Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 2531–2535, April 1990
Genetics

Polyadenylylation signal of the mouse thymidylate synthase gene was created by insertion of an L1 repetitive element downstream of the open reading frame

[poly(A)/RNA processing/sequence evolution/long interspersed repetitive element/repetitive DNA]

CHRISTOPHER J. HARENDZA*† AND LEE F. JOHNSON‡§

Departments of Biochemistry and Molecular Genetics‡ and The Molecular, Cellular and Developmental Biology Program*‡, Ohio State University, Columbus, OH 43210

Communicated by Maxine F. Singer, January 11, 1990 (received for review October 18, 1989)

ABSTRACT **The mouse thymidylate synthase (TS; EC 2.1.1.45) mRNA is unusual in that the poly(A) tail is added at the translation stop codon. To determine the sequence requirements for 3′ processing of this mRNA, we constructed TS minigenes with deletion and point mutations in potential regulatory sequences. The minigenes were transiently transfected into cultured cells and the effect on 3′ processing was determined by S1 nuclease protection assays. These analyses revealed that at least two elements are required for efficient polyadenylylation at the stop codon. The first is an upstream AUUAAA sequence. When this was changed to AUCAAA, polyadenylylation at the stop codon was blocked. However, when it was changed to the canonical AAUAAA hexanucleotide, the amount of TS mRNA increased severalfold. The second element is a stretch of 14 consecutive uridylate residues 32 nucleotides downstream of the stop codon. This U-rich region is absent from the human TS gene, which explains why the human TS mRNA is not polyadenylylated at the stop codon even though the two genes are otherwise almost identical through this region. The most surprising observation was that the U-rich region corresponds to the 3′ end of a 360-nucleotide mouse L1 repetitive element that was inserted in opposite orientation to the gene more than 5 million years ago. Thus the polyadenylylation signal of the present mouse TS gene was created by the transposition of a repetitive element downstream of a cryptic polyadenylylation signal.**

The 3′ ends of most eukaryotic mRNAs are formed by a processing reaction involving site-specific endonucleolytic cleavage followed by the addition of 200–300 adenylate residues by poly(A) polymerase. The sequences that direct this process generally consist of an AAUAAA hexanucleotide (or slight variations) located 10–30 nucleotides upstream of the processing site and loosely conserved G+U- and/or U-rich sequences that are located downstream of the processing site (1, 2). The various elements must be properly positioned relative to each other for optimal polyadenylylation (2, 3). Although the functions of the upstream and downstream elements have not been completely defined, they are required for the recognition of the polyadenylylation site by the trans-acting factors that participate in the 3′ processing reactions (2, 4–8).

Our laboratory has been studying the structure and expression of the gene for mouse thymidylate synthase (TS; 5,10-methylenetetrahydrofolate:dUMP *C*-methyltransferase, EC 2.1.1.45). The predominant (80%) TS mRNA is highly unusual in that it lacks a 3′ untranslated region; the stop codon, UAA, is immediately followed by the poly(A) tail (9, 10). The stop codon in the mouse TS gene is UAG (11), indicating that the final A of the UAA stop codon is added by poly(A) polymerase. Sequences that conform to polyadenylylation consensus elements are present in the vicinity of the stop codon of the TS gene. These include a variant hexanucleotide, AUUAAA, in the coding region, a G+U-rich sequence immediately downstream of the stop codon, and 14 consecutive uridylate residues beginning 32 nucleotides downstream of the stop codon (10). The human and mouse TS genes are very similar across the coding region and differ at only one position between the AUUAAA and G+U-rich sequences (see Fig. 1*B*) (10, 12). In spite of this, the human TS mRNA is polyadenylylated not at the stop codon but rather 500 nucleotides downstream (12). It was suggested that the human TS mRNA was not polyadenylylated at the stop codon because it lacked the oligo(U) region (10).

In this paper, we identify the upstream and downstream sequences that are important for the polyadenylylation of mouse TS mRNA. We show that variations in these sequences lead to large changes in the efficiency of utilization of the major and minor polyadenylylation sites. The oligo(U) region was found to be essential for directing efficient polyadenylylation of TS mRNA at the stop codon. The oligo(U) region and sequences distal to it correspond to the 3′ end of a mouse L1 repetitive element that was inserted in opposite orientation relative to the direction of transcription of the TS gene.

