A genetic polymorphism within the third poly(A) signal of the DHFR gene alters the polyadenylation pattern of DHFR transcripts in CHL cells

Honghao Yang and Peter W. Melera*

Graduate Program in Molecular and Cell Biology and Department of Biological Chemistry, University of Maryland, Baltimore, MD 21201, USA

Received February 4, 1994; Revised and Accepted June 1, 1994

ABSTRACT

Two polymorphic dihydrofolate reductase (DHFR) alleles, termed 20 K and 21 K, exist in Chinese hamster lung cells. Three major transcripts of different lengths are transcribed from each allele, and the expression of these transcripts differs dramatically between the alleles as a result of differential utilization of three poly(A) sites. Transcripts from the 20 K allele are preferentially polyadenylated at the first poly(A) site, while those from the 21 K allele are preferentially polyadenylated at the third site. In this study, transient expression experiments were used to demonstrate that a 2.1 kb genomic fragment containing the three DHFR poly(A) sites is sufficient to reproduce the allele-specific polyadenylation pattern on transiently expressed CAT-DHFR transcripts in COS cells. Site-directed mutagenesis allowed identification of the sequence elements which are responsible for this allele-specific polyadenylation. These studies indicate that a single-base change in the third poly(A) signal sequence, which alters the consensus AAUAAA signal in the 21 K allele to a weak AAUAAU signal in the 20 K allele, is primarily responsible for the dramatic difference in polyadenylation between the two alleles. Thus, as a result of this single-base change in the third poly(A) signal sequence, utilization of the first poly(A) site, located 1.2 kb upstream, changes dramatically.

INTRODUCTION

The 3' termini of most eukaryotic mRNAs are generated by a sequence-specific cleavage of the primary transcript followed by addition of a long poly(A) tail (1, 2). A highly conserved hexanucleotide, AAUAAA, is required for cleavage and polyadenylation (3–5). In addition, sequences downstream from the cleavage site, which are generally GU or U-rich, are also thought to be essential for accurate and efficient cleavage (6–9). Variants of the consensus AAUAAA sequence, containing single-base changes, exist in nature (1, 10, 11). However, such variants are weak poly(A) signals compared to the consensus AAUAAA

signal (1, 3, 12). While most genes in higher eukaryotes contain a single functional poly(A) site, there are several examples of genes that contain multiple poly(A) sites (13). Utilization of multiple poly(A) sites can produce different mRNAs that encode functionally distinct gene products. Examples include the adenovirus major late gene, the calcitonin gene, and the immunoglobulin heavy-chain gene (14–16). On the other hand, mRNA produced using alternative poly(A) sites can still maintain the same protein-coding region, while extending or shortening the length of the 3' untranslated region. Examples include bovine prolactin gene, the mouse α -amylase gene and the mouse, human and hamster dihydrofolate reductase (DHFR) genes (10, 11, 17–19). In the later cases, the functional significance of the differential utilization of multiple poly(A) sites is not clear.

Two DHFR alleles, have been identified in the Chinese hamster lung (CHL) fibroblast cell line, DC-3F (20). As a result of selection with the antifolate methotrexate (MTX), either allele may be found to be amplified, resulting in overexpression of DHFR and commensurate drug resistance (21, 22). This mutually exclusive amplification of one allele or the other provides an ideal system in which expression of either can be independently analyzed. Two of the more well documented MTX-resistant cell lines are DC-3F/A3 and DC-3F/MQ19, each of which respectively amplify the 21 K or 20 K allele over one hundred fold. Three major different-sized mRNAs (1000, 1650 and 2150 nt) are expressed from each of the two DHFR alleles (Fig. 1A) (23–25). However, the steady state level of the three mRNA species is specific to the allele expressed and is consistently so in independently derived MTX-resistant sublines. Hence, the shortest (1000 nt) species is predominantly expressed in sublines amplifying the 20 K allele, while the longest species (2150 nt) is preferentially expressed in sublines amplifying the 21 K allele. Analysis of the 3' ends of the three mRNAs from each allele demonstrate that they differ primarily in the lengths of their 3' untranslated regions (UTRs), which in turn arise as a result of utilization of three different poly(A) sites (19). Moreover, the allele-specific expression profiles of the DHFR mRNAs are believed to result from differential utilization of these poly(A) sites (19).

*To whom correspondence should be addressed

We have used transient expression from intronless CAT reporter gene constructs whose transcripts are polyadenylated by direction of the poly(A) signals present in the 3' end of the two DHFR alleles, to study the mechanism whereby the changes in the allele specific polyadenylation pattern occur. The results of these experiments show that a single-base change in the third poly(A) signal sequence, which alters the consensus AAUAAA sequence in the 21 K allele to AAUAAU in the 20 K allele, is primarily responsible for the dramatic differences in polyadenylation between the two alleles. More specifically, our results show that an alteration in the third poly(A) signal not only decreases utilization of that signal but also increases the extent to which transcripts derived from the first signal, located 1.2 kb upstream, account for the composition of the DHFR mRNA population. Thus, the degree to which the first signal is utilized is dependent upon its intrinsic strength and upon the integrity of the third signal as well.

MATERIALS AND METHODS

Cell culture, RNA extraction and poly (A)⁺ mRNA purification

The CHL cell line, DC-3F, and its independently derived MTX-resistant sublines DC-3F/A3 and DC-3F/MQ19 (21, 22) were grown in α MEM-F12 (GIBCO) medium supplemented with 5% fetal calf serum and 50 μ g/ml and 10 μ g/ml of MTX, respectively. COS-1 cells used for transfection were maintained in Dulbeccos modified Eagle medium (GIBCO) supplemented with 10% fetal calf serum. Total RNA was isolated by the RNazol method (26), whereas poly (A)⁺ mRNA was prepared by oligo d(t) chromatography of total RNA (27).

