Nucleic Acids Research, 1992, Vol. 20, No. 24 6597–6604

Differential utilization of poly (A) signals between DHFR alleles in CHL cells

Kathleen W.Scotto, Honghao Yang[1], Joseph P.Davide[+] and Peter W.Melera[1,*]
Molecular Pharmacology Program, Memorial Sloan-Kettering Cancer Center, New York 10021 and [1]Department of Biological Chemistry, The Cancer Center and the Graduate Program in Molecular and Cell Biology, University of Maryland, Baltimore, MD 21201, USA

Received September 10, 1992; Revised and Accepted November 16, 1992 GenBank accession nos M83108 and M83109

ABSTRACT

The Chinese hamster cell line, DC-3F, is heterozygous at the DHFR locus, and each allele can be distinguished on the basis of a unique DNA restriction pattern, protein isoelectric profile and in the abundancy of the DHFR mRNAs it expresses. Although each allele produces four transcripts, 1000, 1650 and 2150 and 6000 nucleotides in length, the relative distribution of these RNAs differs for each; the 2150 nt mRNA represents the major (60%) species generated from one allele, while the 1000 nt mRNA is the major species generated from the other. The allele that predominantly expresses the 2150 nt transcript is preferentially overexpressed when DC-3F cells are subjected to selection in methotrexate. We have analyzed the 3′ ends of both DHFR alleles and have found that the three major mRNAs arise by readthrough of multiple polyadenylation signals. A four base deletion in one allele changes the consensus polyadenylation signal AAUAAA to AAUAAU, resulting in the utilization of a cryptic polyadenylation signal lying 21 bp upstream. Surprisingly, this mutation in the third polyadenylation signal appears to affect not only the utilization of this signal, but also the efficiency with which the first signal, located 1171 bp upstream from the third site, is utilized.

INTRODUCTION

The Chinese hamster lung cell line, DC-3F, is heterozygous at the dihydrofolate reductase (DHFR) locus (1). When selected with the antifolate methotrexate (MTX), drug-resistant sublines that amplify and overexpress one or the other DHFR allele can be readily isolated (2,3). These sublines have provided an excellent system in which to analyze the expression of individual DHFR alleles in an identical somatic cell nuclear environment, where differential expression would be expected to be mediated by cis-acting, rather than trans-acting, factors.

We have previously shown that the two DHFR alleles differ with respect to DNA restriction patterns, the biochemical properties of the enzymes they encode and in the relative steady-state levels of the multiple mRNAs which they each generate (1,4,5–8). A total of 42 MTX-resistant sublines have been isolated from the parental DC-3F cell line, which itself expresses both DHFR alleles (1, 5). Each drug resistant subline, however, overexpresses one or the other, but not both alleles, and in every case the mechanism of overexpression is gene amplification. Interestingly, thirty-two of the sublines overexpress the 21K allele (so called because the DHFR it encodes displays an apparent molecular weight of 21KD by SDS-PAGE), while only ten overexpress the 20K allele (5, McKissick and Melera, unpublished), suggesting a selective advantage of the one over the other.

The two drug-resistant derivatives of DC-3F studied most extensively are sublines DC-3F/A3 and DC-3F/MQ19, which amplify and overexpress the 21K, or the 20K allele, respectively (5, 5a). The DC-3F/A3 subline (and other indepedently derived sublines from DC-3F which overexpress the 21K allele, (5)) contains three major DHFR mRNAs, 2150, 1650 and 1000 nt in length (7), the most abundant (60%) of which is the 2150 nt mRNA (1). The DC-3F/MQ-19 subline and other indepedently derived 20K allele overexpressors, contain the same three major DHFR mRNAs (7). However, in these cases the most abundant transcript (60%) is the 1000 nt species (1, 4). Since alterations in the processing of DHFR transcripts could play a role in mediating the selective advantage of one allele over the other, we have extended the analysis of these polymorphic DHFR genes to defining mutations in cis elements which may contribute to differences in the relative efficiency with which each allele processes its transcripts.

We had previously isolated two CHL cDNA clones, representing the 1000 nt DHFR mRNA generated by the 20K allele and one of the larger species generated by the 21K allele (7), and have shown that they differ in the length of their 3′ non-translated regions; we now present further analysis of clones representing the 3′end of the two DHFR alleles which verifies that this is a result of readthrough of multiple polyadenylation signals, similar to what has been observed with DHFR genes in human (9) and mouse (10, 11, 12). What is unique to the hamster system, however, is that signals which are identical in

* To whom correspondence should be addressed

+ Present address: Merck, Sharp and Dohme Research Laboratories, West Point, PA 19486, USA

both alleles are used with strikingly different efficiency. In this report, we show that a 4-base deletion which alters the 3'-most polyadenylation signal in the 20K allele by a single nucleotide not only inactivates that signal, but also appears sufficient to alter the utilization of another signal which lies 1171 nucleotides upstream.

MATERIALS AND METHODS

Cell lines and culturing

The characterization and maintenance of the Chinese hamster lung cell line, DC-3F, and its independently-derived methotrexate-resistant sublines DC-3F/A3, DC-3F/MQ19 and DC-3F/MQ8 have been described (2,3). Cells were grown in MEM/F12 (GIBCO) supplemented with 5% fetal calf serum and the appropriate concentration of MTX.

RNA isolation; Northern analysis and DNA sequencing

Total cellular RNA was extracted by the guanidinium isothiocyanate/cesium chloride gradient method (13). Northern blot analyses were performed as described previously (1). Probes were radioactively labelled using the random primer protocol (14) or by nick-translation (13). Poly $(A)^+$ RNA was prepared by oligo d(t) chromatography of polysomal RNA (6). DNA sequencing was carried out according to Maxam and Gilbert (15) with modifications (7) and Sanger and Coulson (16) as described in Devine et. al. (17).

Genomic library preparation and screening

A genomic library of DC-3F/A3 was constructed in lambda 1059 (18) using sucrose density gradient purified 20–25 Kb DNA fragments obtained from Bam HI partial digests of nuclear DNA (19). The library was plated in *E.coli* P358 and screened as described by Benton and Davis (20) using the 3' non-translated region of the DHFR cDNA clone, pDHFR A3-35, as probe (19).

Antisense RNA probe preparation by PCR and RNase protection assays

All reagents for oligonucleotide synthesis were from Applied Biosystem. Taq DNA polymerase and deoxynucleotides for PCR were supplied with the GeneAmp PCR Reagent Kit by Perkin-Elmer Cetus. Ribonuclease T2, Sp6 and T7 RNA polymerases, RNasin and RNase-free DNase I were either from Promega or Bethesda Research Laboratories. Yeast RNA was purchased from Sigma.

