Volume 14 Number 10 1986 Nucleic Acids Research

Structural features of the murine dihydrofolate reductase transcription termination region: identification of a conserved DNA sequence element

Elizabeth G.Frayne and Rodney E.Kellems

The Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, Houston, TX 77030, USA

Received 15 January 1986; Revised and Accepted 23 April 1986

ABSTRACT

Structural features of the transcription termination region for the mouse dihydrofolate reductase gene have been determined and compared with those of several other known termination regions for protein coding genes. A common feature identified among these termination regions was the presence of a 20 bp consensus DNA sequence element (ATC$^{A}_{G}$AA$^{T}_{A}$TAG$^{G}_{A}$AA$^{G}_{T}$AGCAAT). The results imply that the 20 bp consensus DNA sequence element is important for signaling RNA polymerase II transcription termination at least in the several vertebrate species investigated. Furthermore, the results suggest that for the dhfr gene and possibly for other genes in mice as well, the potential termination consensus sequence can exist as part of a long interspersed repetitive DNA element.

INTRODUCTION

Studies in prokaryotes indicate that transcription termination sites can be located within operons as well as at the end of operons. Control of termination at internal sites (i.e., attenuation) is an important mechanism for regulating prokaryotic gene expression (1,2). A common feature of many prokaryotic terminators is the presence of a GC-rich dyad symmetry followed by a stretch of T residues in the noncoding strand, close to where the transcript ends. Much less is known about transcription termination in eukaryotes. Eukaryotic cells contain three classes of RNA polymerase that perform separate transcription functions. RNA polymerase III which transcribes both 5S and tRNA genes requires a short stretch of T residues, often flanked by GC-rich sequences for accurate termination (3,4,5). RNA polymerase I which transcribes ribosomal RNA genes appears to terminate at a discrete site located 565 bp beyond the 3' end of the mature 28S rRNA and involves interaction of factors with repeated sequences in the 3' spacer (6). Transcription termination for RNA polymerase II, however, is poorly understood largely because of the rapid RNA processing that occurs at the 3' ends of transcripts for protein

coding genes. Processing generally involves cleavage of a precursor followed by poly(A) addition (7,8). Even the 3' end of histone mRNA which lacks a poly(A) tail is generated by processing (9,10). In the lower eukaryote yeast, however, it has been proposed that transcription termination and polyadenylation might be coupled (11). Although the sequence requirements for termination and/or polyadenylation in yeast have been investigated, no consistent structural features have emerged (11,12,13,14).

In higher eukaryotes RNA polymerase II transcription termination appears to occur several hundred to several thousand nucleotides downstream of the site(s) of polyadenylation (7). For several genes a region within which transcription presumably terminates has been mapped by hybridizing nascent labeled RNA to an ordered set of cloned genomic DNA fragments (15,16,17,18,19). From such hybridization studies, transcription termination has been located for these genes to within regions ranging in size from 170-1210 bp. Additional fine mapping studies conducted in two instances, however, failed to detect discrete sites of termination (15,17). Functional studies of the mouse β^{maj}-globin termination region indicate that a 1.5 kb fragment spanning the termination region blocks transcription 80-90% when inserted in the appropriate orientation into a foreign gene (20). While these studies suggest that a genetic signal directs transcription termination, the consensus sequence elements likely to be involved in this signaling have not yet been identified.

The transcription unit boundaries for the mouse dhfr gene have recently been determined in cells with amplified copies of the dhfr gene. The mouse dhfr transcription unit is approximately 36 kb in length. Transcription initiates at multiple sites within a promoter region that is shared with a divergently transcribed gene (21,22). Transcription termination occurs within a 900 bp region located downstream of at least seven poly(A) sites and near a large region of repetitive DNA (16). To identify structural features that may be important for signaling termination, we have further characterized the mouse dhfr termination region and have compared the DNA sequence of this region with that of other known termination regions. Our present findings indicate that RNA polymerase II termination regions of genes from such diverse species as chickens, rabbits, and mice share a striking DNA sequence homology. Based on these observations, we postulate a 20 bp consensus DNA sequence element impor-

tant in signaling termination for vertebrate gene transcription units. Our results further imply that for the dhfr gene, and possibly other genes in mice as well, the potential termination consensus sequence is present as part of a long interspersed repetitive DNA element.