Northern blot analysis

Northern blot analyses were carried out as described by Zinn *et al.* (28). Antisense single-stranded RNA probes were used and labeled by the PCR-mediated *in vitro* transcription method (29).

Construction of transfection vectors

pSCAT6, a kind gift from Dr K. Agarwall at University of Chicago, is a derivative of pSV2CAT from which the SV40 t-intron, splicing signals and polyadenylation regions have been removed (30). We replaced those sequences by insertion of a 2.1 kb *Bam*HI genomic fragment obtained from the 3' end of either the 21 K or 20 K DHFR allele that contains the three poly(A) signals of interest (19). As a result, the final constructs, pSCAT-21 and pSCAT-20 (Fig. 2) carry the 3' end of the 21 K and 20 K alleles, respectively, cloned downstream of a CAT reporter gene. Expression, in transfected COS cells, is driven by the SV40 promoter. In this configuration, polyadenylation of CAT-DHFR transcripts is directed by the signals contained in the hamster genomic DNA.

Transfections and CAT assays

Transfection was carried out by the DEAE-dextran method (31). Routinely, cells grown in 10-cm dishes were transfected with 10 μ g of the appropriate plasmid DNA, and co-transfected with 3 μ g of pRSVgal, an internal control vector which contains a β -galactosidase reporter gene (32). Total RNA were extracted 72 hours after transfection and used for RNase protection analysis. At the same time, 100 μ l of cell lysate were prepared for assay of CAT activity (33). The Bradford assay (34) was used to determine protein concentrations. Based upon the protein

measurement, equal amount of sample of each construct was used for CAT assay. The level of β -galactosidase activity was also determined as described (35) and was used as a measure of transfection efficiency. The final values of CAT activities were adjusted based upon their different efficiencies of transfection measured by β -galactosidase assay.

Antisense RNA probe preparation by PCR and RNase protection assays

A PCR-modified RNase protection assay was used in this study (29). A pair of primers were designed to complement the sequences flanking the desired poly(A) sites. A SP6 RNA polymerase promoter sequence (PPS) was attached to the 5' end of the antisense primer. As a result of the ensuing PCR amplification of genomic DNA, a linear DNA template containing a terminal PPS was generated. Radiolabeled transcripts initiated from the PPS were then used as antisense RNA probes. 5×10^5 cpm of probe were hybridized at 30°C with 10 μ g cellular RNA from MTX-resistant cells or 2 μ g poly A⁺ RNA from the transfected COS cells. Thirty-five μ g/ml RNase T2 was used to remove unprotected segments of RNA. Protected

Figure 1. A. Schematic diagram of the 3' end of DHFR gene. **B.** Northern blot analysis of DHFR mRNA. A 232 nt single-stranded ³²P-radiolabeled antisense RNA probe complementary to the sequence across the first DHFR poly(A) site was used for the Northern blot hybridization. About 5×10^7 cpm of probe was applied in this experiment and 10 μ g of total RNA was used for each sample. Lanes 1, DC-3F; 2, DC-3F/A3 and 3, DC-3F/MQ19. PCR followed by *in vitro* transcription was used to synthesize the antisense RNA probe.

fragments were analyzed on 6% denaturing acrylamide gels containing 7 M urea in parallel with DNA sequence ladders as size markers. The intensity of the radioactivity in each band was measured by densitometric scanning of exposed X-ray films.

Site-directed mutagenesis

Eight mutants were generated via site-directed mutagenesis using Promega's Altered Site™ system (36). In brief, the 2.1 kb *Bam*HI fragment containing the three poly(A) sites was cloned into a phagemid vector, pSELECT-1 (Promega), followed by single-stranded DNA extraction. Hybridization was carried out between the single-stranded DNA template and two synthetic oligonucleotides. The first one, containing a mismatch or mismatches in the middle of the fragment, was used to incorporate a point mutation(s) into the complementary DNA strand. The second one, an ampicillin resistant correction primer (Promega), was used to confer ampicillin resistance to the mutant DNA strand for selection. Following hybridization, both oligonucleotides were extended with DNA polymerase (Promega) to create a double-stranded structure. The remaining nicks were then sealed with

DNA ligase (Promega) and the duplex structure transformed into an *E. coli* host. Mutants were then selected by ampicillin resistance and verified by sequencing. The resulting mutant 2.1 kb inserts were then recloned into the original transfection vector.

