the *dhfr* first intron) per dish at 33°C, and harvested 3 days after transfection. The RNA was hybridized to the probe pGDSS9, as described for Fig. 3. Lane M represents 0.5 μg of 3T3-R500 total RNA hybridized, the next two lanes are 80 μg RNA of two different preparations of p*dhfr*15ori, and the last two lanes are 80 μg RNA of two different preparations of p*dhfr*14.7ori.

splice point at the 3′ end of exon 1, as previously demonstrated (1). The strong bands at approximately 325, 354, and 392 nt in the transfected RNA are only faintly seen in S180-M500 RNA. These bands in S180-M500 RNA, as previously shown (1), represent RNA with 5′ ends identical to the set of lower bands, but which is unspliced. Thus we believe that the higher molecular weight bands in the transfected RNA also represent RNA that does not use the first exon splice site. This RNA could either be unspliced, or could use both the donor and accepter small t splice sites.

Effect of *dhfr* intron I on transcription.

In earlier experiments performed with much smaller *dhfr* minigenes, there seemed to be no reproducible differences with the presence or absence of the first intervening sequence on production of Dhfr+ colonies (15). In order to test the effect of intron I on transcription of *dhfr* from the enlarged minigene, we deleted the intron as described in Materials and Methods to produce p*dhfr*14.7ori. The only difference between p*dhfr*15ori and p*dhfr*14.7ori is the precise

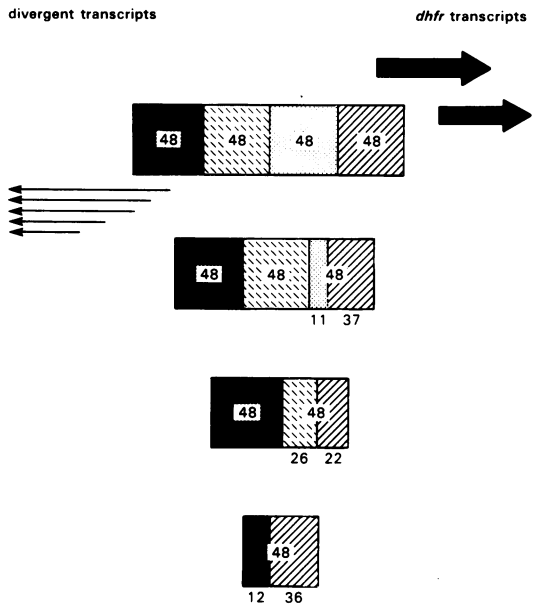

Figure 8. Diagram of promoter repeat constructions. Shown are the compositions of the repeat regions of minigenes with 1,2,3 or 4 copies of the 48 bp repeat of the promoter. Each copy of the repeat differs slightly from the others, and is shown by a different pattern in the diagram.

deletion of intron I in p*dhfr*14.7ori. Various preparations of both minigenes were transfected and measured as described above, with the results shown in Fig. 7. There is a reproducible difference in the two minigenes of 7-8 fold in *dhfr* mRNA production. The minigene copy number was measured in the transfected cells and was not affected by the absence of intron I (results not shown). The divergent transcripts were also measured in the transfected cells and were quantitatively the same from both minigenes (results not shown).

Effect of variable numbers of 48 bp repeats on transcription.

Although the 48 bp repeats of the promoter region are extremely similar, they are not identical. We took advantage of the sequence differences to construct minigene derivatives containing 1, 2, and 3 copies of the 48 bp repeat. A diagram of the resulting repeats in shown in Fig. 8. The resulting minigenes were identical with the exception of the promoter region.