Pairs of primers whose sequences were complementary to regions upstream and downstream of the respective polyadenylation signals were designed for each PCR amplification, with either an Sp6 or T7 RNA polymerase promoter sequence added to the 5'-end of the antisense primer (21,22). For example, to map the first poly (A) site of the DHFR primary transcript, the sense primer (5'-GCCACTCCCCAAAGTCATGC-3') was designed to complement the sequence 85 nucleotides upstream of the putative poly (A) signal, whereas the antisense primer (5'GAATT GGATTTAGGTGACACTATAGAATACGAATACCTGCTG-GGGAGCCACTTGAGGCCGCATGGGA-3') was complementary to the sequence 80 nucleotides downstream of that signal (the attached SP6 RNA polymerase promoter sequence is underlined). An additional four or five nucleotides were routinely attached to the 5'-end of the promoter sequences to facilitate polymerase binding. Also, six additional nucleotides (5'-GAATA-C-3' and 5'-GGGAGA-3', in the case of the Sp6 and T7 promoters, respectively) were added between the 3'-end of the promoter and the 5'-end of the specific antisense primer, such that the single stranded antisense RNA probe was routinely 6 nucleotides longer than the longest possible RNA fragment protected and, therefore, distinguishable. All oligonucleotide primers were synthesized by phosphoramidite chemistry (23) using an Applied Biosystems synthesizer model 380A and purified by use of Oligonucleotide Purification Cartridges (OPC) supplied by Applied Biosystems Corp. The primers used for the second poly (A) site were 5'-GCCTTTCCTATCTCAG-3' (complementary to nucleotides 721 through 736 of p21K-2.1) and 5'-AATTTAATACGACTCACTATAGGGAGAGAATAC-GGATATGAAGGTAGTGGAGGGAATGGGAGGG-3' (complementary to nucleotides 847 through 877 of p21K-2.1; the T7 polymerase recognition sequence is underlined). For the third site, the primers used were 5'-CTCTTCCCTATCCTTCA-GGC-3' (complementary to nucleotides 1183 to 1202 of p21K-2.1) and 5'-GAATTGGATTTAGGTGACACTATA-GAATACGGATTAGTGCTGCTCATAACTCTGGTC-3' (complementary to nucleotides 1376 through 1402 of p21K-2.1; the Sp6 polymerase promoter sequence is underlined).

PCR amplification was performed using the Perkin-Elmer Cetus GeneAmp PCR Reagent Kit. One ng cloned DNA was

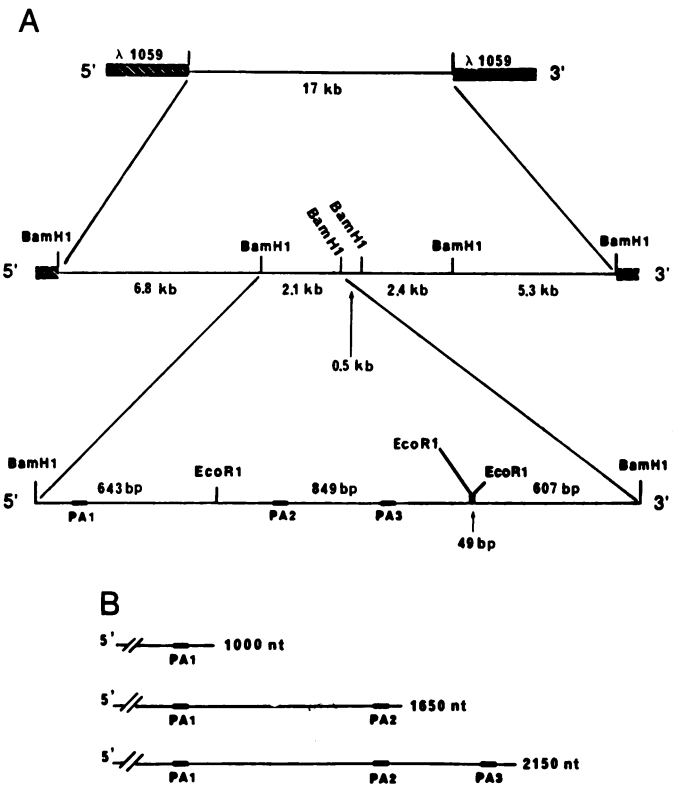


Figure 1. Diagram of λ 1059-DHFR-21K-15. **A.** Bam HI restriction map of the λ 1059-DHFR-21K-15 17 Kb insert and an Eco RI restriction map of the 2.1 Kb Bam HI fragment. Details of the mapping experiments can be found in reference 38. The 5' and 3' designations refer to the orientation of the 17 Kb genomic DNA fragment with respect to its DHFR transcripts. **B.** Approximate location of the polyadenylation sites utilized in the three different sized DHFR mRNAs produced by Chinese hamster cells as determined by preliminary S_1 nuclease experiments (41) and by analyzing the sequences of pDHFR MQ19-97 and A3-35 (7).

suspended in a total volume of 100 ul reaction buffer (50mM KCl, 10 mM Tris−HCl, PH 8.3, 1.5 mM $MgCl_2$, 0.01% (w/v) gelatin) which also contained 1.0 μM of each primer, 200 μM each of dATP, dCTP, dTTP and dGTP, and 2.5 units of Perkin-Elmer Cetus AmpliTaq DNA polymerase. The reaction was carried out for 35 cycles in a Perkin-Elmer Cetus Thermalcycler (the sample was denatured at 94°C for 1 min, annealed at 50−54°C for 2 min, and extended at 72°C for 3 min). The PCR-amplified DNA template was purified by filtration through a G-50 Sephadex spin column (Boehringer Mannheim Biochemicals) (24). Approximately 1 μg of the PCR-amplified DNA template was used for each *in vitro* transcription reaction, in a reaction mixture containing 40mM Tris−HCl (pH7.5), 6mM $MgCl_2$, 2mM spermidine, 10mM NaCl, 10mM DTT, 1 unit/μl of RNasin, 500μM of each rATP, rUTP, rGTP and 12μM rCTP, 50uCi radiolabeled α-^{32}P-rCTP (New England Nuclear Corp.) and 5−10 units of SP6 or T7 RNA polymerase, incubated at 40°C for 1hr. 10 units of DNase I was then added to remove the DNA template, and after digestion for 1 hr at 37°C the resulting RNA probe was purified by phenol extraction followed by ethanol precipitation.

10μg of total cellular RNA and 5×10^5cpm of the antisense probe were mixed in 30μl of hybridization buffer containing 80% formamide, 0.4M NaCl, 40mM PIPES (PH 6.4) and 1mM EDTA. The RNA was denatured at 85°C for 5 min, then allowed to reanneal at 35−45°C overnight. 30 units/ml of T2 RNase was added, and following digestion at 30° for 1 hr to remove single stranded unprotected RNA, the protected RNA-RNA hybrids were extracted, precipitated and electrophoresed through a 6 or 8% denaturing polyacrylamide gel with DNA sequence ladders as size markers. The gel was then fixed, dried and exposed to Kodak XAR-5 film overnight at −70°C with an intensifying screen.

RESULTS

Cloning and characterization of the 3′ ends of the two DHFR alleles

A genomic library was prepared from DC-3F/A3 cells and screened with pDHFR A35-700 (19), a cDNA clone containing sequences within the 3′ untranslated region of the DC-3F/A3-35 cDNA (7) derived from the 21K allele. A single genomic clone, designated lambda 1059/21K-DHFR-A15 (lambda A15) and representing the 21K allele, was isolated from 5×10^5 non-amplified colonies. Additional clones representing the 21K allele were later isolated from a DC-3F/A3 genomic library prepared in the vector EMBL-3 (Forma, F. and P.W.Melera), allowing us to confirm the results obtained with lambda A15.

