Stimulation of gene expression by introns: conversion of an inhibitory intron to a stimulatory intron by alteration of the splice donor sequence

Martin Korb⁺, Yunbo Ke and Lee F. Johnson^{*}

Departments of Molecular Genetics and Biochemistry, The Ohio State University, Columbus, OH 43210, USA

Received September 16, 1993; Revised and Accepted November 18, 1993

GenBank accession no. L11364

ABSTRACT

Efficient expression of many mammalian genes depends on the presence of at least one intron. We previously showed that addition of almost any of the introns from the mouse thymidylate synthase (TS) gene to an intronless TS minigene led to a large increase in expression. However, addition of intron 4 led to a reduction in minigene expression. The goal of the present study was to determine why TS intron 4 was unable to stimulate expression. Insertion of intron 4 into an intron-dependent derivative of the ribosomal protein L32 gene did not lead to a significant increase in expression, suggesting that its inability to stimulate expression was due to sequences within the intron. Deleting most of the interior of intron 4, improving the putative branch point, removing purines from the pyrimidine stretch at the 3' end of the intron, or removing possible alternative splice acceptor or donor sites within the intron each had little effect on the level of expression. However, when the splice donor sequence of intron 4 was modified so that it was perfectly complementary to U1 snRNA, the modified intron 4 stimulated expression approximately 6-fold. When the splice donor site of TS intron 1 (a stimulatory intron) was changed to that of TS intron 4, the modified intron 1 was spliced very inefficiently and lost the ability to stimulate mRNA production. Our observations support the idea that introns can stimulate gene expression by a process that depends directly on the splicing reaction.

INTRODUCTION

The primary transcripts of almost all mammalian genes are interrupted by introns which must be removed during the processing reactions that lead to the formation of mature mRNA. Much progress has been made in understanding the biochemistry of the splicing reaction (1,2). However, information about the biological significance of introns has emerged more slowly.

Introns have been shown to be directly responsible for a variety of qualitative changes in gene expression. For example, alternative splicing of introns from primary transcripts can give rise to multiple mRNA (and protein) species from a single gene (reviewed in 3).

Introns can also have quantitative effects on gene expression. Some introns contain promoter or enhancer elements and stimulate expression by increasing the rate of gene transcription (4–12). Regulating the efficiency of intron splicing also appears to play an important role in the autogenous regulation of yeast ribosomal protein genes (13), and in the growth-regulated expression of several mammalian genes for S-phase proteins (14–18). Studies with chimeric genes or with intronless derivatives of normal genes have shown that addition of an intron to an intronless gene can lead to a large increase in the level of expression (19–25). In many cases, the increase is not due to a change in the rate of transcription and is observed even when foreign or chimeric introns are inserted into the intronless genes. Buchman and Berg (26) showed that sequences located in the vicinity of the intron/exon boundaries are responsible for the stimulatory effects of some introns. These and earlier studies are consistent with the possibility that the splicing reaction itself is required for efficient production of a variety of mRNA species.

We have been studying the effects of introns on the expression of the mouse thymidylate synthase (TS) gene. The mouse TS gene is 12 kb in length and consists of a 1 kb open reading frame that is interrupted by 6 introns (27). Intronless derivatives of the mouse TS gene direct the production of small amounts of normal TS enzyme and mRNA when transiently or stably transfected into TS(–) hamster V79 cells (28). However, when introns 5 and 6 were included in the minigene at their normal locations in the coding region, expression increased approximately 8-fold. The stimulatory effect is not intron-specific since inclusion of introns 1 and 2, 3 and 4 or intron 3 alone also led to a 3–20-fold increase in the level of expression (29,30). The effects of the introns are not additive; rather, the level of expression of the minigene appears to be determined by the intron providing the highest degree of stimulation. Deletion of almost all of the interior

^{*} To whom correspondence should be addressed

⁺ Present address: Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104, USA

of introns 5 and 6 had no effect on their ability to increase TS gene expression, which argues against the presence of a transcriptional enhancer sequence within the introns.

Unexpectedly, inclusion of intact intron 4 alone at its normal position in the TS coding region led to a decrease in the level of expression relative to that observed with the intronless TS minigene (29). TS intron 4 represents a highly unusual example of an intron that is unable to stimulate expression of an intron-dependent gene. When another intron was included with intron 4 in the TS minigene, expression was elevated relative to the intronless minigene. This indicates that the inhibitory effect of intron 4 is recessive to the stimulatory effect of other introns.

The goal of the present study was to determine why intron 4 is unable to stimulate expression. Analyses of this unusual intron might provide unique insight into the mechanisms by which introns stimulate and inhibit gene expression. Our approach was to attempt to convert the intron into a stimulatory intron by deleting sequences within the intron that may be responsible for inhibiting expression or by making specific changes in sequences that are important for splicing. Our results suggest that the inhibitory effect of TS intron 4 is due to inefficient splicing of the intron which is caused by inefficient recognition of the splice donor sequence by U1 snRNA.

MATERIALS AND METHODS

Tissue culture and transfection

TS(−) V79 chinese hamster fibroblasts (31) and COS-1 cells (32) were cultured in Dulbecco's modified Eagle's medium (GIBCO) supplemented with 10% NuSerum (Collaborative Research). The medium for the V79 cells was also supplemented with 50 μ M thymidine. Cells were transiently transfected by the calcium phosphate method (28,29). Routinely, 12 pmole of the test plasmid and 3.0 μ g of control plasmid [pSV2cat, pSI₅₆S, or TI₅₆G(d)] were used per 10 cm tissue culture dish. The cultures were harvested for protein or RNA analysis two days after transfection.

