MOLECULAR AND CELLULAR BIOLOGY, Aug. 1997, p. 4562–4571 Vol. 17, No. 8
0270-7306/97/$04.00+0

G Triplets Located throughout a Class of Small Vertebrate Introns Enforce Intron Borders and Regulate Splice Site Selection

ANDREW J. MCCULLOUGH,* AND SUSAN M. BERGET

Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, Houston, Texas 77030

Received 19 March 1997/Returned for modification 2 May 1997/Accepted 23 May 1997

Splicing of small introns in lower eucaryotes can be distinguished from vertebrate splicing by the inability of such introns to be expanded and by the inability of splice site mutations to cause exon skipping—properties suggesting that the intron rather than the exon is the unit of recognition. Vertebrates do contain small introns. To see if they possess properties similar to small introns in lower eucaryotes, we studied the small second intron from the human α-globin gene. Mutation of the 5′ splice site of this intron resulted in in vivo intron inclusion, not exon skipping, suggesting the presence of intron bridging interactions. The intron had an unusual base composition reflective of a sequence bias present in a collection of small human introns in which multiple G triplets stud the interior of the introns. Each G triplet represented a minimal sequence element additively contributing to maximal splicing efficiency and spliceosome assembly. More importantly, G triplets proximal to a duplicated splice site caused preferential utilization of the 5′ splice site upstream of the triplets or the 3′ splice site downstream of the triplets; i.e., sequences containing G triplets were preferentially used as introns when a choice was possible. Thus, G triplets internal to a small intron have the ability to affect splice site decisions at both ends of the intron. Each G triplet additively contributed to splice site selectivity. We suggest that G triplets are a common component of human 5′ splice sites and aid in the definition of exon-intron borders as well as overall splicing efficiency. In addition, our data suggest that such intronic elements may be characteristic of small introns and represent an intronic equivalent to the exon enhancers that facilitate recognition of both ends of an exon during exon definition.

Vertebrate genes, on average, have small exons and much larger introns (15). Splicing of pre-mRNA precursors from such genes, therefore, requires recognition of the smaller exonic domains without interference from similar looking sequences within larger introns. This task is aided both by the concerted recognition of the splice sites bordering an exon during what is termed "exon definition" (2, 4, 36, 37) and by the binding of accessory splicing factors to additional enhancer sequences located within or near exons (3, 6, 7, 11, 16–18, 21, 23, 24, 30, 35, 38, 39, 41, 43–45, 48, 50, 52, 53). Exon definition and recognition of enhancer sequences occur during the earliest steps of spliceosome assembly, thereby committing regions of a precursor RNA for use as exons. One of the hallmarks of splicing in an exon definition mechanism is the phenotype of 5′ splice site mutations (2, 4, 36, 37, 46). Such mutations result in inactivation of the exon that they border, leading to exon skipping.

The pairing of splice sites across vertebrate exons during early spliceosome formation stands in strong contrast to the situation occurring in precursor RNAs in which introns are smaller than exons. In lower eucaryotes such as *Drosophila melanogaster* or *Schizosaccharomyces pombe*, introns are often only 50 to 100 nucleotides and flanking exons are larger. Splicing in these pre-mRNAs does not utilize exon definition and the pairing of splice sites across exons. Instead, concerted recognition of the splice sites located at the termini of the small introns occurs during early spliceosome assembly, a process we have termed "intron definition" in this report (2, 14, 19, 20, 29, 51). In these introns, 5′ splice site mutations lead to intron inclusion instead of exon skipping.

Examination of gene architecture from flies to humans indicates that small and large introns can be mixed within individual genes. In vertebrates, exon and intron expansion experiments have indicated that vertebrate splicing signals can support initial pre-mRNA recognition via either exonic or intronic pairing of splice sites (42). In lower eucaryotes, however, the signals within small introns only function well in an intronic orientation. This restriction is revealed by the inhibition of splicing of small *S. pombe* and *D. melanogaster* introns when they are internally expanded (14, 31, 47, 51). Even very modest increases in size in a *D. melanogaster* intron can result in either inactivation or activation of a cryptic splice site within the expansion cassette.

The above restrictions suggested that small introns within vertebrate pre-mRNAs might be recognized via a mechanism that differs from that used in pre-mRNAs with a more standard exon-intron architecture. Indeed, an examination of the splicing phenotypes of human disease genes indicates that intron inclusion can be observed in genes containing a splice site mutation within a small human intron (32).

To examine the sequence requirements for splicing of small vertebrate introns, we collected a set of small human introns with lengths of less than 200 nucleotides to permit both a statistical look at small intron sequences and an informed choice of experimental system. Within this population of introns, we were surprised to find an appreciable number of introns that were G rich, specifically that contained multiple G triplets or quartets. Over one-third of the population had a noticeably high concentration of G triplets, suggesting that G triplets might play a common role in small intron recognition. To examine this possibility, we studied the human α-globin

* Corresponding author. Mailing address: Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030. Phone: (713) 798-4622. Fax: (713) 795-5487. E-mail: andrewm@bcm.tmc.edu.

gene. The second intron of α-globin is very G rich (42%) compared to its flanking exons (27 or 22% for exons 2 and 3, respectively). These intronic G residues exist primarily as GGG or GGGG sequences. Here, we demonstrate that the G triplets studded throughout the α-globin second intron are minimal intron-enhancing sequences that additively improve α-globin splicing and spliceosome complex formation. A short competitor RNA containing two G triplets specifically inhibited α-globin intron 2 but not intron 1 splicing, suggesting that G triplets are recognized by a saturable *trans*-acting factor, used only for the splicing of G triplet-containing introns. We suggest that G triplets act as an intronic equivalent of an exon enhancer in small introns whose recognition operates via intron definition.

More importantly, our experiments indicated that G triplets have a role in splice site selection as well as in splicing efficiency. G triplets positioned between duplicated 5′ or 3′ splice sites promoted usage of the outside site, suggesting that G triplets act to define intron borders and to delineate exonic from intronic sequences. Although our data specifically relate to small introns, statistical analyses of vertebrate exons have previously suggested a role for an exon-proximal, simple purine repeat containing G. Sequences immediately downstream of vertebrate 5′ splice sites are G rich, with G triplets representing the most common motif (12, 33, 34). In fact, a parameter for G triplets has been included in computer algorithms designed to predict 5′ splice site locations in raw genomic sequence data, and such inclusion enhances exon identification (40). These statistical observations coupled with our experimental results, as well as observations in several alternative exons (3, 8, 39), suggest that G triplets represent a common intron recognition element that is important for 5′ splice site recognition. We suggest that intron enhancers containing G triplets are used in many vertebrate genes to increase the complexity of splice site sequences so as to provide greater specificity of recognition of exon-intron borders.

MATERIALS AND METHODS

Statistical analysis of G content. The minimal frequency of G, C, U, and A triplets within the collected set of human introns with lengths of 200 nucleotides or less was determined. For this purpose, each triplet, quartet, or pentet was scored as a single triplet, and each sextet was counted as two triplets. Each calculated triplet number was divided by the length of the intron and multiplied by 1,000 to determine the triplet frequency per kilobase.