The structure of lambda A15 and its subclones is shown in Figure 1. These results are consistent with the known structure of the CHO DHFR gene reported by Carothers et al (25) and Milbrandt et al (26). The five BamHI fragments were subcloned into the pUC8 vector and named according to the allele from which they were derived and their size (i.e. p21K-6.8, p21K-2.1, etc). Southern blot analysis of DC-3F and DC-3F/A3 DNA using each of these five fragments as probe verified that four were derived from DNA which was amplified in the drug-resistant cell line. The same pattern of amplified fragments was also observed in DC-3F/MQ19 DNA, indicating that no major rearrangements occurred between the two alleles in this region (data not shown). The insert of p21K-5.3, however, contained no apparent amplified DNA sequences, consisting instead of both unique and

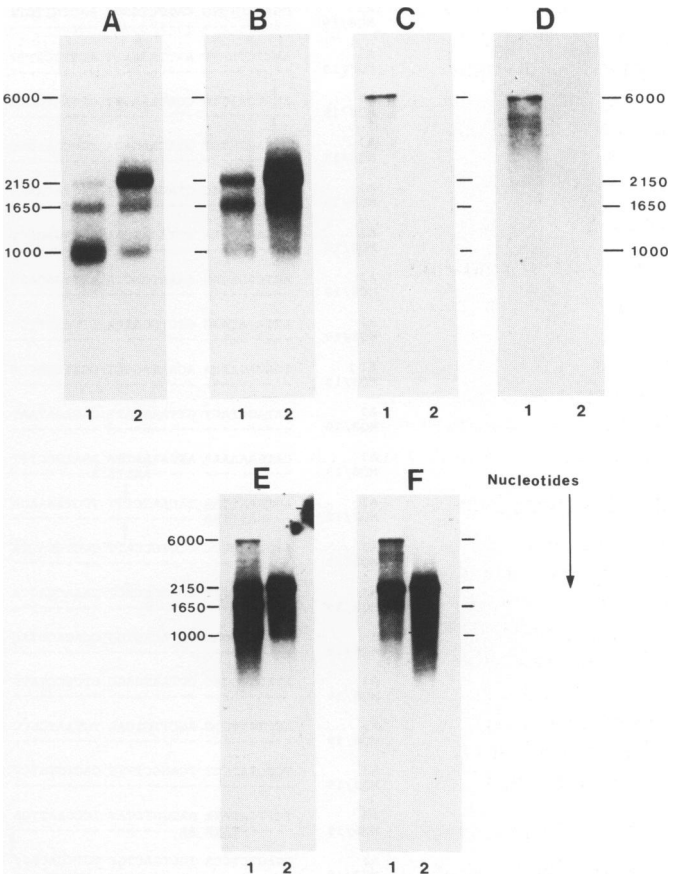


Figure 2. Northern blot analysis of DHFR transcripts. Polysomal poly $(A)^+$ RNA prepared from DC-3F/MQ19 and DC-3F/A3 cells that overexpress the 20K and 21K alleles 120- and 50-fold respectively was hybridized with the four amplified Bam HI genomic DNA fragments obtained from λ 1059 21K-15 (fig. 1). In each case, 1 μg of poly $(A)^+$ polysomal RNA was electrophoresed through a 1.4% formaldehyde gel at 40 volts for 14 hours. After transfer to nitrocellulose, the resulting filters were hybridized with 100 ng of probe labeled with [^{32}P] to a specific activity of 3×10^8 cpm. After 16 hr at 65°C the filters were washed and then exposed to Kodak XAR film for 24 hours. In each panel, lane 1 contains RNA from DC-3F/MQ19 cells 20K allele) and lane 2 contains RNA from DC-3F/A3 cells 21K allele). The probes used in each experiment were: Panel A, p21K-6.8; Panel B, p21K-2.1; Panel C, P21K-0.5; Panel D, p21K-2.4 Overexposure of the blots in Panel A and B show the presence of the 6 Kb transcript (not shown). Panels E and F show the results of a 36 hr exposure of the same RNA samples hybridized with p21K-643 and p21K-849 respectively.

repetitive sequences common to both the parental and resistant cells. Although this most likely represents a cloning artifact, the possibility that the p21K-5.3 represents the 3′ insertion site for the DHFR amplicon cannot be ruled out.

We have previously described the isolation and characterization of the cDNA clones, pDHFR MQ19-97 and pDHFR A3-35 (7) derived from the 20K and 21K alleles, respectively. pDHFR MQ19-97 ends in a tract of adenosine residues, preceded by the polyadenylation signal AUUAAA, and represents all but the 5′ leader sequence of the 1000nt mRNA generated by the 20K allele. pDHFR A3-35 contains a DHFR-specific insert which extends approximately 450 nt beyond this site of polyadenylation despite the presence of the same signal in the same location as it is found in the 20K allele, and therefore was derived from one of the larger mRNAs generated from the 21K allele. Southern blot analysis of lambda A15 using both these clones as probes showed