MATERIALS AND METHODS

M13 cloning and DNA sequencing. mDHA5.6 and mDAH3.5 were derived from mDH9 (16), a previously characterized M13 recombinant containing a 900 bp HindIII fragment that corresponds to the dhfr termination region. Digestion of the 900 bp HindIII fragment with AccI yielded two fragments, both of which were cloned into M13 derivatives to generate mDHA5.6 and mDAH3.5, respectively. Single-stranded phage DNAs from mDH9, mDHA5.6, and mDAH3.5 were prepared and sequenced by the dideoxynucleotide chain termination method described by Sanger et al. (23). Sequencing reactions were performed with [^{35}S]dATP as the radioactive nucleotide and were run on 8% and 6.5% polyacrylamide sequencing gels. Two synthetic oligonucleotides (5' ATACTTCTGAATTGAAACTAAACAC 3', 5' AAGAATTTAAAGGTGTTTAGTTTCA 3') were synthesized by standard phosphoramidite chemistry (24) on an Applied Biosystems DNA Synthesizer (model 380A) and used to prime internally in both directions the region corresponding to mDHA5.6 as shown in Figure 1.

Dot blot hybridizations. In vitro transcription in isolated nuclei and subsequent RNA isolation were performed as described previously (16). Nuclei were incubated in 200 µl reactions with 400 µCi of uridine [α-^{32}P]triphosphate (3700 Ci/mmol, ICN) and cold UTP added to a final concentration of 2 µM. The labeled RNA was then hybridized to nitrocellulose filters that had been spotted with 4 µg of single-stranded M13 recombinant phage DNA. Nitrocellulose filters containing bound phage DNA were prepared using a Schleicher and Schuell filter manifold according to the manufacturer's instructions. Hybridization conditions were as described previously (16). Blots were exposed to Kodak XAR-5 film with an intensifying screen for 1 to 2 days.

Sequence analysis. The sequence of the dhfr termination region was analyzed on an IBM PC-XT computer using Beckman Instrument's Microgenie Sequence Analysis Program developed by Queen and Korn (25). The Make Search function was used to screen Genbank (Nov., 1985) for sequences exhibiting extensive homology with the dhfr termination region. The Compare function was used to identify the short stretches of homology shared between the different termination regions examined. The Compare

function was also used to screen all the eukaryotic sequences in Genbank for 20 bp sequences with 85% or more homology to the 20 bp consensus sequence ATC$\begin{matrix}A\\G\end{matrix}AA\begin{matrix}T\\A\end{matrix}TAG\begin{matrix}G\\A\end{matrix}AA\begin{matrix}G\\T\end{matrix}$AGCAAT or to the complement of this sequence (search parameters were: MINMATCH=17, LOOPLEN=0, and MINPER=85).

The random probability of finding within each particular termination region a 20 bp sequence with as good or better homology to the consensus sequence as the one observed was estimated using the following equation which employs basic permutation methods:

$$\sum_{m=n-q}^{n} 1 - \left\{ \left[1 - \left(\frac{n!}{m!(n-m)!} (1/4)^m (3/4)^{n-m} \right) \right]^{[N-n+1]} \right\}$$

Relative to a particular n base pair long sequence (the 20 bp consensus sequence), this equation estimates the probability of finding homologous sequences with up to q mismatched bases or m matching bases from m=n-q to m=n, within a string N bases long (length of the termination fragment). For example, for the chicken ovalbumin termination region N=170; therefore, the random probability of finding, as observed, a 20 bp homologous sequence with at most 3 mismatched bases (or q=3) relative to the 20 bp consensus sequence within the ovalbumin termination region would be 8 x 10^{-5}. This equation was modified to estimate the probability of finding a sequence at least as homologous as the sequence CTCAAAATAGGAATAG-AAT within the rabbit ß1-globin termination region. In particular, we estimate the probability of finding a sequence with a ratio of matched bases relative to the 20 bp consensus sequence of .85 or better, allowing for at most a one base pair gap, within the rabbit ß1-globin termination region would be 2.5 x 10^{-4}.