MATERIALS AND METHODS

Deletion and Site-Directed Mutagenesis. The TS expression vectors used are shown in Fig. 1*A*. Deletion mutations were created in an intronless mouse TS minigene (pTSMG3) that consists of TS cDNA linked to 0.3 kb of 5′ and 3′ flanking DNA from the mouse TS gene (13). Point mutations were created in TS minigene pI56, which is similar to pTSMG3 except that it contains introns 5 and 6 of the TS gene at their normal positions and has 1 kb of 5′ flanking DNA from the TS gene (14).

The upstream ATCAAA (hex-1) mutation was created by the protocol of Eckstein and coworkers (15), using a kit supplied by Amersham. The pTSMG3 insert was cloned into plasmid pBS(+) (Stratagene). Single-stranded DNA was rescued by using M13KO7 helper phage and was mutated by using the synthetic oligodeoxynucleotide GCCATTTC-CATTTTGATCGTTGG. Plasmids bearing the desired mutation were identified by the presence of a new *Sau*3A1 restriction site. The 336-bp *Cla* I–*Sac* I fragment was cloned back into pI56 to create the mutant TS minigene. The

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: TS, thymidylate synthase.

†Present address: Laboratory of Reproductive Physiology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA 19104.
§To whom reprint requests should be addressed.

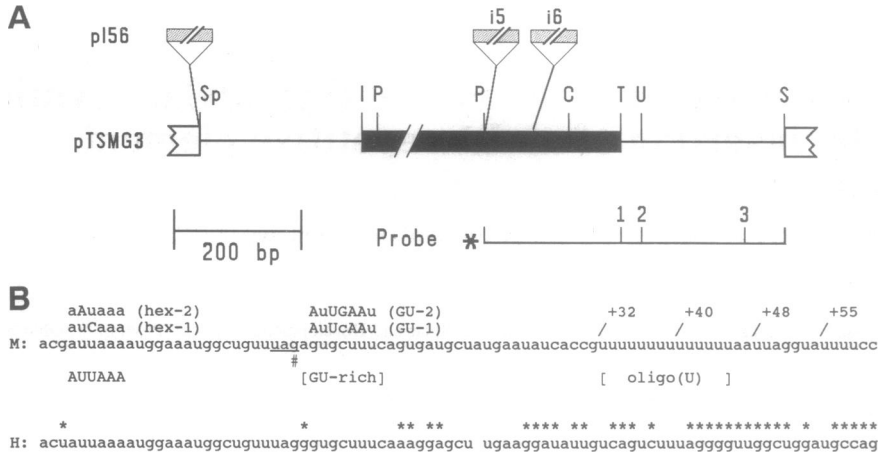

FIG. 1. (*A*) Structures of the TS minigenes and S1 nuclease probes. The restriction map of the intronless minigene, pTSMG3, is indicated. The coding region (black bar) is 1 kilobase (kb) long. The minigene is cloned into pUC18 (open bar). The structure of pI56 is the same except for the presence of introns 5 (i5, 1 kb) and 6 (i6, 0.8 kb) at their normal locations in the coding region and about 0.7 kb of additional 5′ flanking sequence upstream of the *Sph* I site. The level of expression of pI56 is about 8 times greater than that of pTSMG3. The patterns of transcriptional initiation and polyadenylylation for both minigenes are the same as those observed with normal TS mRNA isolated from mouse cells (13, 14). Abbreviations: Sp, *Sph* I; P, *Pst* I; C, *Cla* I; S, *Sac* I; I, AUG initiation codon; T, UAG termination codon; U, U-rich region; bp, base pairs. The S1 probe is shown below the minigenes. The probe was end-labeled (*) at the *Pst* I site in the coding region and extended to the *Sac* I site in the 3′ flanking region. The locations of the 3′ ends of fragments 1, 2, and 3 are indicated on the probe. (*B*) The 3′ ends of the mouse and human TS genes. The sequence of the mouse TS gene (M; shown as RNA) begins 3 bases upstream of the AUUAAA hexanucleotide and extends 15 bases beyond the oligo(U) region. The UAG stop codon is underlined. The changes introduced by the point mutations and the deletion endpoints are shown above the sequence. The upstream hexanucleotide, G+U-rich region, oligo(U) region and the poly(A) addition site (#) for the mouse TS polyadenylylation signal are indicated on the middle line. The sequence of the human (H) TS gene in the same region is shown on the bottom line. Nucleotide differences are indicated by stars. The G of the stop codon is designated +1.

fragment was sequenced (16) before and after insertion into pI56 to confirm the presence of the desired mutation and the absence of other mutations.