RESULTS

Allele-specific DHFR RNA expression demonstrated by Northern blot analysis

The allele-specific expression profiles of DHFR transcripts found in sublines of DC-3F are demonstrated by the Northern blot shown in figure 1B and have been reported previously (23–25). Three major transcripts, 1000, 1650 and 2150 nt in length formed by use of three poly(A) sites (Fig. 1A) are found in these cells and the abundance of each is different between the two alleles. In the case of the 21 K allele represented by DC-3F/A3 (Fig. 1B, lane 2), the longest transcript (2150 nt) is expressed at high levels, while the 1000 nt and the 1650 nt transcripts are expressed at relatively lower levels. On the other hand, expression from the 20 K allele represented by DC-3F/MQ19 (lane 3) is

Figure 2. Construction of pSCAT-21 and pSCAT-20. The 2.1 kb genomic fragment from both DHFR alleles containing the three DHFR poly(A) sites were purified from plasmids p21K-2.1 and p20K-2.1 (19) through *Bam*HI digestion. This 2.1 kb fragment was then cloned into the vector pSCAT6 at *Bam*HI site which is located downstream of the CAT reporter gene. The orientation of the insert was determined by restriction mapping and sequencing.

characterized by predominant expression of the shortest transcript (1000 nt), while the 1650 nt and the 2150 nt species are expressed at the lower levels. A minor transcript of approximately 6 kb in length found in poly(A)⁺ RNA preparations, but not readily observed in blots of cellular RNA as shown in figure 1, is also formed from the 20 K allele, presumably through polyadenylation at a site located 4 kb further downstream of the third poly(A) site. No such transcript is found to be expressed from the 21 K allele (19). DC-3F (lane 1) is the parental cell line used here as a control. No specific DHFR transcripts can be seen in this lane due to their low level of expression. However, use of poly(A)⁺ RNA and long exposure times has shown that the same three transcripts are expressed in DC-3F cells and that their distribution is representative of a mixture of the transcripts from the two alleles (20).

Although some variation in the relative amount of the three RNA species is observed between different RNA samples, the overall difference in the distribution profile of the three major transcripts expressed from the two alleles remains consistent. By determining the relative amounts of each DHFR species by multiple Northern blot analyses of total cellular RNA (data not shown), the distribution of these transcripts is as follows: from the 21 K allele, the 1000 nt transcript accounts for 15%, the 1650 nt transcript 35%, and the 2150 nt transcript 50% of the total DHFR RNA, whereas from the 20 K allele, the 1000 nt transcript accounts for 45% of the total DHFR RNA, and the 1650 and the 2150 nt transcripts account for 35% and 20%, respectively. Analysis of poly(A)⁺ RNA indicates that the 6000 nt transcript expressed from the 20 K allele accounts for less than 5% of the DHFR RNA expressed from that allele (19).

Figure 3. CAT assay. The level of CAT activity was quantitated by determining the extent of conversion of [¹⁴C]chloramphenicol to the monoacetate and diacetate forms. The quantitation was carried out by a Dot Analyzer (Betagene Corp). The result was then normalized by comparison with the co-expressed β -galactosidase activity. Lanes: 1, pRSVgal; 2, pSV2CAT; 3, pSCAT6; 4, pSCAT-21; 5, pSCAT-20.

CAT activity as a measurement of polyadenylation efficiency of poly(A) sites in transfected COS cells

The 2.1 kb *Bam*HI genomic fragment that in each DHFR allele contains the three poly(A) sites (19) was cloned into pSV2CAT downstream of the CAT reporter gene (Fig. 2). In this configuration, CAT-DHFR transcripts are polyadenylated at the three DHFR poly(A) sites contained within the cloned 2.1 kb fragment. The level of CAT expression, as measured by the level of CAT activity post transfection, depends upon the efficiency of polyadenylation. As seen in Figure 3, pSV2CAT (lane 2), containing a SV40 poly(A) site and a small t intron, was used as a positive control, and produced the highest CAT activity (100%). Deletion of the intron and replacement of the SV40 poly(A) site with the three DHFR poly(A) sites from either allele generates essentially the same level of CAT activity: 97 % from pSCAT-21 (lane 4) and 93 % from pSCAT-20 (lane 5). Low activity (15.5 %) is observed following transfection of pSCAT6, a negative control vector that contains neither a SV40 poly(A) site nor the DHFR poly(A) sites (lane 3). This low activity is probably due to the expression of a small amount of CAT-DHFR transcript polyadenylated at cryptic sites present within the pSCAT6 sequence (30). No CAT activity is seen after transfection with pRSVgal (lane 1), a vector which does not contain the CAT gene. That the expression of pSCAT-21 or pSCAT-20 in COS cells both generate essentially the same level of CAT activity is consistent with previous RNase protection studies in which it was demonstrated that the same amount of polyadenylated CAT-DHFR transcript is produced from pSCAT-21 and pSCAT-20 during transient expression in COS cells (Yang, Hussain and Melera, submitted).

Mapping the poly(A) sites found in CAT-DHFR transcripts expressed by pSCAT-21 and pSCAT-20 transfectants

To map the poly(A) sites of the CAT-DHFR transcripts expressed from either pSCAT-21 or pSCAT-20, three antisense RNA probes were designed to complement the sequence spanning each of the three poly(A) sites. The results are shown in figure 4. To map the first poly(A) site, a 171 nt antisense RNA probe was used (Fig. 4A). A 165 nt fragment protected by the probe represents readthrough (RT) transcripts polyadenylated at the second and third sites. Four protected fragments, 95, 101, 105 and 108 nt in length, represent the transcripts polyadenylated at the first poly(A) site (19). The occurrence of the same-sized fragments in DC-3F/A3 (lane 2), DC-3F/MQ19 (lane 3), pSCAT-21 (lane 4), and pSCAT-20 (lane 5), indicates that the same sites were used for polyadenylation of the transiently expressed short CAT-DHFR transcripts in COS cells and the 1000 nt DHFR transcript in CHL cells. The four fragments are much less abundant in the RNA from pSCAT-21 than in the RNA from pSCAT-20. This is consistent with the difference observed *in vivo*, in which the majority of 21 K allelic DHFR transcripts are polyadenylated at the third poly(A) site while those from the 20 K allele are primarily polyadenylated at the first site (Fig. 1B and 19).