RESULTS

The dhfr termination region overlaps with a truncated mouse L1 repeat family member. Transcription termination for the dhfr gene was previously located to within a 900 bp HindIII fragment that either abuts or overlaps a large region of repetitive DNA sequences (16). To examine the location of transcription termination with respect to repetitive DNA sequences, experiments were performed to map both dhfr transcription and repetitive DNA sequences within the 900 bp HindIII fragment. To investigate the level of transcriptional activity within the HindIII fragment, two M13

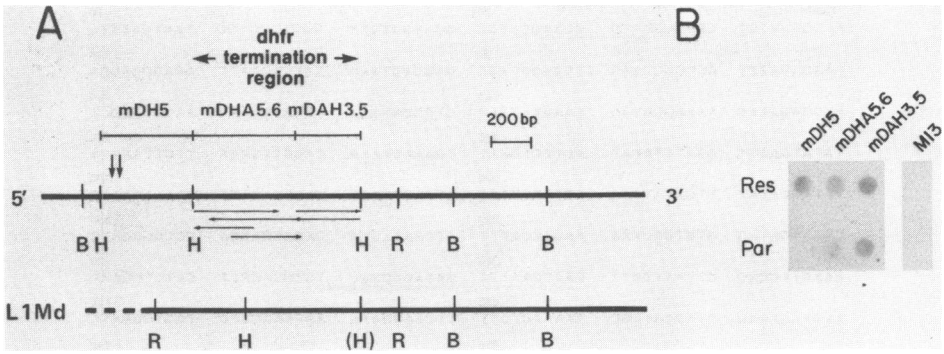

FIG. 1. Location of dhfr transcription termination with respect to repetitive DNA sequences. (A) Map of the 3' end and flanking region of the dhfr gene. The restriction enzyme map surrounding the dhfr termination region (33) is compared with the relevant portion of a consensus map for the mouse L1Md repetitive DNA family (34,35). Parentheses indicate a site in L1Md repeat sequences that occurs less frequently. The restriction enzyme sites are designated as follows: B=BamHI; H=HindIII; R=EcoRI. The vertical arrows above the map indicate the locations of two additional poly(A) sites beyond poly(A) site seven that correspond to two minor mRNAs recently mapped in this region by Hook and Kellems (manuscript in preparation). The horizontal arrows indicate the direction and extent of strands sequenced. DNA fragments cloned into M13 and used for sequencing and/or hybridization experiments are shown above the map. The previously characterized mDH5 recombinant (16) contains a 500 bp insert. mDHA5.6 and mDAH3.5 were prepared as described in Materials and Methods and contain, respectively, hybridizable inserts of 561 bp and 352 bp in length. (B) Measurement of transcriptional activity within the 907 bp termination region. Nascent labeled RNA prepared from isolated nuclei incubated in vitro with [^{32}P]UTP was hybridized to nitrocellulose bound phage DNAs. Equivalent cpm (10^6) of labeled RNA from resistant S180-500R (Res) nuclei or from parental S180 (Par) nuclei were hybridized to the dhfr strand-specific probes indicated as well as to M13 mp10. Transcriptional activity associated with the amplified DNA in lanes 1-3 can be approximated by observing the difference in the intensity of hybridization signal when resistant RNA is used versus when parental RNA is used.

recombinants were constructed containing genomic DNA inserts corresponding to 5' (mDHA5.6) and 3' (mDAH3.5) portions of this fragment (see Fig. 1A). Single-stranded phage DNA from the M13 recombinants was spotted on nitrocellulose paper and used in dot blot hybridization to nascent labeled nuclear RNA obtained by labeling isolated nuclei. Similar to our previous studies, each probe was hybridized to RNA isolated from the nuclei of dhfr gene amplification cells (S180-500R) as well as from the parental or unamplified cells (S180). Although the dhfr gene is expressed in the parental cells, it is transcribed at a level too low to detect under the