Each of the minigenes was transfected into COSts cells and the resulting RNA analyzed by nuclease protection analysis for *dhfr* and 5′ ends. The effect of variable number of repeats is for the most part quantitative rather than qualitative. As can be seen in Fig. 9A, the 1-repeat minigene has only 20-25% of the *dhfr* RNA present in p*dhfr*15ori whereas the 2 and 3-repeat minigenes have approximately equal and intermediate levels of transcripts. The result is reproducible and is not a function of the copy number of the minigenes (results not shown). The ratios of the -55 and -115 transcripts varies between the constructs; the minigene with 2 repeats produces more than 1 band at approximately -115. In all cases the homology of the probe extends beyond the bands produced. Transcription of the divergent gene gives similar results, as shown in Fig. 9B. There is little transcription from the minigene with one repeat but, unlike

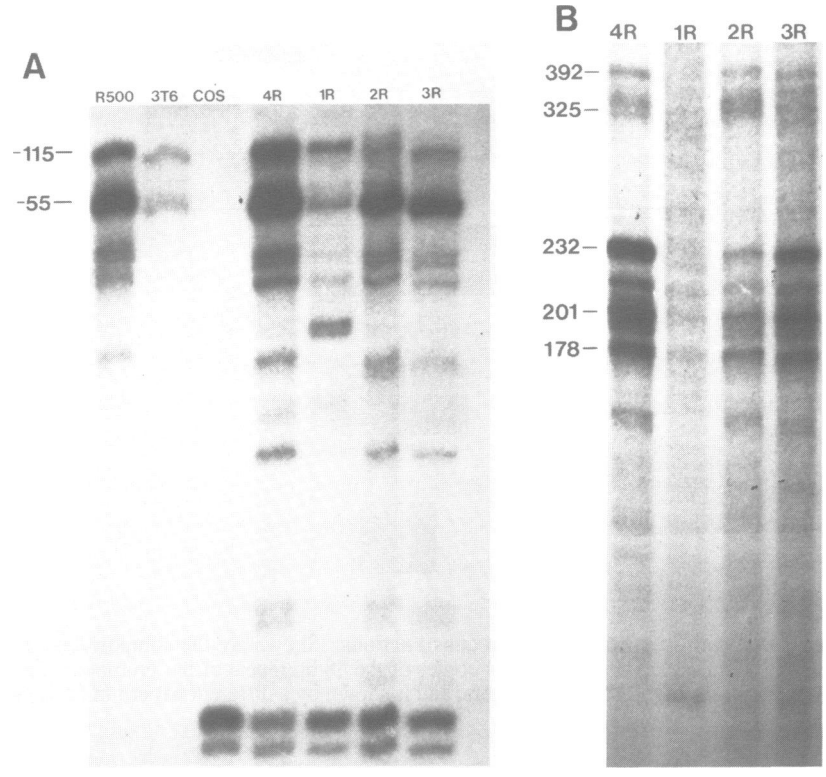

Figure 9. Protection analysis of *dhfr* and divergent 5′ ends from minigenes with variable numbers of 48 bp repeats. COSts cells were transfected with 20 μg per dish of the indicated minigene constructs at 33°C and RNA harvested 2 days after transfection. *dhfr* and divergent 5′ ends were analyzed as in Fig. 3 and 7 respectively. A. Analysis of *dhfr*. The lanes represent protection of 4 μg of 3T3-R500 RNA, 40 μg of 3T6 RNA, and 80 μg of COSts RNA or RNA from cells transfected with minigenes with 1,2,3 or 4 copies of the 48 bp repeat in the promoter region (1R, 2R, 3R, and 4R respectively). B. Analysis of the divergent transcript. The lanes represent protection of 50 μg of RNA from cells transfected with the same minigenes as in A.

dhfr, there is a noticeable difference in the quantity of transcripts produced by the 2 and 3-repeat minigenes. There appears to be no significant qualitative differences in the transcripts produced by the various minigenes.

DISCUSSION

p*dhfr*15ori contains a *dhfr* minigene much enlarged over the previous constructs (15). Except for introns 2-5, all *dhfr* sequences known to be transcribed or to have an effect on transcription are present. These sequences are in a plasmid which replicates both in *E. coli* and in mammalian cells supplying T antigen. The SV40 origin of replication used in the plasmid is a minimal origin sequence, lacking all of the 21 bp repeats and the 72 bp enhancer sequences. In

addition, it lacks early promoter activity, and has only a minor late promoter activity (15,38). Therefore, we expect little, if any, effect on *dhfr* transcription due to the SV40 origin sequences. For example, we know that the orientation of the SV40 origin sequences makes no difference on transcription in the minigene (results not shown). The situation with other replicating systems is quite different, for it has apparently not been possible to design systems making use of replicons such as Bovine Papilloma Virus (18) or Epstein Barr Virus (19) which lack added promoters or enhancer elements. In COSts host cells, the copy number and replication of the plasmid can be controlled by the full or partial temperature inactivation of the T antigen. In addition, the replicated plasmid appears in a Hirt supernatant (39) (results not shown). Any size limit on the plasmid to be replicated must be greater than 17 kb, as there is no evidence of rearrangement or deletion in the replicated plasmids reported here.