Minigenes

The TS minigenes used in these studies contained 1 kb of 5' flanking DNA and 0.25 kb of 3' flanking DNA linked to the first and last exons of the TS coding region, respectively. TS minigenes that are cloned into pUC18 are designated with the prefix 'p' whereas those cloned into pBS(+) (Stratagene) have the prefix 'bs'. The intronless minigene (pTTT) has been described (30). pTI₁T contains intact intron 1 at its normal location in the coding region (18). In some cases, the minigenes were tagged with a 57 bp deletion between two adjacent *Bam*HI sites in exon 3. Such minigenes were indicated with the suffix '(d)'. The control minigene, pSI₅₆S, contains the SV40 early promoter and polyadenylation signals linked to the TS coding region that contains introns 5 and 6 at their normal locations. This minigene was constructed by replacing the TS polyadenylation signal of SI₅₆T (30) with the SV40 early polyadenylation signal, which leads to increased expression. The control TS minigene, pTI₅₆G(d), was described previously (18). It contains the human β -globin polyadenylation signal in place of the normal TS signal to boost the level of expression.

Intron-deficient derivatives of the mouse ribosomal protein L32 gene (21) were provided by Robert Perry. The rpL32 Δ 123 gene (Fig. 1) lacks all three introns, while rpL32 Δ 12 lacks introns 1 and 2. Δ 123-I1d2 (Fig. 1) was constructed by isolating the

Figure 1. Intronless and intron-containing rpL32 minigenes. The construction of the minigenes is described in Materials and methods. The black boxes represent the exons and the lines represent the intron or the 5' and 3' flanking regions. The $\Delta 12$ minigene retains intron 3 (I3) of the rpL32 gene at its normal location. The $\Delta 123$ minigene lacks all three introns. The TS introns I1d2 or I4 were inserted at the *Sma*I site (Sm) in exon 4. The end-labeled (*) probe (243 nucleotides) that was used in the S1 nuclease protection assay and the 83 nucleotide fragment protected by rpL32 mRNA are indicated. Abbreviations: Sc, *Sac*I; Sm, *Sma*I; Su, *Sau*3A.

*Sph*I–*Eco*RI fragment that contains the internally deleted TS intron 1 from the pTI_{1d2}T (see below), blunt-ending the fragment with T4 DNA polymerase, and ligating the fragment into the *Sma*I site in rpL32 Δ 123. The *Sph*I site is 22 nucleotides upstream of intron 1, while the *Eco*RI site, which was created by site directed mutagenesis (18), is 8 nucleotides downstream of the intron 1/exon 2 boundary. $\Delta 123$ -I4 was constructed by ligating a blunt-ended *Pvu*II–*Bgl*III fragment containing the entire TS intron 4 into the *Sma*I site of rpL32 Δ 123. The *Pvu*II site is 73 nucleotides upstream of intron 4, while the *Bgl*III site is 4 nucleotides downstream of the intron 4/exon 5 boundary.

S1 nuclease protection assays

Whole cell RNA was isolated using the guanidinium isothiocyanate procedure (33). The RNA was ethanol precipitated 3 times to remove any residual guanidinium salt. Total cytoplasmic RNA was isolated by lysing the cells with buffer containing Nonidet P-40, removing the nuclei by centrifugation and purifying the cytoplasmic RNA by phenol-chloroform extraction (34). In some experiments, poly(A)⁺ RNA was isolated by oligo(dT)–cellulose column chromatography (35). S1 nuclease protection assays were performed with total cytoplasmic RNA or with poly(A)⁺ RNA as described previously (29,36) using the end-labeled probes described in each figure. The reaction products were analyzed on a 6% polyacrylamide sequencing gel and quantitated with a Betascope 603 beta scanner (Betagen).

DNA sequencing

The sequence of intron 4 was determined in both directions by the Sanger procedure (37) using sequencing kits obtained from USB.

Enzyme assays

TS enzyme level was determined by measuring the formation of the covalent inhibitory ternary complex between TS, 5,10-methylene-tetrahydrofolate and [³H]-FdUMP and expressed as cpm [³H]-FdUMP bound/ μ g protein (38). Chloramphenicol acetyltransferase (CAT) enzyme activity was determined using a thin layer chromatography assay (29,39). The chromatograms were quantitated in the beta scanner.

(exon 4) GAACCCAAAAG

Deletion and site-directed mutagenesis

Intron 4 alterations: Restriction fragments that contained the region to be altered were cloned into pBS(+). Site-directed mutagenesis was performed by the procedure of Kunkel (40), using the positive strand of the plasmid as the template. Single-stranded, uracil-containing DNA was rescued from *E. coli* CJ236 with the use of VCS M13 helper phage (Stratagene). The mutated regions were sequenced to confirm the presence of the desired alteration and the absence of additional changes.

Mutations in sequences of intron 4 that were likely to be important for splicing were made using the following mutagenic oligonucleotides (in 5' to 3' orientation): improvement of the branch point (bp), GGTGGATATATCTGTTAGTAAAGGTAATAAG; creation of an upstream branch point (upbp), CAAGCCAACCAGTTAGTAAACAGCCAGC; removal of the potential alternate splice acceptor site (nag), GGTGGATATAGGGATGAGAAAGG; creation of a consensus U1 snRNA binding site (u1), CTGGGATGCTACTTACCTGTTGGGTTCC. Another mutation, in which all purines were changed to pyrimidines at the 3' end of intron 4 (pp), used C-TGAGGAGAGAAGGGGAAAGAGAGAGGGATGAG as the mutagenic oligonucleotide and intron 4 with the (nag) alteration as the template. The large internal deletion of intron 4 was created by removing a 469 bp *BalI*–*SylI* fragment from the intron, filling in the overhang with *E. coli* DNA polymerase (Klenow fragment) and ligating the blunt ends. Small internal deletions at the 5' end of intron 4 were created by linearizing pTI₄T or bsTI_{4u1}T with *BalI*. The product was incubated with Bal 31 for different times. The DNA was then ligated and transformed into *E. coli*. The deletion end points were determined by sequence analysis.