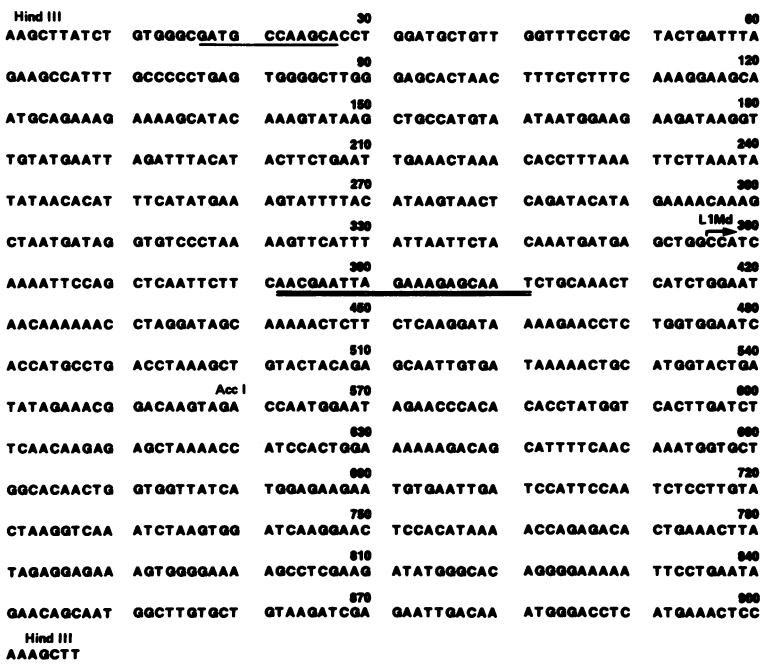

FIG. 2. Sequence of the dhfr termination region. Numbering starts at the 5' end of the 907 bp HindIII fragment located approximately 400 bp from the upstream poly(A) site shown in Figure 1. Sequences homologous to those observed in other termination regions are underlined with one or two lines. Sequences homologous to L1Md repeat sequences are detected beginning at position 356 as indicated by the horizontal arrow and essentially extend through to position 907 (see Table 1).

assay conditions used. Therefore, hybridization of parental RNA to these genomic probes was useful in distinguishing between transcriptional activity associated with the amplified DNA and transcriptional activity from homologous repeated DNA sequences located elsewhere in the genome.

Figure 1B shows the results from such a dot blot hybridization experiment. The results from this experiment and others can be interpreted as indicating that transcription of the dhfr gene continued into the mDHA5.6 region and terminated predominantly within this region. Hybridization of nascent labeled amplified RNA to mDHA5.6 was consistently less than that to a slightly smaller upstream probe (mDH5), indicating that transcriptional activity decreased within the mDHA5.6 region. Furthermore, not all of the apparent transcriptional activity associated with mDHA5.6 was specific for the amplified DNA because this probe also

Table 1. Extent of homology between the dhfr termination region and mouse L1 repeat sequences.

Mouse L1 clone[a]	907 bp HindIII termination fragment (position)	Homology (%)
1. L1Md (C37)	356-506	95.4
2. L1Md (PMD006)	567-633	85.1
3. L1Md (4)	693-907	95.8

[a]Name refers to specific mouse L1 clone (observed in Genbank) exhibiting the highest level of homology to the dhfr nucleotides indicated. NT 507-566 and 634-692 were less than 80% homologous to other known L1 repeat clones. The L1 clones are designated in accordance with recently proposed simplifying nomenclature (26,34) and have been described elsewhere (35,36,39).

hybridized to a certain extent with nascent labeled parental RNA. The adjacent 3' probe, mDAH3.5, hybridized significant and roughly equivalent amounts of nascent labeled amplified RNA and parental RNA. Because little if any of the apparent transcriptional activity associated with mDAH3.5 was specific for the amplified DNA, the majority of transcription presumably terminated within the mDHA5.6 region. The results further suggested that the repetitive DNA sequences flanking the dhfr gene begin within the mDHA5.6 region and encompass the mDAH3.5 region. This was evident because both mDHA5.6 and mDAH3.5 hybridized to nascent labeled parental RNA while the upstream probe mDH5, as expected from previous studies, did not.

The location and type of repetitive DNA in the 900 bp HindIII fragment was determined directly by DNA sequence analysis. The HindIII fragment was determined to be 907 bp in length as shown in Figure 2. Sequences homologous to the long interspersed mouse L1 repeat family (also referred to as the BamHI or MIF-I repeat family) were detected between nucleotides (NT) 356 and 907 (see Table 1). In addition, the restriction enzyme map 3' but not 5' of the sequenced region was similar to that of a consensus L1Md (Mus domesticus) family member (see Fig. 1). Taken together the results indicated that the repetitive DNA flanking the dhfr structural gene most likely represented a truncated L1Md family member. Members of this family are often truncated at their 5' ends (26). For the repeat flanking the dhfr structural gene, the breakpoint occurred within the HindIII fragment probably at or near nucleotide 356. Thus 36% of the mDHA5.6 region and all of the mDAH3.5 region overlapped with the truncated L1Md repeat family member flanking tne dhfr gene.