DNA constructions. To generate the wild-type α-globin intron 2 construct used in these studies, a PCR product containing exon 2, intron 2, and exon 3 of the human α-globin 2 gene was inserted into the pCDNA3 vector (Invitrogen) with *Bam*HI and *Eco*RI restriction sites introduced during PCR. G triplets and splice site mutations were introduced by PCR-mediated site-directed mutagenesis. To facilitate construction of the 5′ splice site duplication substrates, a *Sac*II-*Nhe*I linker was inserted into exon 2, five nucleotides upstream of the intron 2 5′ splice site. The 5′ splice site duplications were subsequently generated by inserting a *Sac*II-*Nhe*I-flanked PCR product which contained the last 5 nucleotides of exon 2 and the first 54 nucleotides of intron 2. To facilitate construction of the 3′ splice site duplication substrates, a *Sac*II-*Nhe*I linker was inserted into exon 3, six nucleotides downstream of the intron 2 3′ splice site. The 3′ splice site duplications were subsequently generated by inserting a *Sac*II-*Nhe*I-flanked PCR product extending from the G4 element in intron 2 to nucleotide +6 of exon 3.

Cell culture and transfection. HeLa cells were maintained in Dulbecco's modified Eagle's medium containing 10% bovine serum. Transfections were performed by lipofection with the Lipofectamine reagent (Gibco BRL). Total cell RNA was purified by the RNAzol B method (Biotecx).

Analysis of transcripts by RT-PCR. The pCDNA3 5′ and pCDNA3 3′ oligonucleotide primers used in reverse transcription (RT)-PCR analysis were complementary to pCDNA3 vector sequences which flank the inserted α-globin sequences. First-strand cDNA synthesis was done with a reaction mixture containing 1 μg of total RNA, 50 mM KCl, 10 mM Tris-Cl (pH 8.4), 2.5 mM $MgCl_2$, 200 μg of gelatin per ml, 200 μM (each) deoxynucleoside triphosphates, 5 U of avian myeloblastosis virus reverse transcriptase (Promega), 2.5 U of *Taq* DNA polymerase (Cetus), 20 U of RNasin (Promega), and 50 pmol of pCDNA3 3′ oligonucleotide primer (unlabeled). First-strand cDNA was synthesized for 30 min at 50°C, after which 50 pmol of the pCDNA3 5′ (^{32}P 5′ end labeled) was added. The cDNAs were subsequently amplified by 15 cycles of PCR. Each PCR cycle consisted of 94°C denaturation for 1 min, 55°C annealing for 2 min, and 72°C extension for 2 min. PCR products were fractionated on nondenaturing 5% polyacrylamide gels and visualized by autoradiography. Signals were quantified by phosphorimaging (Molecular Dynamics). Each reported splicing efficiency represents the average of at least four independent transfection experiments with their corresponding standard errors.

In vitro splicing and spliceosome complex assembly. RNA transcripts for in vitro splicing and spliceosome assembly reactions were synthesized by T7 RNA polymerase (Gibco-BRL) in the presence of 50 μCi of ^{32}P-UTP (DuPont). RNAs were gel purified, eluted, and precipitated in ethanol. Splicing reactions were performed in a volume of 50 μl containing 40% (vol/vol) HeLa cell nuclear extract, 2.0 mM ATP, 20 mM creatine phosphate, 3.0 mM $MgCl_2$, 0.6% polyethylene glycol, 0.15 mM dithiothreitol, and 5×10^4 cpm of ^{32}P-labeled RNA. Splicing reaction mixtures were incubated at 30°C, and 10-μl aliquots were taken at 0, 30, 60, and 90 min, extracted with phenol-chloroform, and ethanol precipitated. The RNAs were analyzed on 5% polyacrylamide gels containing 8.3 M urea. For the analysis of spliceosome complex assembly, aliquots were taken at 0, 5, and 15 min and resolved on native gels consisting of 4% polyacrylamide (80:1) and 1% agarose. For competition experiments, reaction mixtures were formulated as described above but contained 2.5×10^4 cpm of ^{32}P-labeled RNA and from 0 to 200 pmol of unlabeled competitor RNA oligonucleotide (Oligos Etc.) in a 25-μl reaction volume. Reaction mixtures were incubated at 30°C for 40 min, extracted with phenol-chloroform, ethanol precipitated, and analyzed by gel electrophoresis.

RESULTS

A class of small G-rich vertebrate introns. In lower eucaryotes, small introns are frequent. For example, 50% of the introns in *D. melanogaster* are less than 100 nucleotides in length, with an average length within the small intron subpopulation of 50 to 60 nucleotides (31). In fact, a display of *D. melanogaster* introns shows a biphasic distribution into a class of very small introns and a second class of introns with a size distribution similar to that seen in vertebrates (15). Vertebrates, however, do not have a high frequency of small introns, cannot effectively splice introns smaller than 70 nucleotides, and display no obvious bipartite distribution of intron sizes. The existence of intron inclusion phenotypes within human disease genes containing small introns, however, suggested that small human introns might represent a mechanistically separated set of vertebrate introns.

To be able to examine the relevance of small introns and intron definition in vertebrates, we collected a set of 693 human introns with lengths of less than 200 nucleotides. This collection had an average length of 128 nucleotides and an overall base composition of 29% G, 25% U, 27% C, and 19% A, slightly biased towards G and deficient in A. Although we had expected to find isolated G triplets in the vicinity of the 5′ splice sites of these introns because of previously reported statistical analyses of human 5′ splice sites (12, 33, 34, 40), we noticed that a number of the introns were studded with G triplets throughout their length. To better analyze the G content of these introns, the number of G, C, U, and A triplets per unit length in each intron of the collection was calculated (Materials and Methods) and is displayed in Table 1. G, U, and C triplets were roughly of equal abundance in the introns. When triplets per unit length were considered, however, only G triplets were clustered within individual introns.

As shown in Table 1 and the histograms in Fig. 1, a number of the introns in the set had a high frequency of G triplets. Of the 693 introns in the set, 136 (20%) had greater than three times the number of G triplets expected in the random sequence (assume 1 G triplet/64 bases). Poisson distribution predicts that only 2% of the introns should contain this many G triplets. In comparison, only eight introns (1%) had a similarly high density of A triplets (Table 1). Examples of introns with a high frequency of G triplets are shown in Table 2. There was no obvious similarity in function or tissue distribution among the genes containing introns with a high content of G triplets,

TABLE 1. Distribution of simple repeats in small human introns

Triplet	% of triplets	No. of triplets in data set[a]				Avg no. of NNN triplets/kb	No. of introns with triplet frequency >98% Poisson expectation
		NNN	NNNN	NNNNN	NNNNNN		
G	29.1	2,360	841	259	68	27.3	136
U	25.0	1,459	511	175	56	23.0	49
C	26.6	2,094	763	199	30	15.9	86
A	19.4	691	181	47	13	7.1	8

[a] Number of triplets in the data set of 693 small human introns. For this calculation, NNN, NNNN, and NNNNN were each counted as a single triplet.

nor were all of the introns in a gene with a G-rich small intron either small or G rich. This observation suggests that G triplet-containing introns are dispersed among human genes and might represent a common intron subclass.