The altered sequences were introduced into the TS minigenes by exchanging a small restriction fragment that contained the altered region for the corresponding fragment from a TS minigene in the Bluescript vector. An appropriate region of the final construction was again sequenced to confirm the presence of the desired mutation. Minigenes that contained mutations at both the 5' and 3' ends of intron 4 were created by substituting the *SylI*–*BglII* fragment of the minigene that contained the u1 mutation in intron 4 with the corresponding fragment from minigenes that contained the desired mutations at the 3' end of intron 4.

Intron 1 alterations: pTI_{1d1}T, previously designated pTI_{1d}T (18), was derived from pTI₁T by deleting a *SmaI*–*PvuII* fragment from the interior of the intron. The deleted intron retains 100 nucleotides at the 5' end and 95 nucleotides at the 3' end of the intron. pTI_{1d2}T was derived from pTI₁T by deleting the *AatII*–*PvuII* fragment from the interior of the intron. The intron retains 35 nucleotides at the 5' end and 95 nucleotides at the 3' end of the intron. The 5' splice sequence of TS intron 1 in pTI_{1d1}T was changed to that of TS intron 4 using the mutagenic oligonucleotide CAGCCTGAAGGTGGACGGGGCTTGG. This minigene was named pTI_{1u1,d1}T.

RESULTS**Sequence of intron 4**

To facilitate the analysis of the sequences that might be responsible for the inability of intron 4 to stimulate expression, the complete nucleotide sequence of the intron was determined (Fig. 2). Lower case letters represent intron sequences while capital letters indicate the surrounding exon sequences. The intron is 673 bp in length. The 5' splice donor site (AAGgtggac) contains

```

gtggacagcatccagtcattttccttggataaaactggcgttagtgcacaaaggaagca * 60
*****
agggtggccactggaactgcctctaatacatgtagttgattgtgtacccttgagcaactcaa 120
BalI
tcacactttgctcttattttcacattcatattttgggggttaggctcatgtgatcctctt 180

tatgcttttatattacatgagcatttgaagtgtaatgttgaattgcctgtgaaaaacacc 240

tatgattttaatttttaatttggagcagatgaagtttcaagtgttaacgcttaactagat 300

tataaaactaaactctacaagggtatagttagtaccaggaaaggacatttgtatttata 360

ataacttaataacagatgtatctcacatgcagtttccctatggccttgcataatgagata 420

atttcccttttatttgggtccatcaacagtggtccttaactcttattttagctcagatgg 480

gagagcattgcttcagcataagtgtatacttgcagatagctaattgacaagtttccatgg 540
SylI
cctctgctggatggttgggaaatagtctttgaaggtatgaattgaggtgtgtaacctgc 600
+++++###
tggctgttttcatggttggcttgttgcctattttaccttctcatagatatatacaaccc 660

ctgtctcctcagATCTTCCCT (exon 5)

```

Figure 2. Nucleotide sequence of TS intron 4. Capital letters represent exon sequences and lower case letters represent intron 4 sequences. The potential 5' and 3' alternative splice sequences are indicated by * and # symbols, respectively. The putative branchpoint sequence is indicated with + symbols. The *BalI* and *SylI* restriction sites that were used to create internal deletions in intron 4 are also indicated.

only a 4 out of 9 base match to a perfect U1 binding domain (CAGgtaagt), which is the lowest match of all of the TS introns (27). Even allowing for G–U base pairs, the match is only 6 out of 9. The mismatches are on both sides of the splice junction. A closely related sequence (aaggtggcc), which might represent an alternative splice donor site, was also found beginning 60 nt downstream of the normal donor site (indicated by ** symbols). The intron lacks a yeast consensus branch point sequence (tactaac box) near the 3' end. However, a possible branchpoint sequence (ttctcat) near the 3' end of the intron (indicated by ++ symbols) was noted. It is flanked on both sides by polypyrimidine stretches. The 3' splice acceptor site (tctctcagAT) lies 28 nt downstream of the putative branchpoint and is preceded by a pyrimidine-rich region. A possible alternative 3' acceptor sequence (agat preceded by a pyrimidine-rich region, indicated by ## symbols) is immediately downstream of the putative branchpoint. A search of GenBank (Release 74.0) for sequences homologous to TS intron 4 was performed at the NCBI using the BLAST network service (41). This search revealed that there were several regions of significant homology with human TS intron 4.

Intron 4 does not stimulate expression of the intronless rpL32 gene

The inability of intron 4 to stimulate expression might be due to sequences within the intron or to the context of the intron within the TS minigene. To distinguish between these possibilities, the intact TS intron 4 was inserted into an intronless version of the mouse ribosomal protein gene rpL32 (Fig. 1). Expression of the rpL32 gene is highly dependent on the presence of introns (21). As a control, a stimulatory intron from the TS gene (an internally deleted version of intron 1, IId2) was inserted at the same location of the intronless rpL32 gene. A second control minigene

Figure 3. Stimulatory effects of TS intron 4 and intron 1 in the rpl32 gene. (A). The minigenes indicated below were transiently transfected into COS-1 cells. Total cytoplasmic RNA (approximately 25 μ g) was analyzed by S1 nuclease protection assays using a probe that extended from the *Sau3A* (Su) site in the coding region to the *SacI* (Sc) site in the 5' flanking region. Protection of the probe to the transcriptional start site should give a band of approximately 83 nucleotides. The positions of molecular weight markers (not shown) are indicated on the left. Lanes: 1, untransfected cos cells; 2, Δ 123; 3, Δ 12; 4, Δ 123-14; 5, Δ 123-Id2. (B). The expression of *Sl56S*, which was cotransfected to monitor transfection efficiency, was analyzed by S1 nuclease protection analysis. The same amounts of RNA were analyzed in panels A and B. The probe was derived from pTTT and was 5' end-labeled at the *BglII* (Bg) site in exon 5 and extended to the *XbaI* site in the 5' flanking region. The lane designations are the same as in panel A. No signal was detected in untransfected cells (not shown).

contained intact intron 3 of the rpl32 gene at its normal position in the coding region.