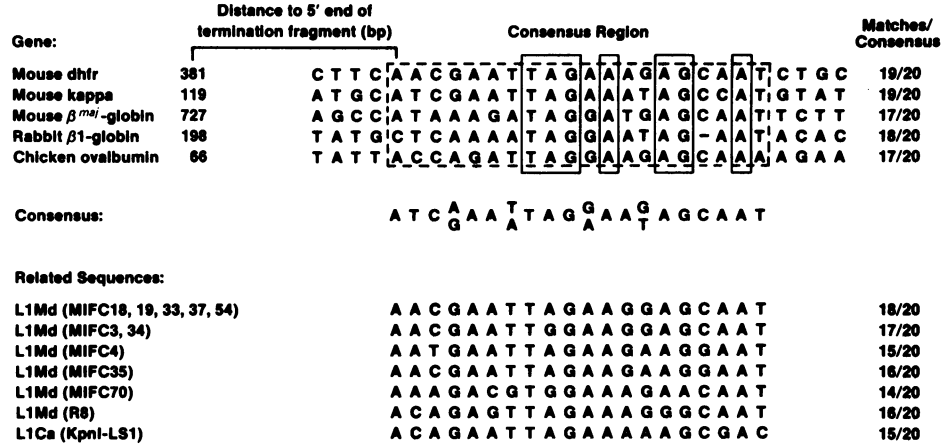

FIG. 3. Alignment of homologous sequences detected in RNA polymerase II termination regions. For the genes shown, transcription termination has been localized to within a sequenced restriction enzyme DNA fragment, in particular, a 907 bp HindIII fragment for the dhfr gene (16, this paper); a 1210 bp BamHI/BglII fragment for the kappa gene (27, X. Ming et al., manuscript submitted); an 806 bp XbaI/BglII fragment for the ß^{maj}-globin gene (15,27); a 568 bp EcoRI/BglII fragment for the ß1-globin gene (19); and a 170 bp HaeIII fragment for the ovalbumin gene (18). The sequences of these termination fragments were analyzed for homology (see Materials and Methods). Shown is a common region of homology identified, boxed with a dashed line. A single one base pair gap was introduced into one sequence to maximize the extent of homology. Nucleotides at seven different positions within the homologous region were invariant and are boxed with solid lines. The derived 20 bp consensus sequence is shown beneath the dashed box. The number of matched bases for each sequence relative to the 20 bp consensus sequence is shown to the right of the dashed box. Beneath the 20 bp consensus sequence are shown the related sequences detected in different L1Md repeat family members (36,37). Six variants were detected among eleven different L1Md family members; all of these occurred at the same relative location within each L1Md repeat member. Also shown is a sequence detected within the corresponding region of the related primate L1 family, specifically within an African green monkey (Cercopithecus aethiops) repeat sequence (38). The L1 clones, as those in Table 1, are designated in accordance with recently proposed nomenclature (26,34).

Other features of the dhfr termination region. The overall base composition of the sequence for the 907 bp HindIII fragment was 37% A, 19% C, 20% G, and 24% T. An especially AT-rich region, 224 bp in length, occurred at NT 127-350. This 224 bp region had an AT content of 73% and was flanked on either side by regions of relatively normal AT content for mammalian DNA (50% 5', NT 1-126; and 59% 3', NT 351-907). Prior to the AT-rich region, a GC-rich inverted repeat occurred (between NT 71-86) capable of forming a particularly stable hairpin (-16 kcal).

The dhfr termination region shares homology with other termination regions. DNA sequence comparisons were made between the dhfr termination region and all other presently known termination regions to identify common structural features that may be important for signaling termination. Specifically, the termination regions or DNA fragments within which transcription terminates for the following genes were examined: the chicken ovalbumin gene, a 170 bp HaeIII fragment (18); the mouse β^{maj}-globin gene, an 806 bp XbaI/BglI fragment (15,20); the rabbit ß1-globin gene, a 568 bp EcoRI/BglII fragment (19); and the mouse kappa light chain gene, a 1210 bp BamHI/BglII fragment (27, X. Ming et al., manuscript submitted). Sequence comparisons were made using a computer program designed to detect short stretches of strong homology; the parameters for the program were set to identify potentially significant homologies (see Materials and Methods).