To map the second poly(A) site, a 191 nt antisense probe was used (Fig. 4B). A 185 nt fragment protected by the probe corresponds to an RT transcript which ends at the downstream poly(A) site. The 116 nt fragment corresponds to the transcript which ends at the second poly(A) site in the cloned DHFR genomic fragment. The same-sized 116 nt fragment is noted for pSCAT-21 and pSCAT-20 (lanes 5 and 6) as well as for

DC-3F/A3 and DC-3F/MQ19 (lanes 2 and 3). Therefore, the second poly(A) site used by the 1650 nt DHFR mRNA is utilized in this case by the middle-sized CAT-DHFR transcripts from both pSCAT-21 and pSCAT-20. Similar amounts of the 116 nt fragments are found for either allele in both transfection samples (lane 5 vs. 6) and control samples (lane 2 vs. 3), suggesting that relatively the same amount of transcript is processed at the second poly(A) site in both alleles.

When a 175 nt antisense RNA probe was used to map the poly(A) site for the long CAT-DHFR transcripts from both pSCAT-21 and pSCAT-20 (Fig. 4C), different-sized fragments were protected: a 103 nt fragment for pSCAT-20 (lane 6) and 120 nt and 124 nt fragments for pSCAT-21 (lane 5). These different-sized fragments localize the third poly(A) site at different positions for both alleles. Compared with the third site in the pSCAT-20 transcript, the third poly(A) site of the pSCAT-21 transcript is shifted 17 nt and 21 nt downstream. The amount of these protected fragments from either allele is also different. The level is higher in pSCAT-21 RNA (lane 5) than in RNA from pSCAT-20 (lane 6). Similar size and abundance differences between the two alleles exist among the DHFR transcripts extracted from DC-3F/A3 (lane 2) and DC-3F/MQ19 cells (lane 3) as well. This similarity between the transfectants and *in vivo* samples supports the conclusion that the same allelic specificity for polyadenylation occurs at the third poly(A) site in both cases. The RT transcript noted as a 169 nt doublet in figure 4C lane 3, represents the minor 6 kb transcript which is polyadenylated approximately 4000 nt downstream of the third poly(A) site (19). The fact that it can only be seen in DC-3F/MQ19 (lane 3) and not in DC-3F/A3 (lane 2), is consistent with our previous finding, suggesting that this minor poly(A)⁺ DHFR transcript is only formed from the 20 K allele (19). No such RT transcripts are observed in the RNA from either pSCAT-21 (lane 5) or pSCAT-20 (lane 6) transfectants. This suggests that no other functional poly(A) sites exist downstream of the third poly(A) site in the cloned 2.1 kb fragments.

Overall, the results confirm earlier *in vivo* DHFR transcript mapping data (19) and in addition demonstrate that the polyadenylation patterns of CAT-DHFR transcripts expressed in transfected COS cells as directed by the 2.1 kb *Bam*HI genomic fragments taken from the 3' ends of the two DHFR alleles faithfully reproduce the polyadenylation patterns seen *in vivo*.

Figure 4. Mapping of the first, second and third poly(A) sites of CAT-DHFR transfectants. Three pairs of the PCR primers were used to map the three poly(A) sites: PA1-5' and PA1-3' (for the first site); PA2-5' and PA2-3' (for the second site); and PA3-5' and PA3-3' (for the third site). DNA sequence ladders run in parallel with RNA samples were used as markers to identify the size of each protected fragment. To accommodate the various sizes of the protected fragments of the CAT-DHFR transcripts at each poly(A) site, different lengths of pRSVgal protected fragments were used as internal controls in the different mapping experiments, i.e. 86 nt for site one, and 98 nt for sites two and three. **A.** Mapping of the first poly(A) site. Lanes: P, undigested probe; 1, yeast tRNA; 2, DC-3F/A3; 3, DC-3F/MQ19; 4, pSCAT-21; 5, pSCAT-20. The horizontal arrows indicate: P, the undigested probe; RT, the readthrough transcript; PA1, the transcripts that are polyadenylated at the first poly(A) site; IC, the co-expressed pRSVgal transcript, used here as an internal control. The blackened triangles represent nonspecific protected fragments. The four fragments (95, 101, 105 and 108 nt) in lane 4 are difficult to see, but can be detected in a longer exposure of the gel. **B** and **C.** Mapping of the second and third poly(A) sites, respectively. Lanes: P, undigested probe; 1, yeast tRNA; 2, DC-3F/A3; 3, DC-3F/MQ19; 4, COS-1 (without transfection); 5, pSCAT-21; 6, pSCAT-20. The same abbreviations and triangle markers as used in panel A are used in panel B and C. Two bands corresponding to the readthrough transcript noted in panel C, lane 3 are considered to be a nuclease digestion artifact of the experiment since they were not seen in our previous RNase protection analyses (19).