Transcription from the *dhfr* promoter region is quite complex, having multiple 5′ ends, multiple 3′ ends, and multiple 5′ ends of a divergently transcribed RNA. We were able to detect in the RNA from transfected cells all of the transcripts which we analyzed in RNA from *dhfr* amplified cell lines, usually in the same relative amounts, although the total level of *dhfr* transcripts is much lower in the transfected cells than in the amplified cell lines. In most cases no extra transcripts are evident. However, extra transcripts are usually observed in the *dhfr* 5′ transcripts; these extra transcripts are clearly seen in Fig. 3 and only faintly seen in Fig. 7. The extra bands at -180 and -300 in the transfected RNA are not visible in S180-M500 RNA. We are not sure why these extra bands are present. The extra DNA inserted upstream in p*dhfr*15ori apparently did not affect their presence. The -180 band would start transcription at the end of the second copy of the 48 bp repeat, and the -300 band would start approximately 50 bp upstream of the first repeat. Other preparations of S180-M500 RNA, using slightly different probes, earlier showed a series of bands near the -300 position (1). The band we see at -300 may be related to those species. Other than these two bands, the patterns of 5′ ends from amplified cell lines and the minigene appear identical.

By making use of the temperature sensitivity of the T antigen, we were able to examine *dhfr* transcription in the presence and absence of p*dhfr*15ori replication. Assuming that the half-life of the *dhfr* mRNA is 6 to 7 h, as it is in amplified cell lines (16), RNA 5′ ends measured in samples held for 12 to 24 h at 40°C would reflect transcription from the plasmids after replication had ceased. It appears, from transfection experiments in COS1 cells, that there are not large differences in *dhfr* RNA stability at 40°C compared to 33°C. Therefore, the results shown in Fig. 4 suggest that there is little effect of replication on the amount or characteristics of the 5′ ends of *dhfr*, as the difference in amount of RNA seems to be a function of the amount of replicated plasmid. However, it should be noted that the majority of the minigene DNA extracted from transiently transfected cells has not replicated. This is true both for DNA isolated as in Materials and Methods, and for DNA isolated according to Hirt (39). It doesn't appear that the unreplicated DNA makes a substantial contribution to the *dhfr* RNA produced, both because of the evidence cited above, and due to the fact that plasmids which do not have an origin (as in Fig. 4B) or which have never replicated produce very small amounts of *dhfr* RNA. We do not know whether the unreplicated minigenes do not transcribe *dhfr* because they have never replicated, or whether they are somehow compartmentalized in the cell such that they neither replicate nor transcribe.

The one noticeable difference in the transcripts from the replicating and the temperature-shifted non-replicating minigenes in the COSts cells is the enhancement of the -55 transcript compared to the -115 transcript in RNA transcribed from replicating plasmids compared to RNA from non-replicating plasmids. It is interesting to note that Farnham and Schimke (40) found that the -55 and -115 start sites were used at similar efficiencies in *in vitro* transcription experiments using a linear template, but that the -55 start site was used at a greatly enhanced

frequency from a supercoiled template. The promoter region of *dhfr* is very GC rich, and replication could help to open up this region for transcription, as could supercoiling.