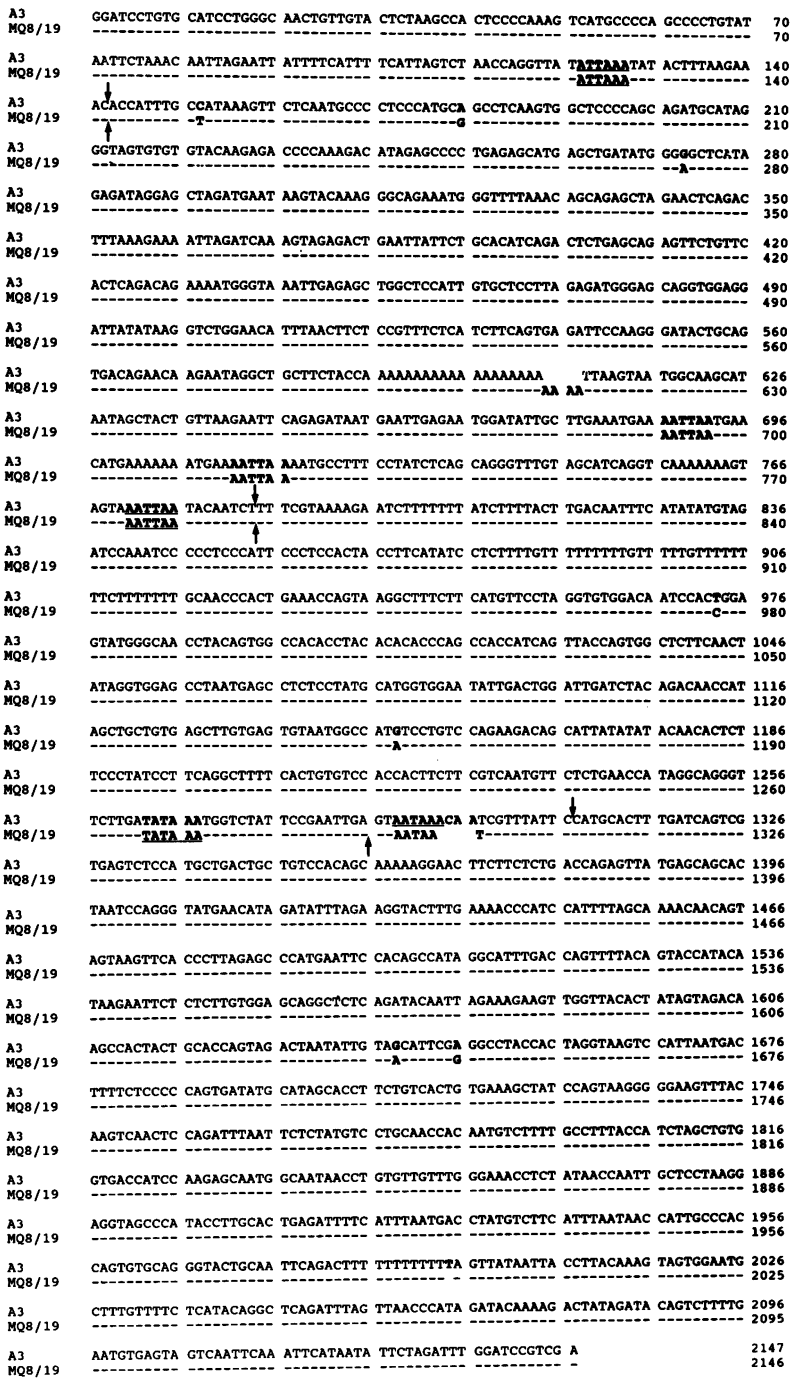


Figure 3. Nucleotide sequence comparison of p20K-2.1 and p21K-2.1. The nucleotide sequences of both the 20K and 21K allele 2.1 Kb Bam HI genomic DNA fragments are shown, as is the sequence of a second 20K allele 2.1 Kb Bam HI fragment cloned from the cell line DC-3F/MQ8. The sequences presented for DC-3F/MQ19 and DC-3F/A3 were derived from two independent clones in each case and each clone was sequenced in both directions. The DC-3F/MQ8 sequence was determined from one clone sequenced in one direction only. As indicated, the DC-3F/MQ19 and DC-3F/MQ8 sequences are identical. The sequence differences between the 2 alleles are indicated in bold type as are the potential polyadenylation signals revealed by preliminary S_1 nuclease assays (41) and by comparisons to cloned cDNAs (7). Those signals underlined are the ones utilized in each allele (see fig. 4). The major polyadenylation sites for each allele are indicated by the arrow heads.

hybridization only to the 6.8 Kb and 2.1 Kb Bam HI fragments (19).

The four plasmids containing Bam HI fragments and representing amplified DNA were used in Northern blot analyses of RNA from DC-3F/A3 and DC-3F/MQ19 cells (Figure 2). During the course of these analyses, an additional minor DHFR mRNA species, approximately 6 Kb in length and co-purifying with poly $(A)^+$ RNA, was identified, but only in DC-3F/MQ19 cells. The three major DHFR mRNAs, as well as the 6 Kb species, hybridize with p21K-6.8 and p21K-2.1; p21K-0.5 and p21K-2.4 only recognize the 6 Kb transcript. Taken together, these results support the notion that the four RNA species differ

by virtue of extensions of their 3' ends, and indicates that the termini of the three major mRNAs are located within p21K-2.1.

In order to further delineate the genomic region which contains the ends of the three major DHFR mRNAs, p21K-2.1 was subcloned into three fragments (p21K-643, p21K-849 and p21K-607 (Figure 1A). Northern blot analyses using these subclones as probe showed that the 3'-end of the 1000 nt species lies within p21K-643 while the ends of the 1650 and 2150 nt mRNAs lie within p21K-849. p21K-607 hybridizes only to the 6 Kb RNA species (data not shown). Southern blot analysis of DNA representing the two alleles revealed an RFLP when DNA was digested with Hpa II and probed with p21K-849; a restriction site for this enzyme is present within this region of the 20K allele, but absent in the 21K allele (19). This observation was fortuitous, since it allowed us to clearly distinguish the two alleles during further analyses.

In order to clone the corresponding 2.1 kB Bam HI fragment from the 20K allele, we electrophoresed Bam HI- digested genomic DNA from DC-3F/MQ19 cells, isolated and cloned fragments migrating from 1.9 to 2.2 Kb, and identified DHFR-containing clones by hybridization with p21K-2.1. Several 20K allele-specific clones (designated p20K-2.1) were obtained by this method and subject to further characterization, as were similar clones obtained from the independently derived cell line DC-3F/MQ8 , which was also known to amplify the 20K allele (ca. 180-fold (5)) (these clones were designated pMQ8-2.1). Northern hybridization analyses using p20K-2.1 and its subclones as probe yielded the same pattern as that seen when p21K-2.1 was used (19, and see fig. 2).

Nucleotide sequence analysis of p21K-2.1 and p20K-2.1

The nucleotide sequences of p21K-2.1 and p20K-2.1 (as well as an additional p20K-2.1 clone from the subline DC-3F/MQ8 that amplifies and overexpresses the 20K allele (1, 5) and is identical to p20K-2.1) are shown in Figure 3; only those nucleotides which differ in the 20K allele are shown. Comparison with the sequence of the full length cDNA, pDHFR-MQ19-97, allowed us to locate the poly (A) addition site for the 1000 nt mRNA to nucleotide 142 or 143 within the 2.1 Kb fragment, 15 bases downstream from the polyadenylation signal AUUAAA. Analysis of the region represented by p21K-849, and therefore containing the cleavage sites of the 1650 and 2150 nt species, revealed the presence of several potential polyadenylation signals for these two mRNAs.

Relatively few sequence differences were noted between the 20K and 21K alleles; using the 21K allele as reference, there were seven point mutations (positions 152, 180, 273, 973, 1149, 1639 and 1646), one insertion (AAAA at position 609) and two deletions (ACAA, present at position 1294 and T at position 1995) in the 20K allele. As a result of these changes, p20K-2.1 is 2146 nucleotides in length, one base shorter than p21K-2.1. There are several AT-rich regions within the sequence, and the addition of AAAA at position 609 in the 20K allele results in an uninterrupted tract of 23 adenosine residues (19 in the 21K allele). It should also be noted that the removal of the sequence ACAA from the 20K allele changes the potential polyadenylation signal AAUAAA to AAUAAU. The significance of this mutation will be discussed.