All subsequent mutations were created by the procedure of Kunkel (17). To construct the upstream AATAAA (hex-2) mutation, uracil-substituted single-stranded DNA that contained the ATCAAA mutation was rescued from *Escherichia coli* CJ236 with the use of M13KO7 helper phage. The mutagenic nucleotide was GCCATTTCCATTTTATTCGTTGGATGTGG. Plasmids that had lost the *Sau*3A1 site at the hexanucleotide were identified, sequenced, and processed as described above. The GU-1 mutation, which creates a new *Eco*RI site, was created using the oligonucleotide CATAGCATCACTGATTGAATTCTAAACAGCCATT and the wild-type TS sequence as template. The GU-2 mutation was created using the oligonucleotide AGCATCACTGATTCAATTCTAAACAGC and GU-1 DNA as template.

To create the 3′ deletion mutations pTSMG3 +55 and +48, the insert of pTSMG3 was subcloned into pBS(+) and single-stranded DNA was isolated as above. Deletions were produced according to the rapid deletion subcloning method of Dale *et al.* (18), using the synthetic oligonucleotide CTCACTATAGGGCGAATTCCC. Clones bearing suitable deletions were identified by restriction analysis, and the exact deletion endpoints were determined by sequence analysis.

The deletion mutations pTSMG3 +40 and +32 were created by a deletion subcloning strategy using BAL-31 exonuclease. A 1.1-kb *Hin*dIII–*Eco*RI fragment containing the entire insert of pTSMG3 +48 was subcloned into the corresponding sites of the pBluescript (pSPT) KS(+) vector, leaving 58 bp of multiple cloning site between the *Eco*RI site and the terminal *Sac* I site. The plasmid was cut with *Eco*RI and digested with BAL-31 exonuclease for various times under conditions that minimize the rate of exonuclease digestion. DNA was then purified and digested at the penultimate restriction site (*Sac* II) so that the distal vector sequence would be identical in each deletion. The termini of the DNA were made blunt using T4 DNA polymerase in the presence of the four dNTPs. DNA was then ligated under conditions that favor unimolecular ligation (<5 μg/ml) and used to transform *E. coli* XL1-Blue. Plasmids bearing suitable deletions were identified by restriction mapping, and the exact deletion endpoints were determined by sequence analysis.

Cell Culture, Transfection, RNA Isolation, and S1 Nuclease Mapping. The TS$^-$ hamster V79 cell line (F88.13) (19) was cultured on plastic dishes (Falcon) in Dulbecco's modified Eagle's medium (GIBCO) supplemented with 10% NuSerum (Collaborative Research) and 10 μM thymidine. The cells were transiently transfected with 40 μg of supercoiled DNA (purified twice on CsCl gradients) by the calcium phosphate method (13, 14). The isolation of cytoplasmic or whole-cell RNA, oligo(dT)-cellulose chromatography, and S1 nuclease mapping were as described (10, 20).

RESULTS

To determine the sequences that are important for the polyadenylylation of mouse TS mRNA, TS minigenes containing various mutations were constructed and transiently transfected into cultured cells. The restriction maps of the wild-type minigenes and the sequences of the normal and mutant TS polyadenylylation signals are shown in Fig. 1. Poly(A)$^+$ mRNA was isolated from the transfected cells, and the patterns of polyadenylylation were analyzed by S1 nuclease mapping of the 3′ ends. The pattern of S1-resistant fragments observed with RNA isolated from TS$^-$ hamster V79 cells transfected with the wild-type mouse TS minigene (Fig. 2*A*, lane 3) was very similar to the pattern observed with normal mouse TS mRNA (lane 1). As shown previously (10, 20), the 210-nucleotide fragment 1 corresponds to TS mRNA that is polyadenylylated at the stop codon and accounts for >80% of the total signal. The 400-nucleotide fragment 3 corresponds to a minor polyadenylylation site located 190 nucleotides downstream of the termination codon. Fragment 2 represents protection of the probe to the 5′ end of the oligo(U) region beginning 32 nucleotides downstream from the processing site. Almost all of this fragment is due to S1 hypersensitivity of TS RNA molecules that extend beyond this region (20). No TS mRNA can be detected in the