The intronless and intron-containing derivatives of the rpl32 gene were transiently transfected into COS-1 cells and the levels of expression were analyzed by S1 nuclease protection assays (Fig. 3). Quantitation of the radioactivity in the various bands using the beta scanner revealed that the presence of rpl32 intron 3 led to a 30-fold increase in mRNA level relative to the intronless rpl32 minigene, in good agreement with previously published results (21). Inclusion of internally deleted TS intron 1 increased expression approximately 8-fold, which is similar to the level of stimulation observed in the TS minigene (Table 1). However, inclusion of intron 4 increased expression less than 2-fold.

Figure 4. Effect of intron 4 alterations on TS minigene expression. Hamster TS(-) V79 cells were transiently cotransfected with a TS minigene that contained various intron modifications as well as a control TS minigene, *TI₅₆G(d)*, to correct for differences in transfection efficiency. Whole cell poly(A)⁺ RNA was analyzed by S1 nuclease protection assays using a probe that contained TS cDNA from the *PstI* (P) site at the 3' end of exon 5 to the *SphI* (Sp) site in the first exon. The probe was 5' end labeled at the *PstI* site and extended to the *NdeI* (N) in the plasmid (pUC18). mRNA derived from the test minigene protects a 548 nucleotide fragment while mRNA derived from the control minigene (with the small internal deletion) protects a 394 nucleotide fragment. The sizes of the protected fragments are indicated at the right, while the sizes of molecular weight markers are indicated at the left of the autoradiogram. Lanes 1, pTTT; 2, *pTI₄T*; 3, *bsTI_{4bp}T*; 4, *bsTI_{4upbp}T*; 5, *bsTI_{4nag}T*; 6, *bsTI_{4u1}T*; 7, *bsTI_{4u1,bp}T*; 8, *bsTI_{4u1,nag}T*. Other abbreviations: B, *BamHI*; Bg, *BglII*.

Table 1. Effect of intron 4 mutations on the level of expression

Minigene	TS enzyme level	TS mRNA level
pTTT	1.0	1.0
bsTTT	1.0	1.0
pTI ₄ T	0.4 ± 0.1	0.6 ± 0.1
pTI _{4d} T	0.2 ± 0.14	0.4 ± 0.2
pTI _{4(-26,+34)} T	0.7 ± 0.06	0.9 ± 0.2
bsI _{4bp} T	0.6 ± 0.14	1.1 ± 0.2
bsTI _{4nag} T	0.1 ± 0.07	0.5 ± 0.04
bsTI _{4pp} T	0.6 ± 0.21	0.9 ± 0.3
bsTI _{4upbp} T	0.4 ± 0.11	0.6 ± 0.1
bsTI _{4u1} T	5.9 ± 1.5	7.7 ± 1.9
bsTI _{4u1,bp} T	8.0 ± 1.6	6.0 ± 1.3
bsTI _{4u1,nag} T	4.9 ± 1.8	3.2 ± 0.7
bsTI _{4u1,pp} T	6.2 ± 1.2	4.7 ± 1.1
bsTI _{4u1,upbp} T	7.2 ± 1.9	5.8 ± 1.3
bsTI _{4u1(-9,0)} T	3.4 ± 0.4	4.7 ± 1.5
pTI ₁ T	13 ± 3.5	
pTI _{1d1} T	15 ± 2.0	
pTI _{1d2} T	12 ± 2.0	

Values are normalized to that observed with the intronless minigene and represent the averages (± standard deviation) of at least three independent experiments.

Therefore, it appears that sequences within intron 4, or immediately adjacent to the intron, are responsible for its inability to stimulate gene expression.

Deletion of the interior of intron 4 has no effect on expression

The inhibitory effects of intron 4 could be due to the presence of internal sequences that had an inhibitory effect on the

A. MUTATIONS NEAR THE 3' END OF INTRON 4

```

          ++++++####
tttca---tggttggtgtgtgtgtctatctttctcatagatatatcaaacccctgtctcctcagATC
      tactaact (upbp)          tactaaga (bp)
                                catgcctat (nag)
                                catgcctgtctctttcccttgt (pp)

```

B. MUTATIONS NEAR THE 5' END OF INTRON 4

```

CCCCAAAGgtggacagcat
      AACAGgtaaatag (u1)

```

C. MUTATIONS NEAR THE 5' END OF INTRON 1

```

CCTGAGAGgtaactggggc
      CTGAGgtgaagggg (I1u1)

```

Figure 5. Site-Directed Mutagenesis of TS Introns 4 and 1. The top lines of panels (A), (B) and (C) represent the wild-type sequence near the intron-exon junctions. Capital letters represent exon sequences and lower case letters represent intron sequences. The sequences indicated by the + symbols and # symbols represent the putative branchpoint sequence and the adjacent potential alternative 3' splice site. The changes that were introduced by site-directed mutagenesis are underlined. The abbreviations used to identify each alteration are: upbp, upstream branchpoint; bp, branchpoint; nag, no alternative 3' splice site; pp, perfect polypyrimidine stretch; u1, perfect u1 snRNA binding site; I1u1, conversion of the u1 binding site of intron 1 to that of intron 4.

production of cytoplasmic mRNA. For example, the intron might contain sequences that block transcriptional elongation, target the primary transcript for rapid turnover, or prevent efficient processing. To explore this possibility, the interior of intron 4 between the *BalI* and *SylI* sites (see Fig. 2) was deleted to form pTI_{4d}T. The minigene was transiently transfected into TS(-) V79 hamster cells and the effect on minigene expression was determined. The results of these analyses showed that deletion of the interior of intron 4 did not convert the intron into a stimulatory intron. Instead, the deletion led to a slight decrease in TS enzyme level and mRNA content (Table 1).