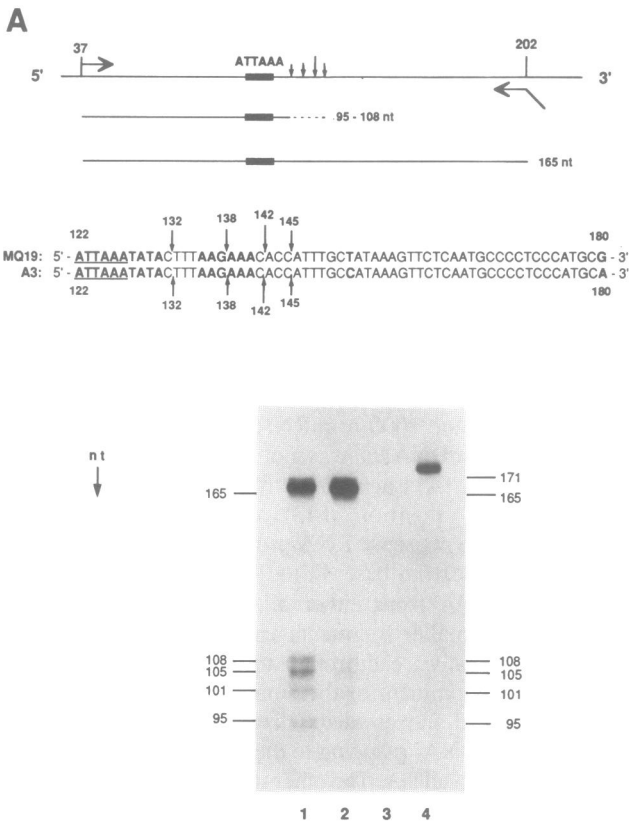
Finally, there is a sequence, 5'-CGTGagTC-3', located 31 nucleotides 3'-ward of the third potential poly (A) signal at position 1325 which is similar to the consensus sequence 5'-YGTGTTYY-3' that has been reported to be required for polyadenylation (27). No similar consensus sequence exists downstream of either the first or second potential poly (A) signals indicated in fig. 3, although an extremely U-rich region does lie downstream of the second signal, beginning at position 800.

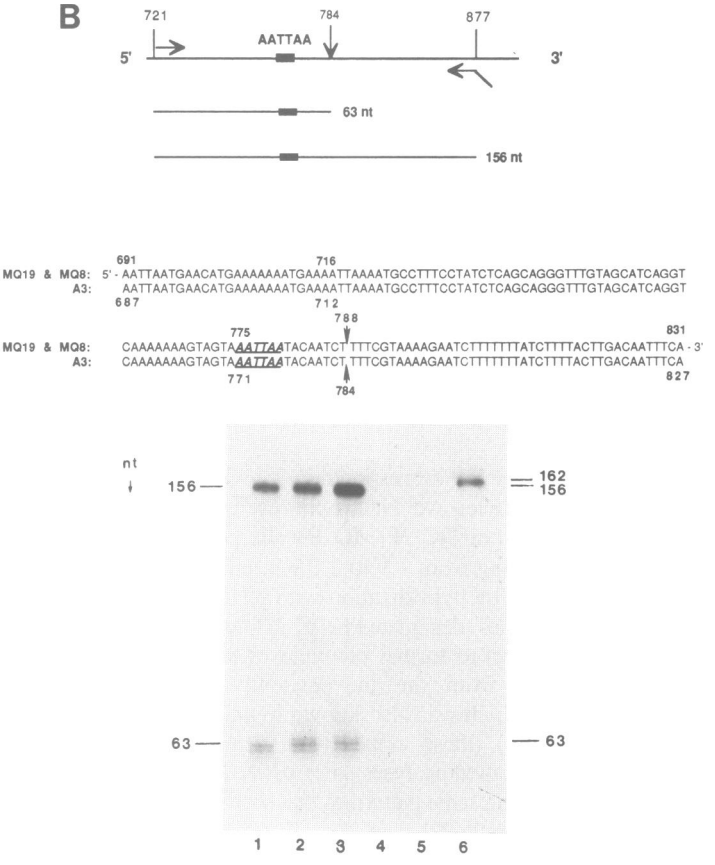
RNase protection analysis of the 3' termini of the major DHFR mRNAs

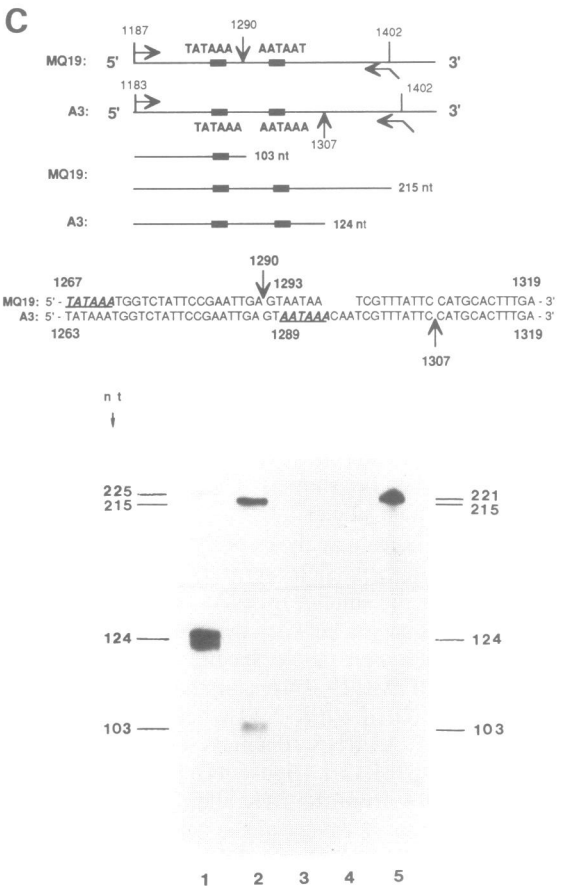
In order to define the 3' termini of the major mRNA species generated by the two DHFR alleles, RNase protection assays were performed on total cellular RNA isolated from DC-3F/A3, DC-3F/MQ19, and in some cases, DC-3F/MQ8 cells (figure 4). Probes for these analyses were generated by PCR amplification of various regions of either pA3-2.1 or p20K-2.1 (Materials and Methods). Although we were able to approximate the 3' poly (A) cleavage site of the 1000 nt mRNA based on comparisons with the full-length cDNA, it was not clear whether the first adenosine in the poly (A) tract was derived from genomic DNA (the adenosine at position 143) or was added post-transcriptionally. The antisense RNA probe used for this analysis represents sequences from base 37 to 202 of p20K-2.1. When hybridized with RNA from either allele (fig 4A), a 165 nt fragment and four smaller fragments are protected. The 165 nt fragment represents the readthrough transcript in both cases, while the most predominant of the smaller fragments (105 nt) reflects the cleavage site predicted from the sequence of the DC-3F/MQ19-97 cDNA, mapping to the 3' side of the C residue at position 142 in both alleles. The other three protected fragments either represent artifacts of the RNAse protection assay or indicate the presence of multiple cleavage sites. Attempts to eliminate this heterogeneity by altering the assay conditions or by altering the length of the antisense RNA probe have been unsuccessful. Since further inspection of the nucleotide sequence surrounding the 5'-AUUAAA-3' signal (fig 3) reveals the presence of two other potential poly (A) signals (5'-AAUAUA, overlapping the AUUAAA signal, and AAGAAA, 9 nt downstream of the signal), it is likely that heterogeneity exists in the 3' processing at this site. Nevertheless, this heterogeneity is identical between the two alleles.