G-rich intronic elements facilitate α-globin RNA processing in vivo. To see if G triplets have a function in splicing, we chose to experimentally alter G triplets in one such intron, the second intron of the human α-globin gene. As shown in Fig. 2A, this 129-nucleotide intron contains seven G triplets located from nucleotides +9 to +103 in the intron. The G triplets in the α-globin intron were mutated singly and in pairs in single-intron constructs designed to monitor in vivo or in vitro splicing. For in vivo analysis of splicing, a DNA fragment containing exon 2, intron 2, and exon 3 was inserted into the mammalian expression vector pCDNA3 (Fig. 2B). This construct (termed α-globin) was transiently expressed in HeLa cells, and the splicing pattern was determined by RT-PCR with oligonucleotide primers complementary to vector sequences flanking the insertion. The vector-derived Neo gene transcript was coamplified as an internal control for transfection efficiency and RNA quality. As shown in Fig. 3B (lane 1), spliced α-globin transcripts and Neo transcripts were readily detected in total RNA isolated from transfected HeLa cells.

Six of the seven G triplets in the α-globin intron are grouped into three pairs with a generic sequence of $GGG(N)_{2-4}GGG$, which we termed G elements G1, G2, and G4, as diagrammed in Fig. 2. To determine if these G elements are important for α-globin RNA splicing, multiple G nucleotides were mutated in each G element or in multiple G elements as diagrammed in Fig. 3A. All mutations were introduced into the pCDNA3 in vivo expression cassette described above. RT-PCR analysis of RNA prepared from HeLa cells transfected with these mutant substrates clearly indicated that each of the G1, G2, and G4 elements is important for maximal spliced α-globin RNA accumulation in vivo (Fig. 3B). Mutation of multiple G's in any of the G1, G2, or G4 elements caused unspliced α-globin RNA to accumulate in vivo. Mutation of a single element (Fig. 3B, lanes 2, 3, and 4) had a modest effect, reducing the percentage of accumulated spliced RNAs to about 89% of wild-type levels (average of G1, G2, and G4).

When multiple elements were mutated, inhibition effects on α-globin-spliced RNA accumulation were additive (discussed below). Cells transfected with α-globin-containing double mutant G elements [G(1+2), G(1+4), or G(2+4)] accumulated only 68% (average) spliced α-globin transcripts compared to wild-type cells (Fig. 3B, lanes 5, 6, and 7). Introduction of a triple mutant caused an even greater reduction in spliced RNA; in this case splicing was reduced to 45% of the level in the wild type (Fig. 3B, lane 8). The average splicing efficiencies from four independent transfections with this panel of mutants are displayed graphically in Fig. 3C.

G-rich intronic elements facilitate α-globin RNA splicing in vitro. The experiments described above indicated that the G1, G2, and G4 G elements were important for efficient α-globin-spliced RNA accumulation in vivo. Because these mutations promote the accumulation of unspliced α-globin RNA, it was likely that the G triplets are important for α-globin splicing per se. Effects on RNA transport, stability, or polyadenylation could, however, produce similar in vivo effects. To differentiate between these possibilities, wild-type and mutant α-globin substrates were spliced in vitro in HeLa nuclear extract. In all cases, substrates containing mutated G triplets were spliced with lower efficiencies than were wild-type substrates. Figure 4A compares in vitro splicing of the wild-type α-globin substrate with that of the G(1+2) mutant α-globin substrate. Quantification of the amount of splicing intermediates and products indicated that the wild-type substrate was spliced with twice the efficiency of the mutant substrate. These in vitro experiments recapitulated the in vivo results and confirm that the G elements are important for efficient α-globin splicing per se.

G-rich intronic elements promote spliceosome assembly in vitro. To determine if G element mutations in the α-globin intron affect spliceosome assembly, in vitro splicing reactions were analyzed by native gel electrophoresis. Figure 4B compares spliceosome assembly with wild-type and G(1+2) mutant

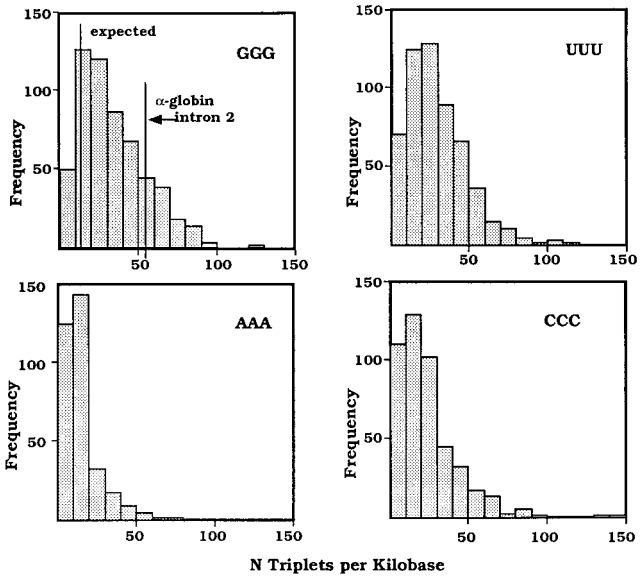

FIG. 1. G-rich small introns. The number of G, U, C, and A triplets in each intron in a data set of 693 human introns with lengths of ≤200 nucleotides was calculated (Materials and Methods). For each intron, the number of N triplets per unit length was calculated. The distribution of these values for each nucleotide is shown. The G plot also indicates the position of the second human α-globin intron in the distribution.

TABLE 2. Small human introns containing G triplets

Intron	Sequence[a]
RNA polymerase II 14.5-kDa subunit, intron 3	gtgagtggcGGGccctgagctGGGGGGcGGGGGtgttggctctggaggctGGGtctgagcgtaattttgcacccccgc gtccctgcag
Ribosomal protein S7, intron 1	gtgagaGGGtttccctGGGGtctGGGGtGGGGGGaggcgccccgcccGGGGGGgaggcggccgcgcgtgtgtt GGGcccGGGGtgctcggacgcgcgctcaGGGtcggtcctgctgttcgttgcttcttag
Na,K-ATPase subunit α-2, intron 9	gtgattGGGtgctccagaGGGGGtggataggattagaggaggctgaGGGcagtggcgtggtGGGGtgagtggttgaga taaaggctctaaaGGGagccacgctcctggttccccctcatttcctcccag
Gastric (H^+ K^+) ATPase	gtgagtGGGGcccaGGGGtccGGGGtgcagtGGGGacaGGGGttgcagtggaatctgagttctccctccctcctcatc ctgtccattcccag
Mytonic dystrophy kinase, intron 11	gtcGGGatcGGGGccGGGGccGGGGccGGGatgcGGGccggtggcaacccttggcatcccctctcgtccggccc ggacggactcaccgtccttacctccccacag
MHC[b] class 1 HLA-AW24, intron 7	gtgagagcttggaGGGcctgatgtgtgttGGGtgttGGGcggaaccgtggacacagctgtgctatGGGGtttctttgcattgg atgtattgagcatgcgatGGGctgtttaaagtgtgacccctcactgtgacagatatgaagttgttcatgaatttttttctatag
HUMMHBA17W′, intron 1	gtgagtgcGGGGtcGGGaGGGaaatggcctctgtGGGGaggagcgaGGGGaccgcaggcGGGGGcgcagga cctgaggagccgcgccGGGaggaGGGtcGGGcGGGtctcagcccctcctcgcccccag
MHC class 1 CD8 α-chain, intron 1	gtgagtttgagacgcccGGGaaggcaGGGGatGGGGcGGGcGGGGGGcGGGGGGctcGGGactgctt gctcagcctcgctgcttgcctcctgcag
Phosphoglycerate mutase, intron 1	gtGGGctgcctttgctGGGaaggcctctGGGaagctgcagagtGGGGagtcGGGtGGGGGcccactggcttGG GaGGGaaagcagcgtgcctgtgtcccccag
Teratocarcinoma-derived growth factor 1, intron 4	gtaagcaattcagaGGGGcGGGGagccgtggagaggagagagaaaGGGaagtggaaatttcagacccaagctatcgcag cttacctgttcattctcag
VH4-immunoglobulin heavy chain variable region, intron 3	gtgagtatctcaGGGatccagacatGGGGatatGGGaggtgcctctgatcccaGGGctcactgtGGGtctctctgttca cag

[a] G triplets are capitalized.
[b] MHC, major histocompatibility complex.