The intron 4 internal deletion did not remove the alternative splice donor site that is located 62 nucleotides downstream of the normal donor site (Fig. 2). To determine if removal of this alternative site would affect the level of expression, an intron 4-containing minigene was created in which intron sequences that encompassed the alternative splice donor site were deleted. The small internal deletion, which was created by *Bal* 31 digestion of pTI₄T at the *BalI* site (see Fig. 2), eliminated 26 nucleotides 5' to the *BalI* site and 34 nucleotides 3' of the *BalI* site and is designated pTI_{4(-26,+34)}T. This deletion also did not convert intron 4 into a stimulatory intron (Table 1).

Effects of point mutations at the splice sequences of intron 4

The above observations indicate that the sequences responsible for the inability of intron 4 to stimulate expression are located in close proximity to the intron-exon boundaries and may correspond to sequences that are important for splicing. In particular, the intron may be spliced inefficiently as a result of suboptimal splice sequences. To test this possibility, we altered the splicing signals at both ends of the full length intron so that they were in agreement with the consensus sequences and determined if any of the improvements converted the intron into a stimulatory intron. The mutations that were introduced are summarized in Fig. 5.

A

B

Figure 6. Inactivation of the U1 snRNA binding site of intron 1 prevents splicing. (A) Hamster TS(-) V79 cells were transiently transfected with the following minigenes: lane 1, pTI_{1u1,d1}T; lane 2, pTI_{1d1}T; lane 3, pTI_{1s6}T. Total cytoplasmic RNA was analyzed by S1 nuclease protection assays. The probe, which was derived from pTI_{1u1,d1}T, was 5' end-labeled at the *Bam*HI (B) site in exon 3 and extended to the *Xba*I (X) site in the 5' flanking region. The band at 84 nucleotides corresponds to RNA that is properly spliced, while the band at 490 nucleotides corresponds to transcripts that retain the intron, as indicated in the diagram. The positions of molecular weight markers are indicated at the left. (B) Hamster TS(-) V79 cells were transiently transfected with 12 pmoles of: lane 1, pTI_{1d1}T(d); lane 2, pTI_{1u1,d1}T(d); lane 3, pTTT(d), as well as 2 pmoles of the control gene, pSI₅₆S. Total cytoplasmic RNA was analyzed by S1 nuclease protection assays. The probe, which was derived from the control gene and which retains the *Bam*HI fragment, was 5' end-labeled at the *Bgl*II (G) site in exon 5 and extends to the *Nde*I (N) site in the vector. The band at 218 nucleotides corresponds to RNA derived from the test minigenes that lack the *Bam*HI fragment while the band at 640 nucleotides corresponds to transcripts that are derived from the pSI₅₆S.

Since earlier studies showed that the yeast consensus branchpoint (tactaac) is very efficiently recognized in mammalian systems (42), we introduced this sequence at two locations near the 3' end of the intron. First, a tactaac box was inserted upstream of the putative branch point to form the minigene bsTI_{4upbp}T. Second, the putative branch point sequence (ttctcat) was converted to a tactaac box to form bsTI_{4bp}T. Third, the 'agat' sequence

that is adjacent to the putative branchpoint was altered to form the minigene bsTI_{4nag}T. The nag mutation might increase splicing efficiency by eliminating steric hinderance between splicing components that bind to a branchpoint sequence and the acceptor site or by eliminating competition between the two acceptor sites. Fourth, all purines in the polypyrimidine stretch preceding the wild-type acceptor site were changed into pyrimidines to form the minigene bsTI_{4pp}T, which was shown to increase splicing efficiency in other genes (43). Unexpectedly, none of the alterations at the 3' end of intron 4 resulted in a significant increase in TS enzyme level or mRNA content (Fig. 4 and Table 1).

As mentioned earlier, the 5' splice donor site of intron 4 exhibits only a 4 out of 9 match to the consensus sequence, which is complementary to U1 snRNA. This sequence was changed so that it would be completely complementary to U1 snRNA (9 out of 9 match) to create the minigene bsTI_{4u1}T. This change resulted in a 6-fold increase in TS enzyme level relative to the intronless minigene, or a 15-fold increase relative to the minigene that contained wild type intron 4. A corresponding increase in mRNA content was also observed (Fig. 4 and Table 1).

To determine if the level of stimulation could be further increased, double mutations were analyzed in which the bsTI_{4u1}T mutation was used in combination with mutations at the 3' end of the intron. No further increases were observed with any of the double mutations that were analyzed (Fig. 4 and Table 1). In addition, deletion of 9 nucleotides upstream of the *BalI* site in pTI_{4u1}T (which eliminated the potential alternative splice donor site) to create bsTI_{4u1(-9,0)}T also did not lead to a further increase in expression.

Alteration of the splice donor site of intron 1

The sequence of the splice donor site of TS intron 1, a stimulatory intron, is GAGgtaact, which is a 7 out of 9 match to the consensus sequence. To further explore the importance of the U1 site for the intron stimulatory effect, the splice donor sequence of TS intron 1 was changed to AAGgtggac, which is the same as that found in intron 4. The internally deleted version of intron 1 (I1d1) was used in this analysis to facilitate the construction of the minigene, which was designated pTI_{1u1,d1}T. The alteration also led to a 1 base deletion in exon 1, so it is not possible to measure TS enzyme levels in experiments involving this mutation.