We also assayed for the presence of the 4 most abundant polyadenylated 3′ ends. (The remaining 2 are present at a much lower level, and we have not yet developed assays for those ends.) Previous results with stably integrated mouse *dhfr* minigenes have demonstrated poor utilization of the normal polyadenylation sites. Kaufman and Sharp found that a SV40 polyadenylation site was necessary in their modular mouse *dhfr* minigenes (41), and Gasser and Schimke found substantial aberrant polyadenylation in various minigenes (42). Recently, CHO *dhfr* minigenes have been found to use the proper polyadenylation sites, but various novel sites were also frequently found (43). In contrast to these results, the pattern of bands of polyadenylated species is identical between amplified RNA and transfected RNA. It may be that 3′ flanking sequences present in p*dhfr*15ori but absent in the other minigenes have some influence on choice of polyadenylation sites. In addition, the relative ratios of all the bands are the same with one major exception. Whereas the smallest mRNA from amplified RNA, giving a band of 134 nt, is one of the most abundant species, this mRNA is barely observed in the transfected RNA. It is surprising that this species is so abundant in the amplified cell, as the polyadenylation signal appears to be UAUAAG, which is a poor match for the canonical AAUAAA signal (44). It is as if cleavage at that position is enhanced in RNA from the chromosomal gene, as opposed to being suppressed in the transfected cells. This polyadenylation site is just 158 bp downstream from the last intron. The last intron is the largest in the gene, approximately 16,500 bp, and in order to splice properly, probably has extensive secondary structure. This secondary structure could create a more favorable environment for cleavage at the first polyadenylation site. The large intron is lacking in the minigene. Another possibility would be a coordination of splicing and polyadenylation. It now seems that snRNPs are involved in the polyadenylation process (44,45). It may be that there is some link between the poly(A) site closest to the 3′ splice site and splicing at that site. Such a linkage might also account for the utilization of what appears to be a poor site. A more minor difference observed is that of RNA with long 3′ ends. Any transcript containing sequences more than 900 nucleotides beyond the fourth poly(A) site hybridizes to the full length of the probe. This band is relatively more intense in the transfected RNA than in the RNA from the amplified cell line; we have not yet identified what 3′ ends are contained on any of the longer RNAs, or how much of that RNA fails to be polyadenylated. We are curious to know whether any of this longer RNA is due to RNA which failed to be polyadenylated at the first polyadenylation site.

Previously, no transfected *dhfr* minigenes have been shown to make any RNA product from the divergent promoter; a stable product would not have been expected, as no polyadenylation site for a divergent transcript was provided in other constructs. It is surprising that divergent promoter activity has not been found in *in vitro* transcription experiments whereas *dhfr* promoter activity is observed. Although one would expect to find a similarly placed divergent promoter in the human gene based on homology (1), no *in vitro* promoter in that region was found (9). Similarly, no divergent promoter activity was found from the mouse promoter region *in vitro* (40). We find that p*dhfr*15ori makes transcripts with the same 5′ ends from the divergent promoter as RNA from a *dhfr* amplified cell line. The majority of the transfected RNA species have a size (from 178 to 232 nt) which suggests that they use the splice site at the end of the first exon. A substantial fraction of the RNA fails to splice at this point as evidenced by its longer length (325 to 393 nt). p*dhfr*15ori has a small t splice from SV40 downstream of the first divergent exon. Therefore we believe that the bands from 178-232 nt result from utilization of the divergent exon splice donor site and the small t splice accepter site, whereas the bands from 325-393 nt are either from unspliced RNA or, more likely, RNA in which the small t splice was used to the exclusion of the divergent exon splice donor site. The important result from this experi-

ment is that the divergent promoter is used, producing stable RNA with the correct 5′ ends.

There has been controversy about the necessity of intervening sequences for mRNA production (46). We have been able to compare directly 2 minigene constructs which differ only in the presence or absence of the first intron, which is only 300 bp in size; there are no other introns in the *dhfr* minigene. It is clear that there is a substantial (7-8 fold) effect of the first intron on *dhfr* mRNA production. Gasser et al. (47) reported a difference in transfection efficiency with mouse *dhfr* minigenes with and without introns. Although in their case the two sets of minigenes were in different vector backgrounds, and contained a subset of sequences used here, they also found that the presence of the first and at least part of the second intron enhanced *dhfr* production. One possible conclusion from our results, that there are enhancer-like sequences in the intron affecting transcription, is less likely because of the lack of any effect of the intron on transcription of the divergent gene. Therefore it appears that the presence of the intron itself is important in the processing of the *dhfr* transcript into stable RNA.