Mapping the poly(A) sites of CAT-DHFR transcripts expressed from mutant plasmids

Site-directed mutagenesis was performed to identify the sequence elements in the 2.1 kb *Bam*HI fragments of the DHFR gene that are primarily responsible for the allele-specific polyadenylation patterns. Eight mutants were generated, and their sequences are shown in figure 5. Comparing the sequences of these fragments from the two alleles, four nucleotides were found to be deleted from the 3' end of the third poly(A) signal in the 20 K allele (Fig. 5A and 19). As a result, the sequence of the third signal changes from AAUAAA in the 21 K allele to AAUAAU in the 20 K allele. The impact of this change was examined by four mutants, three of which were made in the *Bam*HI fragment from the 21 K allele (21M3-1, 21M3-2 and 21M3-3) and one in the fragment from the 20 K allele (20M3-1). The first two mutants changed the consensus AATAAA sequence of the third poly(A) signal to either AATAAT (21M3-1) or AACAAA (21M3-2). In the third mutant, a double point mutation was introduced. Our previous studies had shown that the four-base deletion that disrupts the consensus AATAAA signal in the 20 K allele results in the use of a cryptic signal, TATAAA, located 26 nt upstream of AATAAA (Fig. 5A and 19). The double mutant (21M3-3) of the 21 K allele disrupts both, converting the TATAAA and AATAAA sequences to TACAAA and AACAAA, respectively. In the fourth mutant, 20M3-1, a point mutation was introduced

A.

The Normal Sequence:

The First Poly(A) Site:

The Third Poly(A) Site:

B.

The Mutants:

	The First Poly(A) Site	The Third Poly(A) Site
21M3-1	—	AATAAA → AATAAI(1294)
21M3-2	—	AATAAA → AACAAA(1291)
21M3-3	—	TATAAA/AATAAA → TACAAA/AACAAA(1265/1291)
20M3-1	—	AATAAI → AATAAA(1298)
20M3-2	—	AATAAT → AATAATCAAT (1299-1303)
20M1-1	AITAAA → AATAAA(123)	—
21M1-1	AITAAA → AATAAA(123)	—
20M1-2	I → C(152)	—

Figure 5. Sequence alterations in the eight mutants created by site-directed mutagenesis. **A.** Only the short sequences downstream of the first and third poly(A) signals, in which the mutations have been made, are shown. The first and third poly(A) signal sequences are bold, italicized and underlined. The numbers shown along with the sequences refer to the nucleotide positions relative to the beginning of the 2.1 kb *Bam*HI fragment sequence (19). The polyadenylation cleavage sites are indicated by vertical arrows and the C/T conversion between the two alleles located downstream of the first poly(A) signal is marked by a rectangular box. **B.** The number in each parenthesis shown in the table represents the nucleotide position of the base that has been altered by site-directed mutagenesis.

Figure 6. Mapping of the poly(A) sites in mutant CAT-DHFR transcripts. The same PCR primers as those used in figure 4A, B, and C were used to map the three poly(A) sites of the mutant transcripts. For panel A, B and C, lanes: P, undigested probe; 1, yeast tRNA; 2, DC-3F/A3; 3, DC-3F/MQ19; 4, COS-1 (without transfection); 5, pSCAT-21; 6, pSCAT-20; 7, 21M3-1; 8, 21M3-2; 9, 21M3-3; 10, 20M3-1; 11, 20M3-2; 12, 20M1-1; 13, 21M1-1; 14, 20M1-2. The same abbreviations, triangle markers and internal control RNA fragments as used in figure 4 are used here. In panel A, the four small protected fragments (95, 101, 105 and 108 nt) are difficult to see in lanes 5 (pSCAT-21) and 10 (20M3-1), but they can be detected in a longer exposure of the gel. In panel A, the 124, 120, 116 and 103 nt fragments represent transcripts polyadenylated at the third site. The 120 and 116 nt protected fragments (labeled as PA3*) noted in mutant 20M3-1 (lane 10) are four bases shorter than their counterparts of 124 and 120 nts (labeled as PA3(21K)), observed in the samples reflecting the normal 21 K allele (lanes 2 and 5). This four base difference reflects the change in length caused by the four-base deletion present in this mutant (see Fig. 5A).

to alter the third poly(A) signal, AATAAT in the 20 K allele, to the consensus sequence AATAAA, which is similar to the sequence present in the 21 K allele. However, in this particular

Figure 7. Summary of the results from the mutagenesis studies. The three different histogram bars represent the CAT-DHFR transcript that is processed at each of the three poly(A) sites. ■, corresponds to the CAT-DHFR processed at the first poly(A) site; ▨, the CAT-DHFR transcript processed at the second poly(A) site; and, □ the CAT-DHFR transcript processed at the third poly(A) site. The numbers on the Y axis represent the percentage of CAT-DHFR transcript that is processed at each of the three sites. The results shown were obtained by quantitation of the radioactivity in specific gel bands by a Molecular Dynamics Scanning Densitometer. Comparisons were then made between the *in vivo* DHFR samples and the normal transfectant CAT-DHFR samples as well as the mutant CAT-DHFR samples. Since the relative fraction of the total DHFR transcript being processed at each of the three poly(A) sites is known from Northern analyses which were quantitated by Dot Analyzer (Betagen Corp.), the approximate amount of CAT-DHFR transcript polyadenylated at each poly(A) site can be assayed by comparison between the *in vivo* DHFR RNA samples and the transfectant CAT-DHFR RNA samples. The data shown here are representative of a typical result obtained from multiple independent experiments conducted during the mapping studies. The results are reported in this manner because, due to the relatively low level of transient expression, clear detection of the three CAT-DHFR mRNAs in one gel by Northern blot analysis was not possible.

mutant, the spatial shortening resulting from the four-base deletion in the 20 K allele was not restored.