As shown in Fig. 6(A), S1 nuclease protection analyses of the cytoplasmic RNA derived from pTI_{1u1,d1}T revealed that the U1 mutation in intron 1 inhibited the splicing of this intron. There was almost no radioactivity at 84 nucleotides, corresponding to normally spliced mRNA. (The small amount of radioactivity at this position does not co-migrate with the band derived from normally spliced TS RNA and may represent background at this position). Instead there was a band at 494 nucleotides, corresponding to RNA that retained the intact intron. Quantitative analysis of cytoplasmic TS RNA (Fig. 6B) revealed that the pTI_{1u1,d1}T minigene was expressed at about the same level as the intronless TS minigene pTTT. In three determinations, the expression level of pTI_{1u1,d1}T was 1.27 ± 0.4 relative to pTTT. Therefore, the U1 alteration of intron 1 eliminated the ability of the intron to stimulate production of TS RNA.

We have also performed similar studies on a TS minigene that contains an internally deleted version of TS intron 2, which is also a stimulatory intron. We have changed the splice donor site of intron 2 from AAGgtaaggg to AGCtaaggg, which should completely inactivate the splice donor site. Unfortunately, this

alteration did not prevent splicing of the intron, presumably due to the activation of a cryptic splice donor site. Nevertheless, expression of this minigene was reduced about 2-fold relative to the minigene that contained the normal splice donor site (data not shown).

DISCUSSION

In the present study, we have taken advantage of the novel properties of TS intron 4 to gain additional insight into the mechanisms by which introns stimulate gene expression. We show that insertion of intron 4 into an intronless version of the mouse rpl32 gene had little effect on expression, whereas insertion of TS intron 1 led to a large increase in expression. This indicates that the inhibitory effects of intron 4 are most likely due to sequences within the intron. Deletion of most of the interior of intron 4 did not convert it to a stimulatory intron, indicating that the defect is related to sequences that are near the intron-exon boundary. These observations are consistent with the possibility that the stimulatory effect is related to the splicing signals as suggested in earlier studies with other intron-dependent genes.

The sequences that are important for splicing of TS intron 4 do not conform to the consensus sequences (44). The polypyrimidine stretch at the 3' end contains a number of purines; the sequences in the vicinity of the putative branch point and 5' donor site are not complementary to U2 and U1 snRNAs, respectively; and there are potential alternative donor and acceptor sequences near the 5' and 3' ends of the intron. Earlier studies have shown that improving the branch site or the splice donor site, removing pyrimidines from the purine stretch preceding the splice acceptor site, or deleting potential alternative splice sites can lead to an increase in efficiency of splicing (42,43,45-47). Therefore, it was surprising that most of the improvements of these sequences in TS intron 4 had no detectable effect on the ability of the intron to stimulate gene expression.

The only modification that converted intron 4 to a stimulatory intron was modification of sequences at the splice donor site. Intron 4 has a poor (4 out of 9) match to the consensus sequence at the splice donor site, which is complementary to nucleotides near the 5' end of U1 snRNA. When the sequence in this region was changed so that it represented a perfect consensus sequence (9 out of 9 match), the intron was converted to a stimulatory intron.

The U1 modification is unlikely to affect promoter strength or stability of the spliced RNA. The most likely explanation for our observations is that the change has improved the efficiency of the splicing reaction. This is in line with earlier *in vitro* splicing analyses with other introns which have shown that changes in the splice donor site that vary the degree of complementarity with U1 snRNA can have significant effects on splicing (46,48-50).

Another possibility is that the stimulatory effect is simply due to the presence of a U1 snRNA binding site within the transcript. This alternative appears unlikely in view of the fact that the TS open reading frame already contains several regions that show reasonable complementarity with U1 snRNA. These include the sequences CAGGTGGAG (between 88 and 96, which represents a 5 out of 9 match with a perfectly complementary sequence), GAGGTCAGG (between 624 and 632, 6 out of 9 match) and CAGGTGATT (between 716 and 724, 7 out of 9 match). The extent of match is at least as good as that found at the splice donor sites of many introns.

To further explore the role of the splice donor site in the stimulatory process, the donor site of a stimulatory intron (TS intron 1) was modified so that it was the same as that found in TS intron 4. This alteration led to nearly complete inhibition of formation of normally spliced mRNA, although unspliced cytoplasmic RNA that retained the modified intron was readily detected. The amount of unspliced RNA was similar to the amount of RNA derived from the intronless TS minigene. These observations further strengthen the idea that there is a correlation between splicing efficiency and the stimulatory effect of introns.

It is interesting that the same splice donor sequence had somewhat different effects in two different introns. In intron 4, the splice donor sequence was capable of participating in accurate splicing, albeit at a low efficiency. However, when the same sequence was introduced into intron 1, it was no longer capable of participating in splicing. Furthermore, cytoplasmic transcripts that retained the modified intron 1 were readily observed, while transcripts that retained intron 4 were not observed. Presumably, other sequences within the intron (or perhaps in the surrounding exons) affect the splicing efficiency of the intron or the stability or cytoplasmic export of the intron-containing transcripts (51).