α-globin substrates. This analysis identified three major complexes when the wild-type α-globin substrate was used (lanes 1 to 3). The H complex represents specific and nonspecific RNA-protein interactions which form immediately upon addition of substrate RNA to nuclear extract. The A complex, or prespliceosome, is an ATP-dependent complex which is defined by the binding of the U2 small nuclear ribonucleoprotein particle to the forming spliceosome. The B complex represents the mature spliceosome. As shown in Fig. 4B, (lanes 4 to 6), wild-type and mutant RNAs formed visually identical complexes, indicating that no obvious changes in spliceosome assembly intermediates were induced by the mutations. Importantly, however, complex formation was less efficient when mutant substrates were used. Formation of both A and B complexes was reproducibly reduced twofold in the mutant, indicating that recognition of the G elements is important very early in spliceosome assembly.

α-Globin splice sites are recognized by intron definition. The observation that elements through a small intron enhance splicing and assembly of the intron suggested concerted recognition of the entire intron via an intron definition pathway. To test this hypothesis, we analyzed the in vivo splicing phenotype of a mutant α-globin gene whose splicing pathway should be indicative of exon or intron definition (Fig. 5). If intron definition is operating in the α-globin gene, disruption of the 5′ splice site in intron 2 should cause intron 2 retention without an effect on intron 1 removal. In contrast, exon definition predicts exon skipping for a 5′ splice site mutation (2, 46). As shown in Fig. 5, when the 5′ splice site of the second α-globin intron was disrupted, transcripts accumulated retaining intron 2 but lacking intron 1. No exon skipping was observed. These results support the hypothesis that the splice junctions in α-globin pre-mRNA are recognized by intron definition.

Short G-rich competitor RNAs inhibit α-globin splicing in vitro. G-rich elements could affect splicing efficiency via either an effect on precursor RNA structure or as sites for the binding of *trans*-acting factors. In vitro competition studies were used to help distinguish these alternatives. In vitro splicing reactions were performed in the presence of increasing amounts of RNA oligonucleotides corresponding to either a single wild-type or mutant G1 element. As shown in Fig. 6, an RNA containing the wild-type element (C GGG CC GGG A) inhibited α-globin splicing, while the mutant competitor (C UGA CC AGC A) had no effect. In contrast, parallel splicing reactions with an

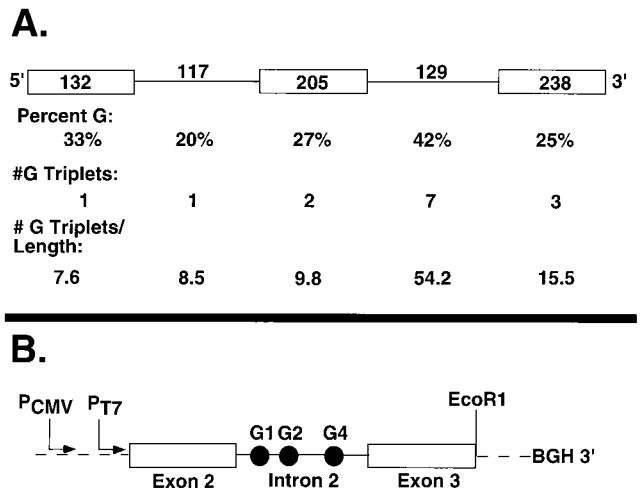

FIG. 2. Architecture of the human α-globin 2 gene. (A) Open boxes represent exons, and lines represent introns. Intron lengths are indicated above each intron, and exon lengths are indicated within the boxes. The percent G of each segment, the number of G triplets found in each segment, and the number of G triplets per unit length (expressed as number of G triplets per kilobase) are shown below the diagram. (B) Structure of the intron 2 splicing substrates analyzed in this work. These constructs contain exons 2 and 3 and intron 2 of the human α-globin 2 gene. In vivo analysis utilized the cytomegalovirus promoter (P_{CMV}) and bovine growth hormone (BGH) gene 3′ end forming signals located in the pCDNA3 vector (Invitrogen). RNA substrates for in vitro analysis were synthesized by using the T7 promoter (P_{T7}) and were run off at the illustrated *Eco*RI restriction site.

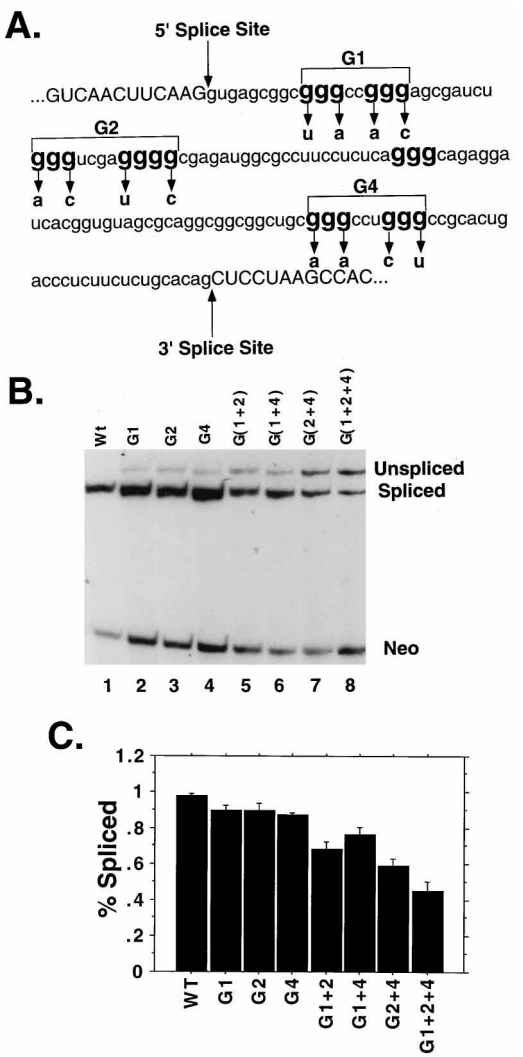

FIG. 3. G element mutations reduce intron 2 splicing in vivo. (A) Complete nucleotide sequence of the α-globin intron used in these studies (lowercase). The sequences of at least three contiguous G residues are highlighted with large, boldface type, and the arbitrarily defined G1, G2, and G4 elements are bracketed. Mutations introduced into these elements and analyzed in panel B are shown below the wild-type sequence. Nucleotides shown in uppercase represent the exonic sequences which border the intron. (B) RT-PCR (15 cycles) analysis of splicing products detected from wild-type (Wt) and mutant α-globin intron 2 constructs following transient transfection of HeLa cells. The positions of PCR products corresponding to spliced and unspliced transcripts and the vector-encoded Neo transcripts used as a transfection control are shown to the right of the gel. The G element mutated in each case is shown above the lane. Wild-type and mutant G element sequences are shown in panel A. (C) Average splicing efficiency (spliced/spliced + unspliced) for each construct tabulated from four independent experiments. Error bars represent the standard error of the mean.