The impact of these four mutations on the pattern of polyadenylation is shown in figure 6A, B, and C, lanes 7 - 10, and the results are summarized in figure 7. Compared with the wild-type, disruption of the third poly(A) signal sequence, AATAAA, in the 21 K allele (mutants 21M3-1 and 21M3-2) dramatically reduces utilization of this site (Fig. 6C, lane 5 vs. 7-9), while, at the same time, substantially increasing the utilization of the first site (Fig. 6A, lane 5 vs. 7-9). In addition, the site of polyadenylation shifts, in an allele dependent manner, either 17 nt or 21 nt upstream where TATAAA is used as the poly(A) signal (Fig. 6C, lane 5 vs. 7 and 8). Presumably due to the destruction of both poly(A) signals, no clearly visible transcript is processed at either site in mutant 21M3-3, in which double point mutations were introduced (Fig. 6C, lane 9). In contrast, when the wild-type AATAAT sequence in the 20 K allele is converted to the consensus AATAAA sequence in mutant 20M3-1, the utilization of the third poly(A) site is dramatically increased to a level similar to that observed in transfectants with the wild-type 21 K allele, even though the spatial shortening caused by the four-base deletion is not restored (Fig. 6C, lanes 5 vs. 10). At the same time, utilization of the first poly(A) site drops accordingly (Fig. 6A, lanes 6 vs. 10), and the amount of RT transcript at the second poly(A) site substantially increases as indicated by the intensity of the 185 nt fragment in figure 6B, lane 10. This is consistent with a shift of polyadenylation from the first site to the third site as reflected by the signals obtained from the control plasmids (Fig. 4B, lane 5 and 6) and the RNA obtained from DC-3F/MQ19 and DC-3F/A3 cells (Fig. 4B, lane 2 and 3) as well. The inability to detect differences in the level of RT transcripts present at the first site (Fig. 4A) results from the need to overexpose the gel in order to adequately visualize the multiple transcripts found at that site (19).

In comparison, none of the four mutations introduced at or near the site of the third poly(A) signal altered utilization of the second poly(A) site (Fig. 6B, lane 5 and 6 vs. 7-10).

To further identify the role of the spatial change caused by the four-base deletion, a four base sequence, CAAT, was inserted immediately following the third poly(A) signal sequence in the 20 K allele (20M3-2), thus restoring the spacing found in the 21 K allele. In this configuration, the nonconsensus AATAAT signal sequence remains unchanged. No alterations in polyadenylation were found at any of the three poly(A) sites resulting from this insertion (Fig. 6A, B, C, lane 11 and Fig. 7).

Point mutations were also placed within the sequence of the first poly(A) signal. At first, the original sequence, ATTAATA, was changed to the consensus sequence AATAAA in both alleles (20M1-1 and 21M1-1). Despite this sequence alteration (20M1-1), no substantial changes in utilization of the three poly(A) sites in the 20 K allele were observed (Fig. 6A, B, C, lane 12; Fig. 7). The lack of change in polyadenylation is probably due to the fact that the utilization of the first site has already predominated in the 20 K allele, and further enhancement of the first signal does not substantially increase that utilization. A dramatic change, however, is observed in the 21 K allele following the same sequence change at the first poly(A) signal (21M1-1). Prior to this sequence change, the third poly(A) site is predominant among the three sites (Fig. 6A, B and C, lane 5). The presence of the mutation, however, shifts the dominant site from the third to the first site (Fig. 6A, B and C, lane 13 and Fig. 7). The utilization of the third site decreases to about

half of the original level (Fig. 6C, lane 13), while the utilization of the second poly(A) site drops dramatically (Fig. 6B, lane 13).

Further comparison of sequence differences between the two alleles also indicates the existence of a point mutation, i.e. a T/C conversion, located 25 nt downstream of the first poly(A) signal (Fig. 5A and 19). This point mutation is located near the putative downstream sequence element of the first poly(A) signal and, therefore, may affect utilization of the first poly(A) site. To examine the potential impact of this T→C point mutation on polyadenylation, the nucleotide T in the 20 K allele was changed to a C in order to mimic the sequence present in the 21 K allele (20M1-2). As shown in lane 14 of figure 6A, B and C, and in figure 7, this sequence alteration did not result in any observed change in polyadenylation, indicating that decreasing the U content in the sequence downstream the first poly(A) signal, albeit by a single residue, has no effect on the processing of transcripts at the first poly(A) site.

DISCUSSION

For this report we have utilized site-directed mutagenesis of the 3' exon of two polymorphic Chinese hamster DHFR alleles coupled with transfection studies in an attempt to clarify the mechanism(s) responsible for the unique patterns of polyadenylation that specifically accompany the expression of each allele (20). Previous studies (19) had identified several nucleotide sequence differences in the genomic DNA containing the multiple poly(A) signals of each allele, and together with RNase protection assays had shown that a mutation in the third poly(A) signal of the 21 K allele, converting AAUAAA to AAUAAU in the 20 K allele, was present (19).

To investigate the extent to which this mutation alone was responsible for the altered polyadenylation patterns, separate CAT reporter gene constructs containing the relevant portions of the 3' exons of each allele were prepared and their expression in COS cells analyzed. Our data show (Fig. 6 and 7) that the single-base change in the third poly(A) signal sequence of the 20 K allele (resulting from a four base deletion), is indeed the major factor responsible for the differences in polyadenylation pattern among transcripts from the two DHFR alleles in CHL cells. That equivalent levels of CAT activity are expressed from each of the CAT-DHFR chimeric constructs and the pSVCAT control as well (Fig. 3), however, suggests that regardless of the dramatic differences in polyadenylation patterns that exist between the two alleles, the total amount of RNA polyadenylated *in vivo* is the same and that the efficient processing afforded by the SV40 poly(A) signal present in pSVCAT (33) can be replaced by the three signals present in either DHFR allele, even though in the case of the 20 K allele, no consensus AAUAAA signal is present. These results are consistent with others to be presented elsewhere showing that in similar transfectants the levels of transient poly(A)⁺ CAT-DHFR mRNA expression from each construct are indistinguishable (Yang *et al.*, submitted). The fact that the polyadenylation pattern of CAT-DHFR transcripts from the two allele-specific constructs faithfully reproduces the patterns of polyadenylation observed from the endogenous DHFR alleles (Fig. 4 A, B and C) also shows that, in contrast to other reports (38), the presence of a terminal 3' intron is not required for accurate poly(A) site selection *in vivo*.