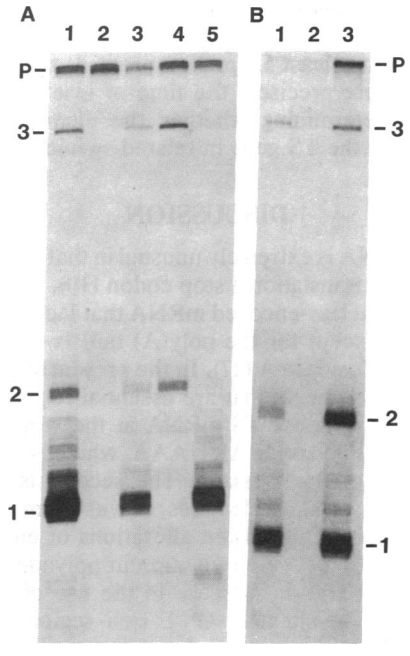

FIG. 2. S1 nuclease mapping of TS mRNA from cells transfected with TS minigenes. (*A*) Effect of upstream point mutations. Wild-type TS minigenes (pI56) or minigenes with point mutations in the upstream hexanucleotide were transiently transfected into TS$^-$ V79 cells. Cytoplasmic poly(A)$^+$ mRNA was isolated and analyzed in a 3′ S1 nuclease protection assay using a probe derived from the wild-type TS minigene. Lanes: 1, mRNA (10 μg) from mouse 3T6 cells; 2, mRNA (10 μg) from untransfected TS$^-$ hamster V79 cells; 3, mRNA (9 μg) from V79 cells transfected with the wild-type TS minigene; 4, mRNA (9 μg) from V79 cells transfected with TS minigenes containing the AUCAAA (hex-1) mutation; 5, mRNA (9 μg) from V79 cells transfected with TS minigenes containing the AAUAAA (hex-2) mutation. Positions of the probe (P) and fragments 1–3 are indicated. (*B*) Mouse polyadenylylation signal is functional in human cells. S1 nuclease protection assays were performed as in *A* with the following RNA preparations. Lanes: 1, cytoplasmic mRNA (50 ng) from mouse LU3-7 cells [a 3T6 cell line in which the TS gene is amplified 50-fold (21, 22)]; 2, total cellular poly(A)$^+$ RNA (8 μg) from HeLa cells; 3, total cellular poly(A)$^+$ RNA (8 μg) from HeLa cells transfected with the wild-type TS minigene (pI56).

untransfected TS$^-$ V79 cells (lane 2), demonstrating that the probe is specific for RNA derived from the transfected TS minigene.

When the mouse TS minigene containing the wild-type polyadenylylation signal was transfected into human (HeLa) cells, the mouse RNA encoded by the minigene underwent normal 3′ processing (Fig. 2*B*, lane 3). Thus the human polyadenylylation machinery is capable of recognizing the mouse polyadenylylation signal and adding poly(A) at the stop codon. Furthermore, since the human and mouse TS genes are almost identical from the upstream hexanucleotide through the downstream G+U-rich regions, this further supports the idea that sequences in addition to these are required for polyadenylylation at the stop codon.

Analysis of the Upstream Polyadenylylation Signal. Alteration of the highly conserved third position of the upstream polyadenylylation signal leads to inactivation of the signal (2). Changing the AUUAAA hexanucleotide of the mouse TS gene to AUCAAA virtually eliminated 3′ processing at the stop codon and led to increased use of downstream polyadenylylation sites (Fig. 2*A*, lane 4). This clearly demonstrates that the upstream element is the variant AUUAAA hexanucleotide.

Converting the upstream hexanucleotide to AAUAAA led to a significant increase in the intensity of fragment 1 (Fig. 2*A*, lane 5). Densitometric analysis revealed that the increase was about 2-fold for this experiment and even greater in subsequent experiments. In addition, there was a great reduction in usage of the downstream polyadenylylation sites (compare lane 5 with lanes 3 and 1). These observations demonstrate that the upstream element is a major factor in determining the amount of stable mRNA produced from the mouse TS gene.