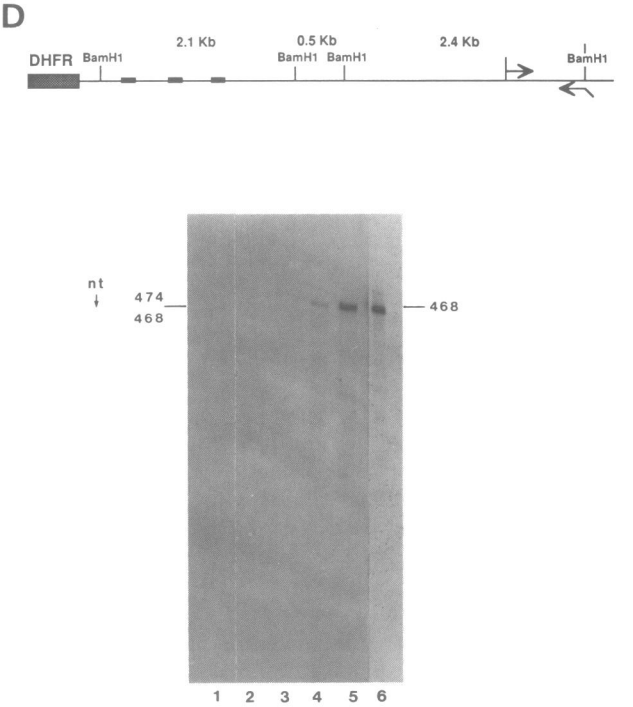
The probe used for analysis of the cleavage site of the 1650 nt mRNA contained sequences from base 721 to 877 of p21K-2.1 (Figure 4B), and protected a 156 nt readthrough RNA fragment and an additional fragment of 63 nt that represents the cleavage site for production of the 1650 nt DHFR mRNA (fig. 4B). This indicated that the polyadenylation signal AAUUAA located at position 771 in p21K-2.1 (775 in the 20K allele) is utilized. Similar analyses using probes spanning the identical signals at positions 712 (716) and 687 (691) had demonstrated that they are not utilized (data not shown). The same protected fragments were found for both alleles.

Results obtained from analyses of the termini of the 2150 base mRNA encoded by each allele were strikingly different than what was observed for the smaller mRNAs. Because of the sequence differences in this region between the two alleles, the probes used for these analyses were allele specific. Hence, the probe for analysis of DC-3F/MQ19 RNA contained sequences from base 1187 to base 1402 of p20K-2.1, whereas the probe used for analysis of DC-3F/A3 RNA contained sequences from base 1183 to base 1402 of p21K-2.1 (figure 4C). Although the fragment protected by RNA from the 21K allele (lane 1) was 124 nt in length, representing a cleavage site at position 1307, the fragment


A
37
202
5'
ATTAAA
3'
95 - 108 nt
165 nt
122
132
138
142
145
180
MQ19: 5' - ATTAAATATACTTTAAGAAACACCATTTGCTATAAAGTTCTCAATGCCCCTCCCATGCG - 3'
A3: 5' - ATTAAATATACTTTAAGAAACACCATTTGCCATAAAGTTCTCAATGCCCCTCCCATGCA - 3'
nt
165
171
165
108
105
101
95
1 2 3 4


B
721
784
877
5'
AATTAA
3'
63 nt
156 nt
691
716
MQ19 & MQ8: 5' - AATTAATGAACATGAAAAAAATGAAAATTAAAATGCCTTTCCTATCTCAGCAGGGTTTGTAGCATCAGGT
A3: AATTAATGAACATGAAAAAAATGAAAATTAAAATGCCTTTCCTATCTCAGCAGGGTTTGTAGCATCAGGT
687
712
775
788
831
MQ19 & MQ8: CAAAAAAAGTAGTAAATTAATACAATCTTTTCGTAAAAGAATCTTTTTTTATCTTTTACTTGACAATTTCA - 3'
A3: CAAAAAAAGTAGTAAATTAATACAATCTTTTCGTAAAAGAATCTTTTTTTATCTTTTACTTGACAATTTCA
771
784
827
nt
156
162
156
63
63
1 2 3 4 5 6


C
MQ19: 5'
1187
1290
TATAAA
AATAAT
1402
3'
A3: 5'
1183
1402
TATAAA
AATAAA
1307
3'
103 nt
MQ19:
215 nt
A3:
124 nt
1290
1267
1293
1319
MQ19: 5' - TATAAATGGTCTATTCCGAATTGAGTAATAA TCGTTTATTC CATGCACTTTGA - 3'
A3: 5' - TATAAATGGTCTATTCCGAATTGAGTAATAACAATCGTTTATTCCATGCACTTTGA - 3'
1263
1289
1307
1319
nt
225
215
221
215
124
124
103
103
1 2 3 4 5


D
DHFR
BamH1
2.1 Kb
0.5 Kb
BamH1
BamH1
2.4 Kb
BamH1
nt
474
468
468
1 2 3 4 5 6

protected by the 20K allelic RNA (lane 2) was only 103 nt long, indicating that cleavage occurred at base 1290. It is interesting to note, then, that the consensus polyadenylation signal, AAUAAA, which is located 19 bp upstream of the cleavage site in the 21K allele, is changed to AAUAAU in the 20K allele due to a four base (ACAA) deletion. Apparently, this mutation severely alters the efficiency of the polyadenylation signal, resulting in the utilization in the 20K allele of a cryptic upstream poly (A) addition site, which is preceded by a putative polyadenylation signal, UAUAAA, 24bp upstream (see the diagram and the sequence in fig. 4C). Indeed, preliminary results from transfection experiments utilizing a construct in which the p20K-2.1 or p21K-2.1 BamH1 fragments have been cloned downstream of a reporter gene have confirmed the cis-acting effect of this deletion (Yang and Melera, unpublished). The 215 nt fragment in lane 2 represents the minor 6 Kb DHFR transcript detected by the northern blots shown in fig. 2, which was observed only in RNA generated by the 20K allele. To confirm that this transcript extends beyond the third poly (A) site in this allele, we prepared an additional 468 nt antisense RNA probe from the downstream terminus of the p20K-2.4 clone indicated in fig. 1. As predicted, this probe protects a 468 nt fragment only in the RNA produced from the 20K allele (fig 4D). The 3′ terminus of the minor 6 Kb transcript appears to lie within a region further 3′ of the amplified sequences present within lambda A15.

One possible explanation for the difference in abundance of the three major DHFR mRNAs generated by each allele is differential pausing or termination of transcription, resulting in a real or temporal difference in availability of the first signal. However, nuclear run-on analysis of newly synthesized RNA from both the DC-3F/A3 and DC-3F/MQ19 cell lines indicated that synthesis proceeded similarly across both alleles and beyond the third polyadenylation signal (data not shown).

DISCUSSION

The majority of eukaryotic mRNAs are polyadenylated (28). Three cis elements have been found to be involved in this process: 1) the polyadenylation signal, AAUAAA, which lies 6−30 bases 5′ to the site of polyadenylation (29−33), 2) the poly (A) addition site, which is usually occupied by an adenosine and often preceeded by a cytosine (34, 35), and 3) a downstream U or UG-rich region (29, 36−43). In addition, a number of trans-acting proteins and RNA factors have been implicated (for review and additional references, see 44−46).