α-globin intron 1 substrate, which has no G triplets, were unaffected by the G1 oligonucleotide. In addition, neither substrate was affected by an oligonucleotide containing U3 small nucleolar RNA sequences (data not shown). Thus, the ability of the wild-type G1 competitor RNA to inhibit α-globin splicing is both competitor and substrate specific. These experiments suggest that the G elements in α-globin are specifically recognized by a saturable *trans*-acting factor which is present in extracts prepared from HeLa cell nuclei.

G-rich intronic elements modulate 5′ intron border definition. The preceding data indicate that G elements stimulate splicing in vivo and in vitro. Most introns do not contain G triplets throughout the intron; instead when such elements are present, they are clustered in the first 50 nucleotides of the intron. Such placement suggests that in addition to boosting splicing efficiency, G elements could function to identify exon-intron borders and distinguish appropriate 5′ splice sites from cryptic sites.

A prediction of this hypothesis is that 5′ splice sites buried within G-rich intronic elements should not be selected for splicing when a usable 5′ splice site exists upstream of the G-rich sequences. To test this prediction, 5′ splice site duplication constructs were prepared. These constructs contained two identical α-globin 5′ splice sites separated by 65 nucleotides of α-globin sequence containing either wild-type or mutant G1 and G2 elements (diagrammed in Fig. 7A [see Materials and Methods for a detailed description of the construct]). If the G elements are important for 5′ intron border definition, the upstream 5′ splice site (distal) should be favored when the duplicated G triplets are wild type, and the proximal site should be activated when these G triplets are mutated. Consistent with this prediction, the distal 5′ splice site was used almost exclusively when the wild-type duplication construct was expressed in transfected HeLa cells (Fig. 7B, lane 1). When the duplicated G1 or G2 element between the 5′ splice sites was mutated, the proximal 5′ splice site was activated, and about 50% of the accumulated transcripts were spliced at the proximal site (Fig. 7B, lanes 2 and 3). When both of the duplicated G1 and G2 elements were mutated, the proximal site became dominant and was used in about 75% of the spliced transcripts (Fig. 7B, lane 4). These results support a

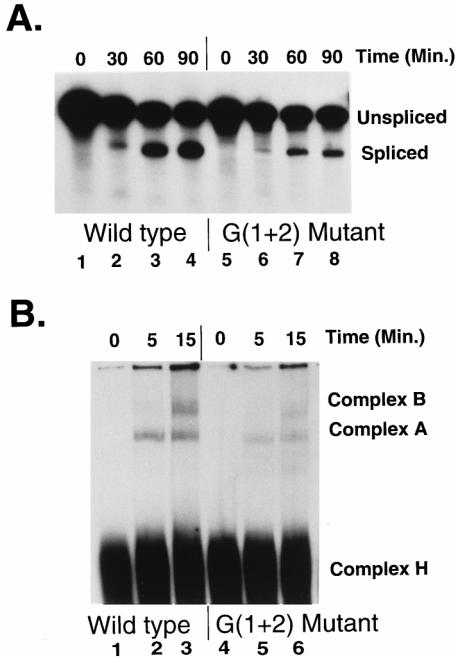

FIG. 4. G element mutations reduce in vitro intron 2 splicing and spliceosome assembly. (A) In vitro splicing. Uniformly labeled wild-type and mutant RNA precursor RNAs containing the α-globin second intron (Fig. 2) were incubated with HeLa nuclear extract under splicing conditions for the times indicated above each lane, and the reaction products were resolved on a denaturing 5% polyacrylamide gel. The positions of unspliced and spliced precursor RNA are shown to the right of the gel. (B) In vitro spliceosome assembly. In vitro splicing reactions of wild-type and mutant precursor RNAs were analyzed by native gel electrophoresis. The locations of the expected A, B, and H complexes are shown to the right of the gel. The G(1+2) mutant is diagrammed in Fig. 3A).

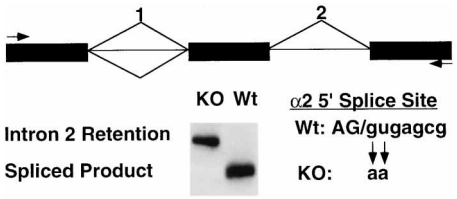

FIG. 5. Intron definition in the α-globin gene. (Top) Diagram of the full-length α-globin construct. The sequences of the wild-type (Wt) and knockout (KO) intron 2 5′ splice sites are shown below the diagram. (Bottom) In vivo splicing RT-PCR (15 cycles) analysis of the splicing products detected from the wild-type and 5′ splice site knockout three exon α-globin constructs. The splice product in the wild-type lane represents fully processed transcripts lacking introns 1 and 2. The intron 2 retention product in the KO lane represents transcripts lacking intron 1, but retaining intron 2. These splicing pathways are illustrated above (Wt) and below (KO) the construct diagram.

strong role for G triplets in α-globin 5′ intron border definition.

To determine if the proximal 5′ splice site represents a usable site in the wild-type duplication substrate, the distal site was weakened by point mutation. Mutation of either the +1 G or +5 G of the distal site inactivated usage of the distal site and resulted in efficient use of the proximal site (Fig. 7B, lanes 5 and 6). Thus, in the absence of a functional upstream site, the proximal site is a usable splice site, even when it is buried within G-rich sequences. These results suggest that G triplets function by promoting usage of the upstream site rather than by blocking usage of the downstream site.

The ability of G triplets to promote distal 5′ splice site usage in α-globin duplication substrates was also tested in vitro in HeLa nuclear extract. As shown in Fig. 7C, the in vitro results effectively recapitulated the in vivo results. When the wild-type duplication substrate was used, the distal site was preferred; when the G(1+2) substrate was used, the proximal site was preferred.

G-rich intronic elements modulate 3′ intron border definition. The experiments described above demonstrated that the G triplets found in α-globin are integral to accurate 5′ splice site definition in this intron. To determine if these sequences also affect 3′ splice site definition, α-globin substrates containing duplicated 3′ splice sites were prepared. These constructs were generated by inserting a 64-nucleotide DNA fragment containing the G4 element and α-globin 3′ splice site into restriction sites engineered in the second exon of the standard α-globin construct as diagrammed in Fig. 8A. This construct effectively duplicated the G4 element, the 3′ splice site (including putative branch point sequence), and the first six nucleotides of the terminal exon. In the construct designated Wt 3′ dup, all of the duplicated sequences were wild type. In the construct designated G4 3′ dup, the downstream G4 element was mutated as shown in Fig. 3A.