Mutations in the G+U-Rich Sequence. A downstream G+U-rich region is an important component of the polyadenylylation signal for many mRNAs (1, 2). As previously noted (10), a G+U-rich sequence, GUGCUUU, is present beginning 2 nucleotides downstream of the stop codon of the mouse (as well as the human) TS gene (see Fig. 1*B*). To determine whether this sequence is important for the polyadenylylation reaction, we altered the sequence to AUUCAAU (GU-1 mutation) and analyzed the 3′ processing pattern. The mutation led to a dramatic reduction in processing at the stop codon and a large increase in use of downstream polyadenylylation sites (Fig. 3*A*, lane 2). It appears that the processing machinery preferred to use the new CA dinucleotide (at +6) that was introduced by the mutagenesis (or possibly the CA at +10) rather than the UA dinucleotide at the stop codon.

To avoid this problem, the GU-1 mutant was altered to form the GU-2 mutant, which has the sequence AUUGAAU in the G+U-rich region. Polyadenylylation at the stop codon was partially restored with this mutation, (Fig. 3*A*, lane 3), even though the sequence of the G+U-rich region had been severely altered. However, there was still considerable heterogeneity in the pattern of S1-resistant fragments. Thus it appears that the G+U-rich sequence immediately downstream of the stop codon is not essential for polyadenylylation at the stop codon, although it does play a role in defining the efficiency of utilization of this processing site. It is also possible that other G+U-rich regions farther downstream

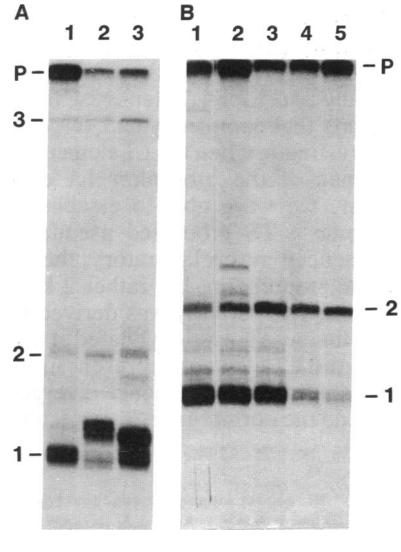

FIG. 3. Effects of downstream mutations. (*A*) Point mutations in the G+U-rich region. Wild-type (pI56) and mutant minigenes were transfected into TS$^-$ V79 cells and RNA was analyzed in a 3′ S1 nuclease assay. Lanes: 1, total cellular poly(A)$^+$ RNA (12 μg) from cells transfected with the wild-type TS minigene; 2, total cellular poly(A)$^+$ RNA (12 μg) from cells transfected with the GU-1 mutation and analyzed with an S1 probe derived from the mutant minigene; 3, same as lane 2 except that the cells were transfected with the TS minigene containing the GU-2 mutation. (*B*) Analysis of 3′ deletions. TS$^-$ V79 hamster cells were transfected with wild-type intronless TS minigene (lane 1) or derivatives that were deleted to +55 (lane 2), +48 (lane 3), + 40 (lane 4), or +32 (lane 5). Cytoplasmic poly(A)$^+$ RNA (8–9 μg) was analyzed with the wild-type probe.

(see Fig. 1) work in conjunction with those adjacent to the stop codon to facilitate 3′ processing at the stop codon.

The Oligo(U) Region Is an Important Downstream Signal. To determine whether the oligo(U) region is important for the polyadenylylation reaction, minigenes with deletions in the 3′ flanking region were analyzed. Deletions to +55 or +48, which leave the oligo(U) region intact, had little effect on 3′ processing (Fig. 3*B*, compare lanes 1–3) except for a slight increase in the amount of fragment 2. However, when the oligo(U) region was partially deleted (to position +40), normal 3′ processing was significantly impaired (lane 4). Deletion to position +32 (lane 5), which eliminates the entire oligo(U) region, led to nearly complete inhibition of polyadenylylation at the termination codon (fragment 1) and an increase in the intensity of fragment 2, which represents the region of divergence between the deleted minigene and the S1 probe. Presumably the RNA was being polyadenylylated at sequences encoded in the plasmid (23). In other experiments, fragment 1 was undetectable with the +32 mutation (data not shown). These analyses show that the oligo(U) region is essential for efficient 3′ processing of TS mRNA at the stop codon.