The remainder of the sequence differences between the two alleles, including the spatial shortening caused by the four-base deletion at the third poly(A) site in the 20 K allele and the T/C

conversion located 25 nt downstream of the first poly(A) signal, have no detectable effect on the utilization of any one of the three poly(A) sites (Fig. 6 and 7). The fact that the spatial shortening per se does not affect polyadenylation is consistent with recent studies (39) indicating that 5 to 7 nt is the minimal length required between the poly(A) signal and the cleavage site to ensure efficient polyadenylation. The four-base deletion in the 20 K allele brings the distance between the upstream poly(A) signal sequence and the downstream cleavage site to 9 nt, well outside this minimum range.

The lack of a negative effect on utilization of the first poly(A) signal in the 20 K allele by decreasing the 'U-richness' of the sequence immediately downstream of the cleavage site (Fig. 6A, lane 14), may reflect the fact that no GU or U-rich region is present (Fig. 5A and 19). Nevertheless, when the first signal is converted to a consensus in the 21 K allele (mutant 21M1-1), its use increases about four fold, allowing it to become the predominant signal (Fig. 6A, lane 13), while use of the third signal, which does contain a GU-like consensus sequence 31 nucleotides downstream (19), decreases by 50% (Fig. 6C, lane 13). Interestingly, this is the only construct tested in which utilization of the second site is appreciably affected, and it is dramatically reduced (Fig. 6B, lane 13).

Surprisingly, converting the first signal to consensus in the 20 K allele (mutant 20M1-1) has very little impact on the overall pattern of polyadenylation (Fig. 6A, B and C, lane 12; and Fig. 7), suggesting that the lack of a downstream GU-rich region may limit the efficiency (40) with which the first signal can be utilized. It must be kept in mind, however, that in the wild-type 20 K allele, the first signal is the dominant signal. Comparing the effect of placing a consensus sequence at the first signal in both alleles indicates that the relative degree to which that signal is utilized is dependent upon the intrinsic strengths of both the first and third signal.

These results raise some interesting points concerning the mechanism of poly(A) site selection in genes that contain multiple poly(A) signals. Although 5' to 3' and 3' to 5' scanning models have been discussed (41), the former cannot easily explain the strict reliance that utilization of the first poly(A) site in the DHFR gene has on the nature of the third site which is located 1.2 kb downstream, nor can the latter explain the result obtained with the mutant construct 21M1-1 in which the first site becomes dominant over the third, even though the third signal remains consensus. Instead, our results are most easily explained by a model suggested previously (19) in which processing of DHFR primary transcripts does not occur to any appreciable extent until transcription proceeds beyond the third site, i.e. processing at the first site and probably the second as well is slow compared to the rate of pol II transcription. Active cleavage and polyadenylation complexes would form at potential processing sites based primarily upon the relative affinities of the processing factors for those sites. Whether or not a site enjoyed an advantage because of its proximity to the promoter would depend upon its strength relative to other sites around it and upon the distance between sites. Relative differences in the efficiency with which active cleavage and polyadenylation complexes formed would determine the pattern of polyadenylation, and all sites at which productive complexes did form would be processed. Hence, in the absence of high affinity sites, lower affinity sites could account for the bulk of the polyadenylated transcripts. The patterns of polyadenylation would change, therefore, according to the nature of the sites present and the distance between them. This

'competition' model explains most of our observations, including the apparent dependence for first site utilization on the strength of the third, the ability of the first site to be predominant in the absence of a consensus third site, the lack of utilization of the second site when both the first and third sites are consensus, and the fact that the same amount of polyadenylated transcript can be generated from either allele. That such competition can apparently occur over the considerable distances reported here, i.e. 1.2 kb, is also consistent with the great size variation known to exist among 3' terminal exons (42).

The model does not, however, readily explain how the second site can account for more transcript than the first site during expression of the 21 K allele and less transcript than the first site during expression of the 20 K allele. Two possible explanations for differential utilization of the second site are 1) the nucleotide sequence differences that do exist between the two alleles (19) may impart structural changes to the primary transcript that alter the affinity of the second site for processing factors or 2) the proximity of the second site to the primary processing sites of both alleles may enhance the opportunity for the second site to bind trans-acting factors concentrated at or near the primary sites. Finally, the model predicts that the bulk of all primary transcripts will be cleaved and polyadenylated only once and that the steady state pattern of polyadenylation reflects the inherent strength of each site, the distance between sites, and the relative rate with which processing at a given site compares to the rate of transcription. Resolution of these issues and verification of the model itself requires further experimentation.

ACKNOWLEDGEMENTS

We thank Arif Hussain, Scott Devine, Myounghee Yu, Jian-fong Ma for their valuable advice and discussions. We thank Carol Mckissick for general support. This work was supported in part by grant CA49538 from NIH to PWM and by SAPEC, SA.