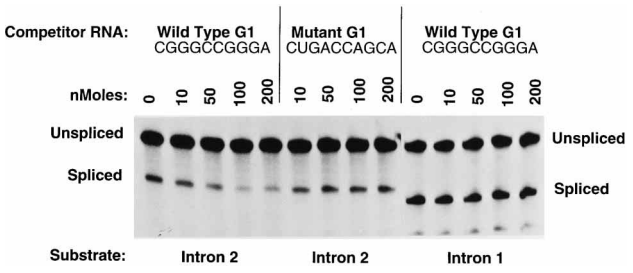

FIG. 6. Short G-rich competitor RNA inhibits α-globin intron 2 splicing in vitro. In vitro splicing reactions were performed in the presence of an RNA oligonucleotide representing either the wild-type or mutant G1 element. The amount of oligonucleotide added (nanomoles per 25 μl of splicing reaction mixture) is indicated above each lane. The intron 1 substrate used as a control contained α-globin intron 1 (which is not G rich) flanked by exons 1 and 2 of the α-globin gene.

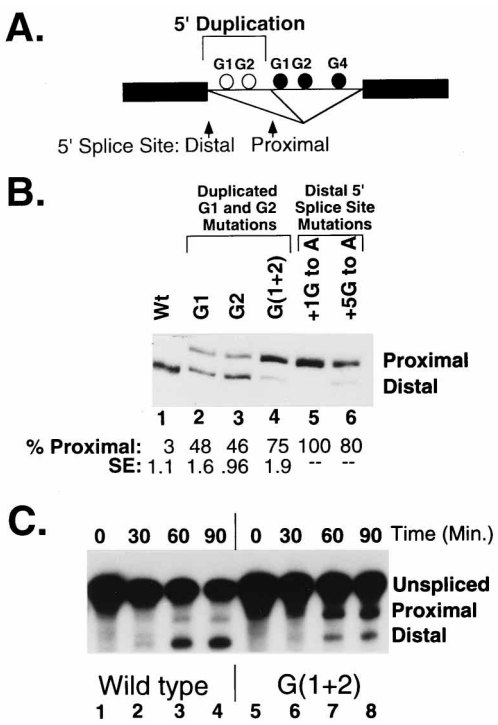

FIG. 7. G elements modulate 5′ splice site selection in vivo and in vitro. (A) In vivo and in vitro splicing constructs containing a single intron with duplicated 5′ splice sites. As constructed, the 5′ splice site duplication substrates contained two identical 5′ splice sites separated by 58 nucleotides containing either wild-type (Wt) or mutant G1 and G2 elements. The G1 and G2 mutations were identical to those indicated in Fig. 3A. The wild-type and mutated G1 and G2 elements are represented by solid and open circles, respectively. The sequences downstream of the proximal 5′ splice site were wild type in each construct. Potential splicing pathways with either the distal or proximal 5′ splice sites are illustrated in panel A. Two other duplication constructs were also used in which the distal 5′ splice site was mutated at position +1 (mutant +1G to A) or +5 (mutant +5G to A). In the latter constructs, the G elements were wild type. (B) In vivo splicing. Results of RT-PCR (15 cycles) analysis of splicing products produced in transiently transfected HeLa cells are shown. The positions of PCR products corresponding to RNAs spliced with the distal or proximal 5′ splice site are indicated to the right of the gel. The mutants used are indicated above the lanes. The percentage of products resulting from usage of the proximal site is indicated below each lane. Each percentage is the average of four transfections, with standard errors (SE) as indicated. (C) In vitro splicing. Uniformly labeled RNA transcripts containing duplicated 5′ splice sites separated by wild-type or mutant G1 and G2 elements were incubated with HeLa nuclear extract under splicing conditions for the times indicated above each lane, and the reaction products were resolved on a denaturing 5% polyacrylamide gel. The positions of unspliced transcripts and transcripts spliced utilizing the distal or proximal 5′ splice site are shown to the right of the gel. The identity of these splicing products was confirmed by RT-PCR and DNA sequence analysis.

These 3′ splice site duplication constructs were expressed transiently in HeLa cells, and their splice site selection patterns were determined by RT-PCR. As shown in Fig. 8B, both 3′ splice sites were active in each construct. The balance of usage was altered when the downstream duplicated G4 element was mutated, however. The internal site was used with significantly greater efficiency when the G4 element had been mutated (32% versus 13% of spliced RNAs). This result supports a role for G elements in 3′ splice site definition as well as 5′ splice site definition.

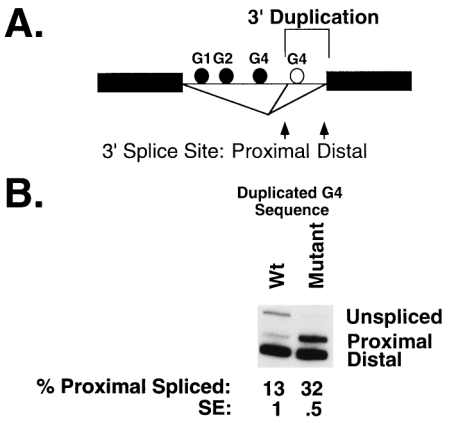

FIG. 8. G elements modulate 3′ splice site selection. (A) Constructs containing duplicated 3′ splice sites and G4 elements. These constructs were generated by inserting a DNA fragment that contained the G4 element and the 3′ splice site from α-globin intron 2 into an engineered restriction site in exon 2. In the wild-type (Wt) construct, all of the duplicated sequences were wild type. In the mutant construct, the downstream G4 element (open circle) had been mutated, and all other sequences were wild type. (B) In vivo splicing. RT-PCR (15 cycles) analysis of splicing products detected from 3′ splice site duplication constructs following transient transfection of HeLa cells was used to determine the effect of G elements on 3′ splice site selection. The positions of PCR products corresponding to transcripts spliced with the distal or proximal 3′ splice site are indicated to the right of the gel.

GGG is the minimal G element. The previous data indicated that a G element of generic sequence $GGG(N)_{2-4}GGG$ facilitates splicing and splice site recognition. All of the mutations shown in previous figures mutated at least four G nucleotides per element. A battery of mutations was created to ascertain the requirement for individual nucleotides within the G element. The assay described in the legend to Fig. 7 testing the ability of G elements to influence 5′ splice site selection was used to provide a sensitive indicator of element function. For these experiments, the G1 element (GGG CC GGG) was subdivided into three domains—the G1-1 triplet (GGG), the pyrimidine spacer (CC), and the G1-2 triplet (GGG) as illustrated in Fig. 9A and B. Each domain was independently targeted for mutation. As detailed below, this analysis demonstrated that the individual G triplets in the G1 element are critical for its function, but the spacer between the G triplets is not important. Deletion of the spacer or mutation of it from CC to AA, GG, or UU had no effect on 5′ splice site selection. In these spacer mutants, the distal site was used exclusively (Fig. 9B). This result suggests that the G triplets within the G elements are the functional portion of the element.