By ribonuclease protection mapping, the sites of poly (A) addition, as well as the sequences signalling polyadenylation for the three major Chinese hamster DHFR mRNAs, have been localized within analogous 2.1 Kb Bam HI fragments within the two polymorphic DHFR alleles of DC-3F cells. For both alleles, polyadenylation of the smallest mRNA species is heterogenous, apparently due to the close proximity of multiple polyadenylation signals. Surprisingly, however, although there is no difference between the two alleles in the signals present or the potential polyadenylation sites recognized, these sites, as a group, are utilized with efficiencies that differ strikingly between the two alleles, representing the major site of polyadenylation in the 20K allele (60%) and the minor site (10%) in the 21K allele. Furthermore, although two single nucleotide sequence differences do exist between the alleles in this region, *in vitro* mutagenesis and preliminary transfection studies have shown that conversion of the T at position 152 (+10 relative to the poly(A) site at position 142) to a C (fig 4A) does not decrease the efficiency of polyadenylation at the first site, nor does it alter the heterogeneity of the cleavage pattern (Yang and Melera, in preparation). It is also unlikely that the G to C transition at position 180 is critical, since it is relatively distant from the major poly(A) addition site and does not lie within a GU-rich region or a GU cluster (fig 3). We conclude, therefore, that differential utilization of the first site is not dependent upon sequence differences between the two alleles within this region.

The 1650 nt mRNA was cleaved and polyadenylated at the third of three closely spaced potential sites, all preceded by the sequence AAUUAA, which has been shown to be a very weak signal *in vitro*, representing only 0.5% of those used in natural mRNAs (47). Mutation of the third major signal, responsible for polyadenylation of the 2150 nt mRNA, from AAUAAA to AAUAAU, eliminated use of this signal entirely in the 20K allele, and exposed a cryptic upstream signal, UAUAAA, which is utilized inefficiently, consistent with its relative strength *in vitro* (~17%) (47).

Although it appears that the utilization of the first signal is significantly altered by a mutation in the third signal, it is not readily apparent how this occurs. A number of models have been proposed to explain the differential utilization of multiple polyadenylation signals, all of which take into account the inherent efficiency of the polyadenylation signal as well as its relative

Figure 4. Mapping of polyadenylation sites. Antisense RNA probe production, hybridization, RNase digestion conditions and autoradiographic procedures are described in Materials and Methods. In each panel the horizontal arrows indicate the positions of the primers used for PCR amplification of the DNA from which antisense RNA was then synthesized *in vitro*. Due to the nature of the promoter constructs used, the antisense RNA probes were routinely six nucleotides longer than the longest RNA fragment protected. The large vertical arrows indicate the sites of polyadenylation. The polyadenylation signals used for each transcript are shown in both the diagrams and in the sequences presented. In the latter, only the signal utilized in each allele is italicized and underlined. The numbers shown in both the diagrams and the sequences refer to the nucleotide positions indicated in fig. 3. For panel A, the probe used was generated from p20K-2.1. For panels B and D the probes used were generated from p21K-2.1. For panel C the probe used for analysis of DC-3F/A3 RNA was also generated from p21K-2.1, whereas the probe used for analysis of DC-3F/MQ19 RNA and the rest of the RNAs shown was from p20K-2.1. Each lane contains 5×10^5 cpm of the appropriate radiolabeled antisense RNA probe plus the indicated amount of total cellular RNA. Panel A. Lanes 1 and 2 contain 10 μg RNA; Lane 1, DC-3F/MQ19; lane 2, DC-3F/A3; lane 3, tRNA; lane 4, undigested probe. It should be noted that based upon the Northern blot analysis of poly(A)$^+$ RNA in lane 1 of Panel A, fig 2, the amount of signal generated by the 1650 and 2150 nt transcripts from DC-3F/MQ19 cells represented by the 165 nt protected fragment in lane 1, is higher than expected. Subsequent Northern blots analysis of the total RNA preparations used for the RNase protection assays (data not shown) has indicated that relatively more 1650 not mRNA is present in total RNA preparations than in poly (A)$^+$ preparations from DC-3F/MQ19 cells. The reason for this discrepancy is under investigation. Panel B. Lanes 1−5 contain 10 μg RNA. Lane 1, DC-3F/A3; lane 2, DC-3F/MQ19; lane 3, DC-3F/MQ8; lane 4, DC-3F; lane 5, tRNA, lane 6, undigested probe. Panel C. Lanes 1−5 contain 10 ug RNA. Lane 1, DC-3F/A3; lane 2, DC-3F/MQ19; lane 3, DC-3F; lane 4, tRNA; lane 5, undigested 221 nt probe from p20K-2.1. Residual undigested 225 nt probe from p21K-2.1 used to analyze DC-3F/A3 RNA can be seen in lane 1. Panel D. Lanes 1−5 contain 10 μg RNA. Lane 1, tRNA; lane 2, DC-3F; lane 3, DC-3F/A3; lane 4, DC-3F/MQ19; lane 5, DC-3F/MQ8; lane 6, undigested probe.

position with respect to the promoter (48), and can be distinguished on the basis of directionality of template scanning and signal recognition. If scanning occurred in a 5′ to 3′ direction, as suggested by the most well-documented model, the recognition and utilization of the first signal would be expected to be independent of a mutation lying over 1 kb downstream, since the choice of sequential polyadenylation signals would depend both on the intrinsic strength of the signal and its relative 5′ to 3′ position.

A second model evokes a 3′ to 5′ scanning mechanism, which was proposed by Denome and Cole (48) to account for the utilization of multiple polyadenylation signals in constructs assayed by transient transfection; this model could explain our data if, when scanning reached the first signal, more RNA was available from the 20K allele than the 21K allele, due to less efficient utilization of the mutated third signal in the 20K allele.

In the third model, the polyadenylation signals are recognized by a soluble factor(s) which interacts with the initial transcript shortly before or following termination. In this case, utilization of signals would rely more heavily on their intrinsic strength (i. e., affinity for the factor) than on their relative position, although the promoter proximal sites may retain some temporal advantage, since they are synthesized earlier. In order for this model to explain our observations, it must be assumed that any given transcript can only be cleaved once and that the third signal in the 21 K allele, AAUAAA, is intrinsically more efficient than the first signal, AUUAAA (consistent with the observations of others (47)). Therefore, elimination of the third signal in the 20K allele would allow the first signal to predominate. Although we cannot yet determine which of these models best describes the mechanism of polyadenylation in the DHFR system, perhaps the most intriguing conclusion derived from these studies is that all models require that the advancing polymerase complex reaches the third signal before preferential utilization of the first site occurs.

Our attempts to understand the molecular basis for the different patterns of DHFR mRNA expression in CHL cells that amplify and overexpress either the DHFR 20K or 21K allele have led us to study 3′-end formation in DHFR genes which are known to have multiple polyadenylation sites. Experiments in progress are aimed at determining the mechanism governing polyadenylation site selection in these genes.