REFERENCES

- Wickens, M. (1990) *Trends Biochem. Sci.* **15**, 277–281.
- Proudfoot, N. (1991) *Cell* **64**, 671–674.
- Wickens, M., and Stephenson, P. (1984) *Science* **226**, 1045–1051.
- Zarkower, D., Stephenson, P., Sheets, M. and Wickens, M. (1986) *Mol. Cell. Biol.* **6**, 2317–2323.
- Montell, C., Fisher, E.F., Caruthers, M.H. and Berk, A.J. (1983) *Nature* **305**, 600–605.
- Gil, A., and Proudfoot, N.J. (1984) *Nature* **312**, 473–474.
- McDevitt, M.A., Imperiale, M.J., Ali, H. and Nevins, J.R. (1984) *Cell* **37**, 993–999.
- Zarkower, D., and Wickens, M. (1988) *J. Biol. Chem.* **263**, 5780–5788.
- Mclauchlan, J., Gaffney, D., Lindsay Whitton, J., and Clements, J.B. (1985) *Nucleic Acids Res.* **13**, 1347–1368.
- Tosi, M., Young, R.A., Hagenbuchle, O. and Schibler, U. (1981) *Nucleic Acids Res.* **9**, 2313–2323.
- Setzer, D.R., McGrogen, M. and Schimke, R.T. (1982) *J. Biol. Chem.* **257**, 5143–5147.
- Wilusz, J., Pettine, S.M. and Shenk, T. (1989) *Nucleic Acids Res.* **17**, 3899–3908.
- Leff, S.E., Rosenfield, M.G. and Evans, R.M. (1986) *Annu. Rev. Biochem.* **55**, 1091.
- Perricand, M., LeMoullec, J.M., Tiollais, P. and Petterson, U. (1986) *Nature* **288**, 174.
- Rosenfeld, M., Amara, S.G. and Evans, R.M. (1984) *Science* **225**, 1315–1320.
- Danner, D., and Leder, P. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 8658–8662.
- Sasavage, N.L., Gillam, S., Woychik, R.P. and Rottman, R.M. (1982) *Proc. Natl. Acad. Sci.* **79**, 223–227.
- Morandi, C., Masters, J.H., Mattes, M. and Attardi, G. (1982) *J. Mol. Biol.* **156**, 583–607.
- Scotto, K.W., Yang, H., Davide, J.P. and Melera, P.W. (1992) *Nucleic Acid Res.* **20**, 6597–6604.
- Lewis, J.A., Davide, J.P. and Melera, P.W. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6961–6965.
- Biedler, J.L., Albrecht, A.M., Hutchison, D.J. and Spengler, B.A. (1972) *Cancer Res.* **32**, 153–161.
- Biedler, J.L., Albrecht, A.M. and Spengler, B.A. (1978) *Eur. J. Cancer* **14**, 41–49.
- Lewis, J.A., Kurtz, D.T. and Melera, P.W. (1981) *Nucleic Acids Res.* **9**, 1311–1322.
- Melera, P.W., Davide, J.P., Hession, C.A. and Scotto, K.W. (1984) *Mol. Cell. Biol.* **4**, 38–48.
- Melera, P.W., Davide, S.P. and Oen, H. (1988) *J. Biol. Chem.* **263**, 1978–1990.
- Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry* **18**, 5294–5299.
- Aviv, H., and Leder, P. (1972) *J. Mol. Biol.* **134**, 743.
- Zinn, K., DiMaio, D. and Maniatis, T. (1983) *Cell* **34**, 865–879.
- Yang, H. and Melera, P.W. (1992) *Biotechniques* **13**, 922–927.
- Sato, K., Ryoji, I., Back, K.-H. and Agarwal, K. (1986) *Mol. Cell. Biol.* **6**, 1032–1043.
- Selden, R.F. (1987) In Ausubel, F.M., Brent, R., Kingston, R., Moore, D.D., Seidman, J.G. and Struhl, K. (ed.), *Current Protocols in Molecular Biology*, Wiley, New York, Vol. 2, pp. 9.2.1–9.2.5.
- MacGregor, G., Mogg, A., Burke, J. and Caskey, T. (1987) *Somatic cell and molecular genetics* **13**, 253–265.
- Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) *Nucleic Acids Res.* **11**, 7631.
- Bradford, M.M. (1976) *Anal. Biochem.* **72**, 248–254.
- Hall, C.V., Jacob, P.E., Ringold, G.M. and Lee, F. (1983) *J. Mol. Appl. Genet.* **2**, 101.
- Technical manual: Altered Sites™ *in vitro* mutagenesis system. (1991) Promega Corp., Madison, WI.
- Caruthers, M.H. (1983) In Weisman, S.M. (ed.), *Methods of DNA and RNA sequencing*. Praeger, N.Y., pp. 1–21.
- Liu, X. and Mertz, J.E. (1993) *Nucleic Acids Res.* **21**, 5256–5263.
- Wigley, P.L., Sheets, M.D., Zarkower, D.A., Whitmer, M.E. and Wickens, M. (1989) *Genes Dev.* **10**, 1705–1713.
- Weiss, E.A., Gilmartin, G.M. and Nevins, J.R. (1991) *EMBO J.* **10**, 215–219.
- Denome, R.M. and Cole, C.N. (1988) *Mol. Cell Biol.* **8**, 4829–4839.
- Hawkins, J.D. (1988) *Nucleic Acids Res.* **16**, 9893–9908.