The Oligo(U) Region Was Introduced by a Mouse L1 Retroposon. The sequences of the mouse and human TS genes diverge shortly after the G+U-rich region (Fig. 1*B*). We considered the possibility that the divergence may have resulted from an insertion or deletion event at some time following the evolutionary divergence of the two species. Sequence comparisons (Fig. 4) revealed excellent agreement between the sequence downstream of the oligo(U) region and the consensus sequence at the 3′ end of a mouse L1 long interspersed repetitive element (LINE) (24). The L1 element is in opposite orientation relative to the direction of transcription of the TS gene, so the poly(A) segment at the 3′ end of the element corresponds to the oligo(U) region downstream of the mouse TS gene. The similarity extends for about 360 nucleotides, from the oligo(U) stretch through the *Sac* I site (Fig. 4) and about 140 nucleotides beyond (not shown), and then diverges. Flanking the region of L1 homology are CACC direct repeats. Although these repeats are uncharacteristically short for L1 elements, there are examples of L1 elements that completely lack direct repeats (25).

It is difficult to estimate when the L1 element was inserted, since the sequence of the progenitor L1 element is not known. However, we were able to establish a minimum estimate. In a mouse TS processed pseudogene recently isolated and sequenced in this laboratory, the poly(A) tail is not located at the stop codon, but rather 2 kb downstream (26). The pseudogene apparently was derived from a minor TS RNA species that was polyadenylylated at a distal site. As shown in Fig. 4, the L1 element is present in the processed pseudogene. By analyzing the sequence divergence between the pseudogene and the normal gene, it was estimated that the pseudogene was formed about 5.6 million years ago (26). Therefore, the L1 element has been present downstream of the TS gene for at least 5.6 million years. It may be possible to measure more precisely the time of insertion of the L1 element by determining whether the element is present downstream of the TS gene in related species.

DISCUSSION

Mouse TS mRNA is extremely unusual in that the poly(A) tail is added at the translational stop codon (10). The only other example of a nuclear-encoded mRNA that lacks a 3′ untranslated region [except for the poly(A) tail] is the mRNA for human α-galactosidase A (27). In the present study, we have identified two elements that are essential for efficient polyadenylylation of mouse TS mRNA at the stop codon. The first is the hexanucleotide AUUAAA, which is located in the coding region of the message. The second is an oligo(U) sequence that is 32 nucleotides downstream of the stop codon. Deletion or significant alterations of either of these sequences led to a severe reduction in polyadenylylation at the stop codon and an increase in the use of downstream polyadenylylation sites. A G+U-rich region immediately downstream of the stop codon does not appear to be essential for polyadenylylation at the stop codon, although it does appear to affect the efficiency of the reaction and the site of poly(A) addition. These observations explain why the human TS mRNA is not polyadenylylated at the stop codon. Even though the human and mouse TS genes are almost identical from the AUUAAA through the G+U-rich region, the human gene lacks the critical downstream oligo(U) sequence.

The most surprising observation was that the oligo(U) sequence corresponds to the poly(A) tail of a mouse L1 long interspersed repetitive element (LINE) that was integrated downstream of, and in opposite orientation to, the mouse TS gene at least 5 million years ago. Thus it appears that the polyadenylylation signal of the present mouse TS gene was created by the activation of a cryptic polyadenylylation signal following the insertion of the L1 element.

L1 elements are a complex family of DNA sequences that are repeated 10^4 to 10^5 times in mammalian genomes (25). The elements are thought to be inserted by a retrotransposition event—i.e., reverse transcription of an RNA species followed by integration at a chromosomal break. The intact, transcriptionally active L1 element(s) has not been identified. However, it is believed to be as large as 7 kb and to have two overlapping reading frames, one of which may encode a protein that has homology to reverse transcriptase (25, 28). Most L1 elements are truncated at the 5′ end, probably as a result of incomplete copying by reverse transcriptase (24). The L1 element 3′ to the mouse TS gene is an example of a severely truncated (360-nucleotide) element.

```
L:                                                              (t)ntttttttattagatattttcctcatttaca
                                                                          *
M: acgattaaaatggaaatggctgtttagagtgctttcagtgatgctatgaatatcaccgtttttttttttttaattaggtattttcctcatttaca
                 x  x           x               x        xx
P: acgattaaaatggagatagctgtttagagagctttcagtgatgctgtgaatatcgacg     ttttttttaattaggtattttcctcatttaca

L: tttcaaatgctatcccaaaagtcccccataccctcccctcccnactcccctacccacccactcccactttttggccctggcgttcccctgtactg
       *        *                                                      *           *     *
M: tttccaatgctataccaaaagtcccccataccctcccctcc  actcccctacccacccactcccacttcttggccctggcattcccttgtactg
                x          x x          xx                             x           x     x
P: tttccaatgctatcccaaaagtccgctataccctcccaccc  actcccctacccacccactcccacttattggccctggcgttcccctgtactg

L: gggcatataaagtttacaagtccaatgggcctctcttcccagtgatggccgactaggccatcttttgatacatatgcagctagagacangagctc
                  *                     *                                               *
M: gggcatataaagtttgcaagtccaatgggcctctctttccagtgatggccgactaggccatcttttgatacatatgcagctagagtcaagagctc
                                                             x
P: gggcatataaagtttgcaagtccaatgggcctctctttccagtgatggccgactaggctatcttttgatacatatgcagctagagtcaagagctc
```