The DNA sequence analysis identified an extensive stretch of homology shared among all of the termination regions examined. When the corresponding sequences were aligned, it was possible to derive a 20 bp consensus sequence element ($ATC^{A}_{G}AA^{T}_{A}TAG^{G}_{A}AA^{G}_{T}AGCAAT$) shown in Figure 3. This consensus sequence element contained two nearly identical pentamers (AGNAA and AGNAA). Within these two pentamers occurred six of the seven invariant nucleotide positions detected among the five 20 bp sequence elements. Each of the five 20 bp sequence elements showed a ratio of matched bases relative to the consensus of .85-.95. The significance of the 20 bp homologous sequence element is substantiated by statistical considerations. In particular, the random probability of finding a sequence with as good or better homology to the 20 bp consensus sequence as the one observed within each of the different termination regions examined ranged from 2.5×10^{-4} for the rabbit ß1 globin gene to 9×10^{-7} for the mouse dhfr gene (see Materials and Methods for calculations).

An additional stretch of homology was identified either within or just upstream (in two cases) of each termination region. When the additional homologous sequences were aligned, it was possible to derive an 11 bp consensus sequence element (GAAGCCAAGCA) shown in Figure 4. The 11 bp sequence element within or near each of the termination regions showed a ratio of matched bases relative to the consensus sequence of .72-1.0. The potential significance of the 11 bp homologous sequence element is enhanced by virtue of its occurring in each case upstream of

Gene:	Consensus Region	Distance to 20 bp sequence element (bp)	Matches/ Consensus
Mouse dhfr	C G A T G C C A A G C A C	354	10/11
Mouse kappa	A T A A G C A A A A A C A A	404	8/11
Mouse β^{maj}-globin	A G A A G C C A A G C A G	150	11/11
Rabbit β1-globin	T G A A T C C A T G A A A	12	8/11
Chicken ovalbumin	A G A A G C C A A G A G C	120	9/11
Consensus:	G A A G C C A A G C A		

FIG. 4. Alignment of upstream homologous sequences. A common region of homology is boxed with a dashed line and was used to derive the 11 bp consensus sequence shown. The distance in bp downstream to the 20 bp sequence element (see Fig. 3) is shown to the right of the boxed region of homology. The number of matched bases relative to the 11 bp consensus sequence is shown to the far right.

the 20 bp sequence element. However, for two genes the 11 bp sequence element was not highly conserved and therefore of questionable significance. It is possible that this sequence element may be an important component of some but not all of the termination regions.

DISCUSSION

We have observed a striking 20 bp consensus DNA sequence element within the RNA polymerase II termination regions of several vertebrate genes, lending support to the idea that these regions are functional domains that serve to define the 3' boundaries of transcription units. A computer search of all eukaryotic sequences in Genbank (see Materials and Methods) for sequence elements showing at least 85 percent homology to the 20 bp consensus sequence revealed only two copies in addition to those listed in Figure 3. One copy is located within the intercistronic region between the divergent tobacco chloroplast genes encoding the large subunit of ribulose bisphosphate carboxylase and the ß-subunit of ATP synthetase (28) and is presumably not involved with transcription termination in this plant organelle. A second copy is located in the 3' flanking region of the Vaccinia Virus early gene and may serve as part of a transcription termination signal (29). It was not surprising that more copies of the 20 bp consensus sequence were not detected at the 3' ends of other cellular genes in the data bank because the sequence information available for most genes does not extend more than a few hundred bp beyond the site(s) of polyadenylation and thus is unlikely to include a termination region. However, as several viral genomes that occur in higher eukaryotes

have been completely sequenced, the absence of such sequences near the 3' ends of viral genes (with the exception noted above) suggests that viruses may utilize alternate genetic signals for specifying termination or may not possess such signals. The latter possibility is supported by studies of viral gene expression in cells infected with either polyoma virus or adenovirus. Such studies indicate for the polyoma virus a significant fraction of late viral transcripts extend more than once around the circular genome (30). Similarly, no discernible termination event has been detected for the adenovirus major late transcription unit as the 3' boundary of this transcription unit occurs at or very near the end of the linear genome (31). In any event, the remarkable conservation of the 20 bp sequence element among the termination regions of genes from such evolutionarily diverse species, suggests that this sequence may function alone or as part of a general termination signal for higher eukaryotic cellular genes transcribed by RNA polymerase II.