In a second set of mutants, double-point mutations were introduced into the first G triplet of the G1 element. Mutation of the first G triplet (denoted G1-1) in the G element from GGG to UGA (mutant G1-1), AGA (mutant G1-1A), CGC (mutant G1-1C), or UGU (mutant G1-1U) activated the proximal splice site (Fig. 9C, top gel), indicating that the first G triplet in the G element is important for 5′ splice site selection. The G1-1 G triplet was further dissected by mutation of single G residues. Mutation of the first G triplet in the G element from GGG to UGG (mutant G1-1-1U) or GGU (mutant G1-1-3U) activated the proximal site to the same degree as the double-G mutations described above (Fig. 9C, bottom gel). Collectively, these experiments indicate that the sequence GGG represents the minimal functional unit of a G element for splice site selection.

G triplets compensate for a weak 3′ pyrimidine tract. Why does the α-globin intron require auxiliary elements (G triplets) for efficient splicing? Both the 5′ splice site (AAG/GUGAGCG) and the putative branch point sequence (CACUGAC) in this intron are reasonable matches to vertebrate consensus sequences. The 3′ pyrimidine tract, however, is C rich and, as a result, may be suboptimal. To test this hypothesis, the U content of the natural α-globin pyrimidine tract sequence (CCCUCUUCUC UGC) was increased. An improved pyrimidine tract (CCCUU UUUUUUUC) was introduced into both the wild-type α-globin construct and a construct in which all G elements were mutated. As shown in Fig. 10, the improved pyrimidine tract effectively alleviated the requirement for G triplets for maximal in vivo splicing. The G(1+2+4) mutant construct containing the improved pyrimidine tract (Fig. 10, lane 4) was spliced as efficiently as the wild-type α-globin construct (Fig. 10, lane 1). The wild-type construct with the improved pyrimidine tract

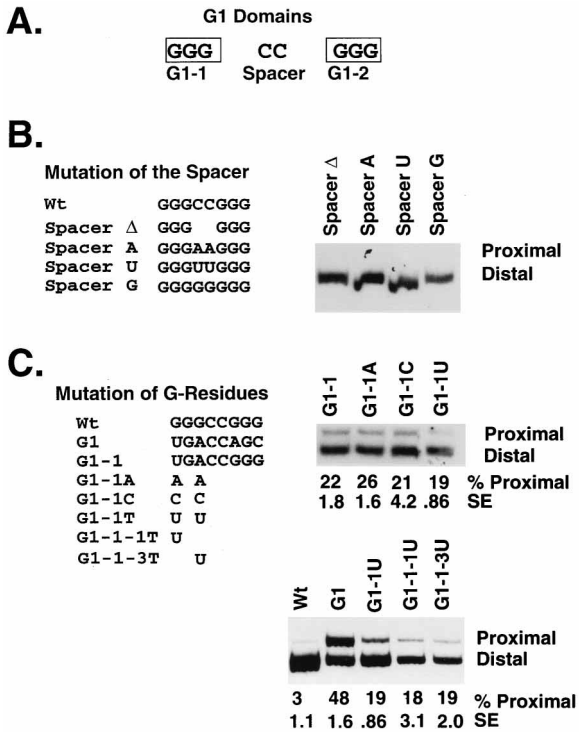

FIG. 9. GGG represents the minimal element within a G element. (A) To further delineate essential sequences in the G1 element, it was divided into three domains designated G1-1, spacer, and G1-2 as diagrammed. Mutations were independently introduced into the first G triplet and spacer of the G1 element present in the 5′ splice site duplication construct diagrammed in Fig. 7A. (B) Mutation of the spacer has no effect on 5′ splice site selection. Introduced mutations are shown on the left. Results of RT-PCR (15 cycles) analysis of splicing products following transient transfection of HeLa cells are shown to the right. The positions of PCR products corresponding to transcripts spliced with the distal or proximal 5′ splice site are indicated to the right of the gel. In each case, only the upstream G1-1 element had been mutated, and all other sequences were wild type (Wt). The percentage of products resulting from usage of the proximal site is indicated below each lane. Each percentage is the average of four transfections, with the standard error (SE) as indicated. (C) Mutation of the G1-1 triplet activated splicing of the proximal 5′ splice site. Introduced mutations are shown to the left. RT-PCR analysis of RNAs produced following transient transfection of HeLa cells as in panel B above is shown to the right. In each case, only the upstream G1 element had been mutated, and all other sequences were wild type. The percentage of products resulting from usage of the proximal site is indicated below each lane. Each percentage indicated is the average of four transfections, with the standard error as indicated.

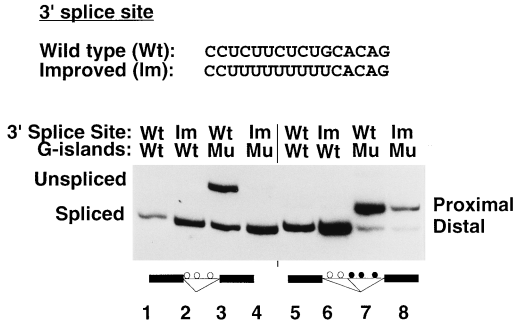

FIG. 10. Improving the α-globin 3′ splice site suppresses the requirement for G elements for maximal splicing but does not affect the ability of G elements to regulate 5′ splice site selection. (Top) Sequences of the wild-type (Wt) and improved (Im) α-globin intron 2 pyrimidine tract and 3′ splice site. (Bottom) In vivo splicing. Results of RT-PCR (15 cycles) analysis of splicing products from wild-type and mutant α-globin intron 2 constructs following transient transfection of HeLa cells are shown. Lanes 1 to 4 show substrates with single 5′ and 3′ splice sites (as in Fig. 3) with either wild-type or mutant (Mu) G elements [the G(1+2+4) mutant diagrammed in Fig. 3B]. Lanes 5 to 8 show 5′ splice site duplication substrates (as in Fig. 7) with either wild-type or mutant upstream G1 and G2 elements [the G(1+2) mutant diagrammed in Fig. 3A]. The pyrimidine tract present in each construct is identified above each lane. The positions of PCR products corresponding to spliced and unspliced simple substrates are shown to the left of the gel. Positions of PCR products corresponding to splicing with the distal or proximal 5′ splice sites are shown to the right of the gel.

also spliced efficiently, indicating that a strong U-rich pyrimidine tract is not incompatible with G elements.

The ability of G elements to direct 5′ splice site choice in the presence of an improved pyrimidine tract was also assayed. For this purpose, the improved pyrimidine tract was also introduced into the wild-type and G(1+2) mutant 5′ splice site duplication constructs described in the legend to Fig. 7. As shown in Fig. 10 (lanes 5 to 8), the 5′ splice site selection patterns in these constructs were unaffected by the improved pyrimidine tract. Thus, the U-rich pyrimidine tract did not affect the ability of G elements to modulate 5′ splice site selection.

DISCUSSION

G triplets enhance splicing efficiency and splice site selection. Statistical analyses have previously identified an abundance of simple G_3 repeats near the 5′ splice sites of primate introns (12, 33, 34, 40). Here we have analyzed a constitutively spliced human intron containing seven G triplets within 100 nucleotides of the 5′ splice site and have experimentally determined that each G triplet functions as a positive intron recognition element in human cells and in HeLa nuclear extract. In addition to their role in enhancing splicing efficiency, the G triplets played a role in determination of exon-intron borders. Placement of G triplets between competing 5′ splice sites caused dominant splicing by pathways that placed the G triplets in an intronic rather than exonic location. The ability of G triplets to enhance both activity and exon-intron border definition as well as the statistically meaningful occurrence of such sequences immediately downstream of 5′ splice sites suggests that G triplets should be considered as part of the primate 5′ splice site in constitutively spliced introns. The presence of an additional recognition sequence as part of 5′ splice sites may explain why vertebrate 5′ splice sites are less well conserved than might have been predicted were they working alone to define 5′ splice sites.