One of the striking differences in the *dhfr* promoter regions of mouse, hamster, and human is the different number of 48 bp repeats, ranging from 4 in mouse to 1 in human. The work reported here suggests that the number of repeats has a quantitative effect on transcription in both directions from the promoter region. It will be interesting to compare the promoter strengths of the various mammalian *dhfr* promoters to determine if there are different relative strengths or if compensating changes in the promoters have been made to balance the lower number of repeats. Given the differences between the 48 bp repeats in the mouse promoter, there are many possible arrangements of the repeats, and we plan to test what sequence differences in the various repeats are important.

It appears that the present *dhfr* minigene contains all the necessary DNA to specify *dhfr* transcripts with the proper 5′ and 3′ ends and to initiate correctly the divergent transcript. Many of the disadvantages of a replicating system have been minimized by the use of a minimal origin sequence deleted for promoter and enhancer sequences and a host cell which permits the inactivation of the protein necessary for replication, thereby allowing transcription in the absence of replication. The assayed plasmids appear to be unrearranged. Whereas in transfection experiments analyzing stably integrated genes there seems to be little correlation with amount of transcript and copy number of the transfected gene (15), we see a direct relation of copy number of replicated plasmid and amount of transcript. There are some differences between RNA from the minigene and from amplified cell lines, particularly with respect to the utilization of the first poly(A) site of *dhfr* and the extra 5′ *dhfr* ends seen in the transfected RNA. We are in the process of testing to see if these differences might be modulated by the addition of selected intervening sequences.

ACKNOWLEDGEMENTS

We are grateful to Liguo New for expert technical assistance, to K. N. Subramanian for the SV40 origin fragment, to R. Tjian for the COSts cells, to Sue Jinks-Robertson and Jeremy Boss for comments on the manuscript and to J. Michael Clark for assistance in the preparation of the manuscript. This research was supported by Public Health Service grant CA40621 from the National Institutes of Health, by American Cancer Society Grant CD-254, and by the Emory University Research Fund. Methotrexate was supplied by the Drug Synthesis and Chemistry Branch, Division of Cancer Treatment, National Cancer Institute.

REFERENCES

1. Crouse, G.F., Leys, E.J., McEwan, R.N., Frayne, E.G. and Kellems, R.E. (1985) Mol. Cell. Biol. 5, 1847-1858.
2. Farnham, P.J., Abrams, J.M. and Schimke, R.T. (1985) Proc. Natl. Acad. Sci. USA 82, 3978-3982.