The complexity of the 20 bp consensus sequence element seems appropriate for its putative function in that such a sequence would not be expected to occur "inappropriately" on a random basis within even the longest polymerase II transcription units. However, other accessory signals, possibly the 11 bp consensus sequence element (see Fig. 4) or a polyadenylation site, may also be required to elicit termination. An intriguing finding was that for the dhfr termination region the 20 bp sequence element was located within the first 50 bp of a truncated L1Md family member. In contrast, the 20 bp sequence elements detected within the other termination regions are not part of known repetitive DNA elements. In this regard, however, it is interesting that for the mouse immunoglobulin μ heavy chain gene, the endpoint for transcription (prior to class switching) is located either within or near repetitive DNA sequences (32). In addition, we have observed that the corresponding 20 bp sequences of other cloned mouse L1 repeats were 60-80% homologous to the consensus sequence, several were just slightly less homologous to the consensus sequence than the dhfr sequence (shown in Fig. 3). Therefore, it is conceivable that some members of the mouse L1 repeat family, present in approximately 20,000 copies in the mouse genome, could play a role in signaling termination for genes in mice other than dhfr.

Further supportive evidence that the homologous sequences we have identified may be important for specifying transcription termination comes from the recent studies of the mouse β^{maj}-globin termination

region (26). In these studies it was demonstrated that a 1.5 kb fragment spanning the globin poly(A) site and distally located transcription termination region or fragment F (807 bp) effectively blocked transcription when placed within the adenovirus E1A gene. Both the 20 and 11 bp homologous sequence elements shown in Figures 3 and 4 are positioned towards the 3' end of fragment F and are included in the functional terminator fragment. In contrast, E1A transcription was not blocked in similar experiments using an internal segment of the 1.5 kb fragment that lacked the poly(A) site as well as the 20 bp and 11 bp sequence elements. These results were originally interpreted as suggesting that an active polyadenylation signal may be required for termination to occur. Based on our findings, an additional possibility would be that the 20 bp sequence element and possibly the 11 bp sequence element as well may be required for function of the β^{maj}-globin terminator. Functional tests of the β^{maj}-globin terminator and those of other transcription units should be used to examine separately the importance of the poly(A) site and the DNA sequence elements identified here.

ACKNOWLEDGMENTS

We thank Dr. William T. Garrard for information prior to publication concerning the mouse kappa termination region. We also thank Drs. Susan M. Berget and John H. Wilson for helpful discussions during the course of this work and, along with Dr. Jeffrey M. Rosen, for comments on the manuscript. We are especially grateful to John H. Wilson and Hiram F. Gilbert for deriving the equation presented under Materials and Methods and for providing advice on statistical analysis of nucleotide sequence information.

This research was supported by Public Health Service grant GM-30204 from the National Institutes of Health. R.E.K. was supported by a Research Career Development Award (1 K04 CA00828). Methotrexate used in maintaining S180-500R cells was supplied by the Drug Synthesis and Chemistry Branch, Division of Cancer Treatment, National Cancer Institute.

REFERENCES

1. Holmes, W.M., Platt, T. and Rosenberg, M. (1983) Cell 32, 1029-1032.
2. von Hippel, P.H., Bear, D.G., Morgan, W.D., and McSwiggen, J.A. (1984) Ann. Rev. Biochem. 53, 389-446.
3. Adeniyi-Jones, S., Romeo, P.H., and Zasloff, M. (1984) Nucleic Acids Res. 12, 1101-1115.
4. Bogenhagen, D.F. and Brown, D.D. (1981) Cell 24, 261-270.
5. Cozzarelli, N.R., Gerrard, S.P., Schlissel, M., Brown, D.D., and Bogenhagen, D.F. (1983) Cell 34, 829-835.
6. Bakken, A., Morgan, G., Sollner-Webb, B., Roan, J., Busby, S., and Reeder, R.H. (1982) Proc. Natl. Acad. Sci. USA 79, 56-60.
7. Birnstiel, M.L., Busslinger, M., and Strub, K. (1985) Cell 41, 349-359.
8. Moore, C.L. and Sharp, P.A. (1985) Cell 41, 845-855.