It is not clear how the addition of an efficiently spliced intron to an intronless minigene leads to an increase in gene expression. Addition of an intron substantially increases the complexity of the reactions that are necessary for conversion of the initial transcript into a mature cytoplasmic mRNA. Therefore, it might seem unlikely that this would lead to increased expression. There are several possible mechanisms by which addition of an intron might increase mRNA production. The splicing reaction itself may stabilize the initial transcript by removing nuclease target sequences that are present in the intron or by masking these target sequences. In the case of intron 4, these target sequences may be inefficiently masked or removed since spliceosomes may assemble very inefficiently on TS transcripts that contain only intron 4. In line with this, we have never been able to detect (by Northern blot or S1 protection analyses) the accumulation of intron 4-containing transcripts within the nuclear compartment. It appears that the transcripts are either spliced or destroyed rapidly after they are transcribed. The fact that transcripts that retain modified intron 1 accumulate in the cytoplasm may be due to the deletion or inactivation of sequences that are responsible for nuclear retention or for rapid destruction of the intron.

Splicing of mRNA precursors is believed to occur in association with discrete structural features of the nucleus (1,52). Pederson's laboratory has recently shown that intron-containing globin RNA is localized to discrete regions in the nucleus which are enriched in splicing factors whereas intron-lacking transcripts do not localize to these regions (53). Thus splicing reactions may be compartmentalized within the nucleus. Lawrence and coworkers (54) have shown that mRNA precursors are localized to well defined tracks in the nucleus that overlap with the site of transcription, and that splicing appears to occur along these tracks. There is also good evidence for coupling between the splicing and polyadenylation reactions (55–58), suggesting that all of the RNA processing reactions may occur within the same compartment. Transcripts within these compartments may be protected from nuclease attack and be efficiently processed and exported to the cytoplasm whereas transcripts outside of these compartments may be inefficiently processed and readily attacked by nucleases. Interaction of the nascent, intron-containing transcript with U1 snRNP at the splice branch point (one of the initial steps in the splicing reaction) may be important for directing

the transcript to these specialized regions. Transcripts that lack introns, or that are poorly recognized by U1 snRNP, may not be targeted to these regions and may therefore be more susceptible to degradation within the nucleus.

The necessity of introns for efficient mRNA production is not universal since some mammalian cellular and viral genes naturally lack introns and are expressed at high levels (59–62), and removal of the introns from certain genes has little effect on their levels of expression, at least in cultured cells (23,63). Furthermore, the ability of introns to stimulate expression also appears to depend on the nature of the promoter that is linked to the minigene (64). These observations suggest that there are at least two pathways for production of cytoplasmic mRNA in mammalian cells. One depends on the presence of introns while the other does not. Elucidation of the differences in these pathways may lead to additional insight into the complexities of gene regulation in mammalian cells.

ACKNOWLEDGEMENTS

We thank R.P. Perry, Y. Li, and T. Deng for plasmids and M. Graham for skilled technical assistance. These studies were supported by grants from the National Institute for General Medical Sciences (GM29356), the American Cancer Society (CB-71) and the National Cancer Institute (CA16058).

REFERENCES

1. Sharp, P.A. (1987) *Science* 235, 766–771.
2. Green, M.R. (1986) *Ann. Rev. Genet.* 20, 671–708.
3. Breitbart, R.E., Andreadis, A. and Nadal-Ginard, B. (1987) *Ann. Rev. Biochem.* 56, 467–495.
4. Atchison, M.L., Meyuhas, O. and Perry, R.P. (1989) *Mol. Cell. Biol.* 9, 2067–2074.
5. Gillies, S.D., Morrison, S.L., Oi, V.T. and Tonegawa, S. (1983) *Cell* 33, 717–728.
6. Banerji, J., Olson, L. and Schaffner, W. (1983) *Cell* 33, 729–740.
7. Rossi, P. and De Crombrughe, B. (1987) *Proc. Natl. Acad. Sci. USA* 84, 5590–5594.
8. Bornstein, P. and McKay, J. (1988) *J. Biol. Chem.* 263, 1603–1606.
9. Horton, W., Miyashita, T., Kohno, K., Hassell, J.R. and Yamada, Y. (1987) *Proc. Natl. Acad. Sci. USA* 84, 8864–8868.
10. Konieczny, S.F. and Emerson Jr, C.P. (1987) *Mol. Cell. Biol.* 7, 3065–3075.
11. Coulombe, B., Ponton, A., Daigenault, L., Williams, B.R.G. and Skup, D. (1988) *Mol. Cell. Biol.* 8, 3227–3234.
12. Slater, E.P., Rabenau, O., Karin, M., Baxter, J.D. and Beato, M. (1985) *Mol. Cell. Biol.* 5, 2984–2992.
13. Dabeva, M.D., Post-Beittenmiller, M.A. and Warner, J.R. (1986) *Proc. Natl. Acad. Sci. USA* 83, 5854–5857.
14. Ottavio, L., Chang, C.-D., Rizzo, M.-G., Travali, S., Casadevall, C. and Baserga, R. (1990) *Mol. Cell. Biol.* 10, 303–309.
15. Alder, H., Yoshinouchi, M., Prystowsky, M.B., Appasamy, P. and Baserga, R. (1992) *Nucleic Acids Res.* 20, 1769–1775.
16. Takayanagi, A., Kaneda, S., Ayusawa, D. and Seno, T. (1992) *Nucleic Acids Res.* 20, 4021–4025.
17. Gudas, J.M., Knight, G.B. and Pardee, A.B. (1988) *Proc. Natl. Acad. Sci. USA* 85, 4705–4709.
18. Ash, J., Ke, Y., Korb, M. and Johnson, L.F. (1993) *Mol. Cell. Biol.* 13, 1565–15710.
19. Hamer, D.H. and Leder, P. (1979) *Cell* 18, 1299–1302.
20. Buckman, A.R. and Berg, P. (1984) *Mol. Cell. Biol.* 4, 1915–1928.
21. Chung, S. and Perry, R.P. (1989) *Mol. Cell. Biol.* 9, 2075–2082.
22. Callis, J., Fromm, M. and Walbot, V. (1987) *Genes Dev.* 1, 1183–1200.
23. Brinster, R.L., Allen, J.M., Behringer, R.R., Gelinas, R.E. and Palmiter, R.D. (1988) *Proc. Natl. Acad. Sci. USA* 85, 836–840.
24. Choi, T., Huang, M., Gorman, C. and Jaenish, R. (1991) *Mol. Cell. Biol.* 11, 3070–3074.
25. Palmiter, R.D., Sandgren, E.P., Avarbock, M.R., Allen, D.D. and Brinster, R.L. (1991) *Proc. Natl. Acad. Sci. USA* 88, 478–482.