3. Frayne, E.G., Leys, E.J., Crouse, G.F., Hook, A.G. and Kellems, R.E. (1984) Mol. Cell. Biol. 4, 2921-2924.
4. McGrogan, M., Simonsen, C.C., Smouse, D.T., Farnham, P.J. and Schimke, R.T. (1985) J. Biol. Chem. 260, 2307-2314.
5. Sazer, S. and Schimke, R.T. (1986) J. Biol. Chem. 261, 4685-4690.
6. Setzer, D.R., McGrogan, M., Nunberg, J.H. and Schimke, R.T. (1980) Cell 22, 361-370.
7. Setzer, D.R., McGrogan, M. and Schimke, R.T. (1982) J. Biol. Chem. 257, 5143-5147.
8. Crouse, G.F., Simonsen, C.C., McEwan, R.N. and Schimke, R.T. (1982) J. Biol. Chem. 257, 7887-7897.
9. Chen, M., Shimada, T., Moulton, A.D., Cline, A., Humphries, R.K., Maizel, J. and Nienhuis, A.W. (1984) J. Biol. Chem. 259, 3933-3943.
10. Mitchell, P.J., Carothers, A.M., Han, J.H., Harding, J.D., Kas, E., Venolia, L. and Chasin, L.A. (1986) Mol. Cell. Biol. 6, 425-440.
11. Yen, J.J. and Kellems, R.E. (1987) Mol. Cell. Biol. 7, 3732-3739.
12. Hofbauer, R., Mullner, E., Seiser, C. and Wintersberger, E. (1987) Nucl. Acids Res. 15, 741-752.
13. Lewis, J.A. and Matkovich, D.A. (1986) Mol. Cell. Biol. 6, 2262-2266.
14. Merrill, G.F., Hauschka, S.D. and McKnight, S.L. (1984) Mol. Cell. Biol. 4, 1777-1784.
15. Crouse, G.F., McEwan, R.N. and Pearson, M.L. (1983) Mol. Cell. Biol. 3, 257-266.
16. Leys, E.J., Crouse, G.F. and Kellems, R.E. (1984) J. Cell Biol. 99, 180-187.
17. Hartzell, S.W., Bryne, B.J. and Subramanian, K.N. (1984) Proc. Natl. Acad. Sci. USA 81, 23-27.
18. Lusky, M. and Botchan, M. (1986) Proc. Natl. Acad. Sci. USA 83, 3609-3613.
19. Reisman, D. and Sugden, B. (1986) Mol. Cell. Biol. 6, 3838-3846.
20. Hamkalo, B.A., Farnham, P.J., Johnston, R. and Schimke, R.T. (1985) Proc. Natl. Acad. Sci. USA 82, 1126-1130.
21. Rio, D.C., Clark, S.G. and Tjian, R. (1985) Science 227, 23-28.
22. Gluzman, Y. (1981) Cell 23, 175-182.
23. Gorman, C.M., Merlino, G.T., Willingham, M.C., Pastan, I. and Howard, B.H. (1982) Proc. Natl. Acad. Sci. USA 79, 6777-6781.
24. Graham, F.L. and van der Eb., A.J. (1973) Virol. 52, 456-467.
25. Soberon, X., Covarrubias, L. and Bolivar, F. (1980) Gene 9, 287-305.
26. Gorman, C.M., Padmanabhan, R. and Howard, B.H. (1983) Science 221, 551-553.
27. Bergsma, D.J., Olive, D.M., Hartzell, S.W. and Subramanian, K.N. (1982) Proc. Natl. Acad. Sci. USA 79, 381-385.
28. Crouse, G.F., Frischauf, A-M. and Lehrach, H. (1983) Meth. Enzymol. 101, 78-89.
29. Cesarone, C.F., Bolognesi, C. and Santi, L. (1979) Anal. Biochem. 100, 188-197.
30. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) Nucl. Acids Res. 12, 7035-7056.
31. O'Farrell, P.H., Kutter, E. and Nakanishi, M. (1980) Mol. Gen. Genet. 179, 421-435.
32. Silhavy, T.J., Berman, M.L. and Enquist, L.W. (1984) Experiments with Gene Fusions, pp. 164-165. Cold Spring Laboratory, New York.
33. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
34. Wahl, G.M., Stern, M. and Stark, G.R. (1979) Proc. Natl. Acad. Sci. USA 76, 3683-3687.
35. Johnson, D.A., Gautsch, J.W., Sportsman, J.R. and Elder, J.H. (1984) Gene. Anal. Tech. 1, 3-8.
36. Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6-13.
37. Lusky, M. and Botchan, M. (1981) Nature 293, 79-81.
38. Byrne, B.J., Davis, M.S., Yamaguchi, D.J., Bergsma, D.J. and Subramanian, K.N. (1983) Proc. Natl. Acad. Sci. USA 80, 721-725.
39. Hirt, B. (1967) J. Mol. Biol. 26, 365-369.
40. Farnham, P.J. and Schimke, R.T. (1986) Mol. Cell. Biol. 6, 2392-2401.
41. Kaufman, R.J. and Sharp, P.A. (1982) Mol. Cell. Biol. 2, 1304-1319.
42. Gasser, C.S. and Schimke, R.T. (1986) J. Biol. Chem. 261, 6938-6946.
43. Venolia, L., Urlaub, G. and Chasin, L.A. (1987) Somat. Cell Genet. 13, 491-504.
44. Birnstiel, M.L, Busslinger, M. and Strub, K. (1985) Cell 41, 349-359.
45. Hashimoto, C. and Steitz, J.A. (1986) Cell 45, 581-591.
46. Gross, M.K., Kainz, M.S. and Merrill, G.F. (1987) Mol. Cell. Biol. 7, 4576-4581.
47. Gasser, C.S., Simonsen, C.C., Schilling, J.W. and Schimke, R.T. (1982) Proc. Natl. Acad. Sci. USA 79, 6522-6526.