ACKNOWLEDGEMENTS

This work was supported in part by grants from NIH and NSF to PWM and by additional support from SAPEC, SA. The authors wish to thank Felicia Forma for excellent technical assistance.

REFERENCES

1. Lewis,J.A., Davide,J.P. and Melera,P.W. (1982) *Proc. Natl Acad. Sci. USA.* **79**,6961–6965.
2. Biedler, J.L., Albrecht,A.M., Hutchison,D.J. and Spengler,B.A. (1972) *Cancer Res.*, **32**,153–161.
3. Biedler,J.L., Albrecht,A.M. and Spengler,B.A. (1978) *Eur. J. Cancer*, **14**,41–49.
4. Lewis,J.A., Kurtz,D.T. and Melera,P.W. (1981) *Nucleic Acids Res.* **9**,1311–1322.
5. Melera,P.W., Hession,C.A., Scotto,K.W., Biedler,J.L., Meyers,M.B. and Shanske,S. (1982) *J. Biol. Chem.* **257**,12939–12949.

5a. Melera,P.W., Lewis,J.A., Biedler,J.L. and Hession,C.A. (1980) *J. Biol. Chem.* 255, 7024–7028.

6. Melera,P.W., Wolgemuth,D., Biedler,J.L. and Hession,C.A. (1980) *J. Biol. Chem.* **255**,319–322.
7. Melera,P.W., Davide,J.P., Hession,C.A, and Scotto,K.W. (1984) *Mol. Cell. Biol.* **4**,38–48.
8. Melera,P.W., Davide,S.P. and Oen,H. (1988) *J. Biol. Chem.* **263**,1978–1990.
9. Morandi,C., Masters,J.H., Mattes,M. and Attardi,G. (1982) *J. Mol. Biol.* **156**,583–607.
10. Setzer,D.R., MacGrogan,M., Numberg,J.H. and Schimke,R.T. (1980) *Cell.* **2**,361–370.
11. Setzer,D.R., MacGrogan,M. and Schimke,R.T. (1982) *J. Biol Chem.* **257**,5143–5147.
12. Yen,J.Y. and Kellems,R.E. (1987) *Mol. Cell. Biol.* **7**,3732–3739.
13. Maniatis,T., Frisch,E.F. and Sambrook,J. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory.
14. Feinberg,A.P.and Vogelstein,B. (1983) *Anal. Biochem.* **132**,6–13.
15. Maxam,A.M. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**,560–564.
16. Sanger,F. and Coulson,A.R. (1975) *J. Mol. Biol.* **94**,441–448.
17. Devine,S.E., Hussain,A., Davide,J.P.and Melera,P.W. (1991) *J. Biol. Chem.* **266**,4545–4555.
18. Karn,J.S., Bremer,L., Barnett,L. and Cesareni,G. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**,5172–5176.
19. Scotto,K.W. (1983) Characterization of DHFR genes in methotrexate-resistant Chinese hamster lung cells. Doctoral Dissertation, Cornell University Graduate School of Medical Sciences, New York, NY.
20. Benton,W.R. and Davis,R.W. (1977) *Science*, **196**,180–181.
21. Chamberlin,M.and Ryan,T. (1982) *The Enzymes* **XV**,87–108.
22. Kassevetis,G.A., Butler,E.T., Roulland,D. and Chamberlin,M.J. (1982) *J. Biol. Chem.* **257**,5779–5787.
23. Caruthers,M.H. (1983) In Weisman,S.M.(ed.), Methods of DNA and RNA Sequencing. Praeyer, N.Y., N.Y., pp.1–21.
24. Maniatis,P., Jeffery,A. and Kleid,D.G. (1975) *Proc. Natl. Acad. Sci. U.S.A.* 72,1184–1187.
25. Carothers,A.M.,Urlaub,G.,Ellis,N.and Chasin,L.A. (1983) *Nucleic Acids Res.* **11**,1997–2012.
26. Milbrandt,J.D., Azizkhan,J.C., Greisen,K.S. and Hamlin,J.L. (1983) *Mol. Cell Biol.* **3**,1266–1273.
27. McLauchlan,J.,Gaffney,D., Whitton,J.L. and Clements,J.B. (1985) *Nucleic Acid Res.* **13**,1347–1368.
28. Nevins,J.R., and Darnell,Jr.J.E. (1978) *Cell*, **15**,1477–1493.
29. Fitzgerald,M.and Shenk,T. (1981) *Cell.* **24**:251–260.
30. Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature (London)* **263**,211–214.
31. Montell,C., Fisher,E.F., Caruthers,M.H. and Berk,A.J. (1983) *Nature (London).* **305**,600–605.
32. Wickens,M. and Stephenson,P. (1984) *Science.* **206**,1045–1051.
33. Zarkower,D., Stephenson,P., Sheets,M. and Wickens,M. (1986) *Mol. Cell. Biol.* **6**,2317–2323.
34. Moore,C.L., Skolnik-David,H. and Sharp,P.A. (1986) *EMBO J.* **5**,1929–1938.
35. Sheets,M.D., Stephenson,P. and Wickens,M. (1987) *Mol. Cell. Biol.* **7**,1518–1529.
35. Proudfoot,N.J. (1989) *TIBS* **14**,105–110.
36. Cole,C.N.and Stacy,T.P. (1985) *Mol. Cell Biol.* **5**,2104–2113.
37. Conway,L.and Wickens,M. (1985) *Proc. Natl Acad Sci, USA.* **82**,3949–3953.
38. Gil,A.and Proudfoot,N.J. (1984) *Nature (London)* **312**,473–474.
39. McDevitt,M.A., Imperiale,M.J. Ali,H. and Nevins,J.R. (1984) *Cell*, **37**,993–999.
40. Qian,Z. and Wilusz,J. (1991) *Mol. Cell. Biol.* **11**,5312–5320
41. Ryner,L.C., Takagaki,Y. and Manley,J.L. Sequences downstream of AAUAAA signals affect pre-mRNA cleavage and polyadenylation in vitro both directly and indirectly.
42. Sadofsky,M., Connelly,S., Manley,J.L. and Alwine,J.C. (1985) *Mol. Cell. Biol.* **5**,2713–2719.
43. Woychik,R.P., Lyons,R.H., Post,L.and Rottmann,F.M. (1984) *Proc. Natl. Acad. Sci. USA.* **81**,3944–3948.
44. Birnsteil,M.L.,Buslinger,M. and Strub,K. (1985) *Cell.* **41**, 349–359.
45. Manley,J.L. (1988) *Biochem. Biophys. Acta.* **950**,1–12.
46. Wickens,M. (1990) *Trends Biol Sci.* **15**,277–281.
47. Sheets,M.D., Ogg,S.C. and Wickens,M.P. (1990) *Nuc Acids Res.* **18**,5799–5805.
48. Denome,R.M.and Cole,C.N. (1988) *Mol. Cell Biol.* **8**,4829–4839.
49. Hamlin,J. L., Milbrandt, J.D., Heintz, N.H., and Azizkhan, J.C. (1989) *International Rev. Cytol.* **90**, 31–82.