FIG. 4. Comparison of the sequences of the 3′ ends of the mouse TS gene (M), processed pseudogene (P), and the mouse L1 repetitive element (L). The M and P sequences start 3 nucleotides upstream of the AUUAAA hexanucleotide and extend to the *Sac* I site at the 3′ end. The L1 sequence is shown in antisense orientation. Star, difference between L and M; x, difference between M and P.

There are many examples of L1 elements in the vicinity of expressed genes. In most cases, insertion of an L1 element within introns or distal flanking regions probably has little effect on gene expression (24, 25). Horie *et al.* (29) showed that a 3.6-kb human L1 element is present in the third intron of the functional human TS gene. However, if the element disrupts an open reading frame or an important regulatory signal, gene expression will usually be severely inhibited. Kazazian *et al.* (30) described two examples of hemophilia A that are caused by the insertion of a human L1 element into exon sequences of the factor VIII gene. A similar situation was observed in the *Drosophila* white locus (31). The L1 element inserted downstream of the mouse TS gene appears to be an extremely unusual case in which a portion of the repetitive sequence forms an essential part of a regulatory signal.

It is curious that in the mouse TS processed pseudogene (which also contains the L1 element) the poly(A) tail is not at the stop codon but rather 2 kb downstream (26). Although the template for the pseudogene may have been a minor TS mRNA that was polyadenylylated at a downstream site, it is also possible that the L1 insertion did not immediately lead to the formation of an efficient polyadenylylation signal. The oligo(U) region of the pseudogene is only 8 nucleotides long. When the oligo(U) region of the normal TS gene was shortened from 14 to 8 nucleotides, polyadenylylation at the stop codon relative to downstream sites was significantly reduced (Fig. 3*B*). If the oligo(U) region was only 8 nucleotides long at the time of the L1 insertion, this may have only partially activated the polyadenylylation signal; elongation of the oligo(U) region (and/or other nucleotide changes) may have been required to fully activate the signal. Unfortunately, since the length of repeated DNA regions changes rapidly during evolution (32), it is not possible to determine the length of the oligo(U) region at the time of insertion of the L1 element.

The upstream element of the TS polyadenylylation signal is unusual in that it is a variant of the canonical AAUAAA sequence and is located in the open reading frame. Our results show that the upstream hexanucleotide is a major factor in determining the amount of stable mRNA produced from a gene. Changing the AUUAAA to AAUAAA led to an increase in TS mRNA content and a great reduction in use of the minor downstream polyadenylylation sites. Since this is the only change in the TS minigene, the increase in mRNA content is almost certainly due to more efficient polyadenylylation. Our results are consistent with a previous *in vitro* analysis of the simian virus 40 late polyadenylylation signal, which showed that the AUUAAA hexanucleotide is only 20% as efficient as AAUAAA (33). Presumably the AAUAAA signal is recognized more efficiently by trans-acting factors (5–7, 34), which leads to more efficient polyadenylylation of TS heterogeneous nuclear RNA and consequently an increase in TS mRNA. Regulation of the efficiency of utilization of a weak polyadenylylation signal (by qualitative or quantitative changes in the polyadenylylation machinery) represents a potential mechanism for posttranscriptional regulation of gene expression. It will be interesting to determine whether this mechanism is relevant to the cell cycle regulation of the TS gene or other genes that have weak polyadenylylation signals.

We thank Drs. Haig Kazazian, Caroline Breitenberger, and Arthur Burghes for helpful discussions and comments on the manuscript.

These studies were supported by grants from the National Institute for General Medical Sciences (GM29356), the National Science Foundation (PCM-8312017), and the National Cancer Institute (CA16058). C.J.H. was supported by a National Institutes of Health Training Grant (CA09498).