26. Buckman, A.R. and Berg, P. (1988) *Mol. Cell. Biol.* 8, 4395–4405.
27. Deng, T., Li, D., Jenh, C.-H. and Johnson, L.F. (1986) *J. Biol. Chem.* 261, 16000–16005.
28. DeWille, J.W., Jenh, C.-H., Deng, T., Harendza, C.J. and Johnson, L.F. (1988) *J. Biol. Chem.* 263, 84–91.
29. Deng, T., Li, Y. and Johnson, L.F. (1989) *Nucleic Acids Res.* 17, 645–658.
30. Li, Y., Li, D., Osborn, K. and Johnson, L.F. (1991) *Mol. Cell. Biol.* 11, 1023–1029.
31. Nussbaum, R.L., Walmsley, R.M., Lesko, J.G., Airhart, S.D. and Ledbetter, D.H. (1985) *Am. J. Hum. Genet.* 37, 1192–1205.
32. Gluzman, Y. (1981) *Cell* 23, 175–182.
33. Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* 162, 156–159.
34. Johnson, L.F., Abelson, H.T., Green, H. and Penman, S. (1974) *Cell* 1, 95–100.
35. Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* 69, 1408–1412.
36. Deng, T., Li, Y., Jolliff, K. and Johnson, L.F. (1989) *Mol. Cell. Biol.* 9, 4079–4082.
37. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
38. Rossana, C., Rao, L.G. and Johnson, L.F. (1982) *Mol. Cell. Biol.* 2, 1118–1125.
39. Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) *Mol. Cell. Biol.* 2, 1044–1051.
40. Kunkel, T.A. (1985) *Proc. Natl. Acad. Sci. USA* 82, 488–492.
41. Altschul, S.F., Gish, W., Miller, W. and Myers, E.W. (1990) *J. Mol. Biol.* 215, 403–410.
42. Zhuang, Y., Goldstein, A.M. and Weiner, A.M. (1989) *Proc. Natl. Acad. Sci. USA* 86, 2752–2756.
43. Dominski, Z. and Kole, R. (1991) *Mol. Cell. Biol.* 11, 6075–6083.
44. Shapiro, M.B. and Senapathy, P. (1987) *Nucleic Acids Res.* 15, 7155–7174.
45. Dominski, Z. and Kole, R. (1992) *Mol. Cell. Biol.* 12, 2108–2114.
46. Zhuang, Y. and Weiner, A.M. (1986) *Cell* 46, 827–835.
47. Spritz, R.A., Jagadeeswaran, P., Choudary, P.V., Biro, P.A., Elder, J.T., deRiel, J.K., Manley, J.L., Gefter, M.L., Forget, B.G. and Weissman, S.M. (1981) *Proc. Natl. Acad. Sci. USA* 78, 2455–2459.
48. Zhuang, Y., Leung, H. and Weiner, A.M. (1987) *Mol. Cell. Biol.* 7, 3018–3020.
49. Kreivi, J.-P., Zefrivity, K. and Akusjärvi, G. (1991) *Nucleic Acids Res.* 19, 6956–6956.
50. Eperon, L.P., Estibeiro, J.P. and Eperon, I.C. (1986) *Nature (London)* 324, 280–282.
51. Reed, R. and Maniatis, T. (1986) *Cell* 46, 681–690.
52. Zeitlin, S., Parent, A., Silverstein, S. and Efstratiadis, A. (1987) *Mol. Cell. Biol.* 7, 111–120.
53. Wang, J., Cao, L.-G., Wang, Y.-L. and Pederson, T. (1991) *Proc. Natl. Acad. Sci. USA* 88, 7391–7395.
54. Xing, Y., Johnson, C.V., Dobner, P.R. and Lawrence, J.B. (1993) *Science* 259, 1326–1330.
55. Niwa, M., Rose, S.D. and Berget, S.M. (1990) *Genes Dev.* 4, 1552–1559.
56. Niwa, M. and Berget, S.M. (1991) *Genes Dev.* 5, 2086–2095.
57. Pandey, N.B., Chodchay, N., Liu, T.-J. and Marzluff, W.F. (1990) *Nucleic Acids Res.* 18, 3161–3170.
58. Huang, M.T.F. and Gorman, C.M. (1990) *Nucleic Acids Res.* 18, 937–947.
59. Hentschel, C.C. and Birnstiel, M.L. (1981) *Cell* 25, 301–313.
60. Hunt, C. and Morimoto, R.I. (1985) *Proc. Natl. Acad. Sci. USA* 82, 6455–6459.
61. Lengyel, P. (1982) *Ann. Rev. Biochem.* 51, 251–282.
62. McKnight, S.L. (1980) *Nucleic Acids Res.* 8, 5949–5966.
63. Gross, M.K., Kainz, M.S. and Merrill, G.F. (1987) *Mol. Cell. Biol.* 7, 4576–4581.
64. Neuberger, M.S. and Williams, G.T. (1988) *Nucleic Acids Res.* 16, 6713–6724.