9. Birchmeier, C., Schumperli, D., Sconzo, G., and Birnstiel, M.L. (1984) Proc. Natl. Acad. Sci. USA 81, 1057-1061.
10. Krieg, P.A. and Melton, D.A. (1984) Nature (London) 308, 203-206.
11. Zaret, K.S. and Sherman, F. (1982) Cell 28, 563-573.
12. Henikoff, S., Kelly, J.D., and Cohen, E.H. (1983) Cell 33, 607-614.
13. Henikoff, S. and Cohen, E.H. (1984) Mol. Cell. Biol. 4, 1515-1520.
14. Sutton, A. and Broach, J.R. (1985) Mol. Cell. Biol. 5, 2770-2780.
15. Citron, B., Falck-Pedersen, E., Salditt-Georgieff, M., and Darnell Jr., J.E. (1984) Nucleic Acids Res. 12, 8723-8731.
16. Frayne, E.G., Leys, E.J., Crouse, G.F., Hook, A.G., and Kellems, R.E. (1984) Mol. Cell. Biol. 4, 2921-2924.
17. Hagenbüchle, O., Wellauer, P.K., Cribbs, D.L., and Schibler, U. (1984) Cell 38, 737-744.
18. LeMeur, M.A., Galliot, B., and Gerlinger, P. (1984) EMBO J. 3, 2779-2786.
19. Rohrbaugh, M.L., Johnson III, J.E., James, M.D., and Hardison, R.C. (1985) Mol. Cell. Biol. 5, 147-160.
20. Falck-Pedersen, E., Logan, J., Shenk, T., and Darnell Jr., J.E. (1985) Cell 40, 897-905.
21. Crouse, G.F., Leys, E.J., McEwan, R.N., Frayne, E.G., and Kellems, R.E. (1985) Mol. Cell. Biol. 5, 1847-1858.
22. Farnham, P.J., Abrams, J.M., and Schimke, R.T. (1985) Proc. Natl. Acad. Sci. USA 82, 3978-3982.
23. Sanger, F., Nicklen, S., and Coulson, A. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
24. Beaucage, S.L. and Caruthers, M.H. (1981) Tet. Lett. 22, 1859-1862.
25. Queen, C. and Korn, C.J. (1984) Nucleic Acids Res. 12, 581-599.
26. Voliva, C.F., Jahn, C.L., Comer, M.B., Hutchison III, C.A., and Edgell, M.H. (1983) Nucleic Acids Res. 11, 3631-3636.
27. Neumaier, P.S. and Zachau, H.G. (1983) Nucleic Acids Res. 11, 3631-3636.
28. Shinozaki, K. and Sugiura, M. (1982) Nucleic Acids Res. 10, 4923-4934.
29. Venkatesan, S., Baroudy, B.M., and Moss, B. (1981) Cell 125, 805-813.
30. Acheson, N.H. (1984) Mol. Cell. Biol. 4, 722-729.
31. Nevins, J.R. and Darnell Jr., J.E. (1978) Cell 15, 1477-1493.
32. Mather, E.L., Nelson, K.J., Haimovich, J., and Perry, R.P. (1984) Cell 36, 329-338.
33. Crouse, G.F., Simonsen, C.C., McEwan, R.N., and Schimke, R.T. (1982) J. Biol. Chem. 257, 7887-7897.
34. Singer, M.F. and Skowrousk, J. (1985) Trends In Biochem. Sci. 10, 119-122.
35. Martin, S.L., Voliva, C.F., Burton, F.H., Edgell, M.H., and Hutchison III, C.A. (1984) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
36. Brown, S.D.M. (1983) J. Mol. Biol. 165, 249-256.
37. Gebhard, W. and Zachau, H.G. (1983) J. Mol. Biol. 170, 255-270.
38. Singer, M.F., Thayer, R.E., Grimaldi, G., Lerman, M.I., and Fanning, T.G. (1983) Nucleic Acids Res. 11, 5739-5745.
39. Dubnick, M., Chou, J., Petes, T.D., and Farber, R.A. (1983) J. Mol. Evol. 19, 115-121.