1. Birnstiel, M. L., Busslinger, M. & Strub, K. (1985) *Cell* **41**, 349–359.
2. Manley, J. L. (1988) *Biochim. Biophys. Acta* **950**, 1–12.
3. Gil, A. & Proudfoot, N. J. (1987) *Cell* **49**, 399–406.
4. Moore, C. L., Chen, J. & Whoriskey, J. (1988) *EMBO J.* **7**, 3159–3169.
5. Wilusz, J. & Shenk, T. (1988) *Cell* **52**, 221–228.
6. Christofori, G. & Keller, W. (1988) *Cell* **54**, 875–889.
7. Takagaki, Y., Ryner, L. C. & Manley, J. L. (1988) *Cell* **52**, 731–742.
8. Wilusz, J., Feig, D. I. & Shenk, T. (1988) *Mol. Cell. Biol.* **8**, 4477–4483.
9. Perryman, S. M., Rossana, C., Deng, T., Vanin, E. F. & Johnson, L. F. (1986) *Mol. Biol. Evol.* **3**, 313–321.
10. Jenh, C.-H., Deng, T., Li, D., DeWille, J. W. & Johnson, L. F. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8482–8486.
11. Deng, T., Li, D., Jenh, C.-H. & Johnson, L. F. (1986) *J. Biol. Chem.* **261**, 16000–16005.
12. Takeishi, K., Kaneda, S., Ayusawa, D., Shimizu, K., Gotoh, O. & Seno, T. (1985) *Nucleic Acids Res.* **13**, 2035–2043.
13. DeWille, J. W., Jenh, C.-H., Deng, T., Harendza, C. J. & Johnson, L. F. (1988) *J. Biol. Chem.* **263**, 84–91.
14. Deng, T., Li, Y. & Johnson, L. F. (1989) *Nucleic Acids Res.* **17**, 645–658.
15. Taylor, J. W., Ott, J. & Eckstein, F. (1985) *Nucleic Acids Res.* **13**, 8764–8785.
16. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
17. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 488–492.
18. Dale, R. M. K., McClure, B. A. & Houchins, J. P. (1985) *Plasmid* **13**, 31–40.
19. Nussbaum, R. L., Walmsley, R. M., Lesko, J. G., Airhart, S. D. & Ledbetter, D. H. (1985) *Am. J. Hum. Genet.* **37**, 1192–1205.
20. DeWille, J. W., Harendza, C. J., Jenh, C.-H. & Johnson, L. F. (1989) *J. Cell. Physiol.* **138**, 358–366.
21. Rossana, C., Rao, L. G. & Johnson, L. F. (1982) *Mol. Cell. Biol.* **2**, 1118–1125.
22. Jenh, C.-H., Geyer, P. K., Baskin, F. & Johnson, L. F. (1985) *Mol. Pharmacol.* **28**, 80–85.
23. Zhang, F., Denome, R. M. & Cole, C. N. (1986) *Mol. Cell. Biol.* **6**, 4611–4623.
24. Voliva, C. F., Jahn, C. L., Comer, M. B., Hutchinson, C. A., III, & Edgell, M. H. (1983) *Nucleic Acids Res.* **11**, 8847–8859.
25. Singer, M. F. & Skowronski, J. (1985) *Trends Biochem. Sci.* **10**, 119–122.
26. Li, D. & Johnson, L. F. (1989) *Gene* **82**, 363–370.
27. Bishop, D. F., Kornreich, R. & Desnick, R. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3903–3907.
28. Loeb, D. D., Padgett, R. W., Hardies, S. C., Shehee, W. R., Comer, M. B., Edgell, M. H. & Hutchinson, C. A., III (1986) *Mol. Cell. Biol.* **6**, 168–182.
29. Horie, N., Nalbantoglu, J., Kaneda, S., Ayusawa, D., Seno, T. & Takeishi, K. (1989) *J. Biochem.* **106**, 1–4.
30. Kazazian, H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G. & Antonarakis, S. E. (1989) *Nature (London)* **332**, 164–166.
31. DiNocera, P. & Casari, G. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5843–5847.
32. Jeffreys, A. J., Wilson, V. & Thein, S. (1985) *Nature (London)* **314**, 67–73.
33. Wilusz, J., Pettine, S. M. & Shenk, T. (1989) *Nucleic Acids Res.* **17**, 3899–3908.
34. Hashimoto, C. & Steitz, J. A. (1986) *Cell* **45**, 581–591.