Simple sequence elements in small introns. This study was initiated to investigate if small vertebrate introns utilized mechanisms that differ from those observed in more standard exon definition pre-mRNAs. We began the analysis by compiling a collection of small vertebrate introns for analysis of sequence composition and element function. While we expected to find G triplets in the vicinity of the 5′ splice sites in this collection given previous statistical observations about vertebrate 5′ splice sites, we were startled by the number of introns in the collection that were both G rich and studded with G triplets—38% of the introns in the collection had a noticeably high frequency of G triplets. In contrast, A or C triplets were represented at high frequency in only 10 to 24% of the introns. The introns containing a high G triplet occurrence were not characterized by an obvious tissue distribution or gene coding function. Instead, the represented introns came from a wide variety of genes. The abundance of G triplets in our data set suggests that G triplets are common elements in small human introns.

Why would a small intron require an accessory sequence for either efficiency or exon-intron border definition? The splice sites within the α-globin intron are not noticeably weak, with the exception that the pyrimidine tract lacks an uninterrupted stretch of uridines. Therefore, it seems puzzling that an additional splice site consensus sequence would be necessary. We observed that the G triplets were only necessary for maximal splicing efficiency if the pyrimidine tract was wild type; improvement of the tract to include contiguous uridines removed the necessity for the G triplets (Fig. 10). This observation suggests that G triplets are an alternative to the presence of an optimal 3′ pyrimidine tract. In fact, when we analyzed our small intron data set, we noticed that the introns in the set with a high G triplet content had pyrimidine tracts that were only 40% uridine, whereas the introns with a lower G content had a 60% uridine content. Thus, the effect of G triplets to enhance splicing of introns with a weak pyrimidine tract may be similar to the ability of enhancer sequences that bind SR proteins to bypass requirements for either U1 snRNPs or U2AF and represent alternative mechanisms for splice site recognition (10, 25, 49).

Improvement of the pyrimidine tract had no effect on the ability of G triplets to regulate 5′ splice site choice. This observation suggests that G triplets play a role in aiding the discrimination between exon and intron sequences, irrespective of their ability to enhance splicing efficiency. Interestingly, G triplets affected both 5′ and 3′ splice site usage. Thus, a centrally located intron element affected both ends of the intron. This property is similar to the ability of exon elements to affect both ends of an exon during exon definition (50, 52). We suggest that the second intron of α-globin is recognized by intron definition utilizing both ends of the intron and a helper G triplet sequence within the intron. If so, the prediction is that α-globin should not demonstrate classical exon definition properties. One prominent exon definition property is the inducement of exon skipping via the introduction of a 5′ splice site mutation (2, 46). When the 5′ splice site of intron 2 of α-globin was mutated, exon skipping was not observed in vivo; instead an RNA product still containing intron 2 but lacking intron 1 accumulated. This result mirrors similar experiments with small intron-defined *Drosophila* and plant introns (29, 47). Therefore, α-globin splicing has the hallmarks of intron definition in which the splice sites that are paired during early spliceosome recognition are an intronic pair and recognition is enhanced by an intron-located enhancer. This interpretation is supported by the observation that mutation of G triplets in α-globin caused a depression in early spliceosome formation.

G triplets during exon definition. Interestingly, splice site-proximal intronic G triplets are also abundant in vertebrate

genes that contain large introns in which exon definition is the predominant mechanism for pairing splice sites (2, 12, 33, 34, 36, 37, 40). In addition, a G triplet-containing sequence has been implicated in alternative exon recognition in the β-tropomyosin gene (39). Perhaps the role of these sequences is to increase the information content of authentic splice sites in order to distinguish them from cryptic sites during exon definition or alternative exon recognition. Indeed splice site recognition in complex vertebrate genes is a daunting task and probably depends on a delicate balance between exonic elements, intronic elements, splice site border sequences, and the relative concentrations of numerous *trans*-acting factors.

G-binding factors. What factors recognize G elements? Several nuclear proteins within the heterogeneous nuclear RNP (hnRNP) family are known to bind preferentially to G-rich sequences, including hnRNP A1 (5), hnRNP F (26), hnRNP H (26), and p54^{nrb} (1). Binding of hnRNP A1 to the sequence UAGAGU, located in the intron downstream of an alternative exon in the hnRNP A1 gene, has recently been shown to correlate with the ability of the sequence to facilitate exon recognition (8). The identified sequence is similar to UAGGGU, a sequence determined to be an optimal hnRNP A1 binding site by iterative selection (5). hnRNPs F and H tightly bind to poly(G) and have been implicated as factors that bind to a neural-specific intron enhancer containing G-rich sequences along with several other proteins (30). Iterative selection has indicated that p54^{nrb} has a preferred binding site of AGGGA (1). All of these proteins are broadly distributed and could potentially participate in the recognition of a common G-rich intron element. The α-globin intron can be UV cross-linked to a number of proteins in the 40- to 60-kDa molecular mass range that characterizes all of these previously identified G-binding proteins. It will be interesting to determine which if any of these factors participate in G element-mediated splicing and splice site specificity.

G-rich elements act additively to both enhance splicing and facilitate exon-intron borders. Seven G triplets are located within the second intron of the human α-globin gene within the 100 nucleotides between the 5′ splice site and the branch point. Analysis of the contribution of individual or multiple G triplets indicated that each G triplet is an additive recognition unit (Fig. 11). Figure 11A shows the relationship between the number of G triplets and the overall α-globin splicing efficiency. From this plot, it is apparent that the effects of mutating multiple G triplets in this intron were additive. Thus, the factors which bind these triplets apparently do not do so in a cooperative manner. The only hint of cooperativity derives from the observation that the G(1+4) mutant, in which the mutated elements are not contiguous, had less of an effect on splicing than did the other two-element mutants [G(1+2) and G(2+4)], in which the mutated triplets were contiguous. While this difference is statistically significant, its biological significance is unknown.

The G triplets also acted additively to define exon-intron borders. Figure 11B indicates that mutation of individual G triplets affected the choice of flanking splice sites additively in a fashion indistinguishable from the effects of individual G triplets on splicing efficiency. Therefore, the α-globin intron is defined by the additive contributions of multiple low-impact interactions that play a role in both overall efficiency and in determining exon-intron borders. Interestingly, when the distal 5′ splice site was mutated in substrates containing all wild-type G triplets, the proximal site was efficiently used for splicing. Thus, a wild-type 5′ splice site positioned downstream of G triplets represents a usable site. This suggests a mechanism in which factors bound to G triplets activate the most upstream functional 5′ splice site as opposed to a mechanism whereby the proximal site is repressed or physically blocked. We cannot yet rule out a model in which the upstream G1 and G2 elements collaborate with the distal 5′ splice site to repress the proximal site. The positive-acting model, however, is also consistent with the ability of these elements to promote efficient splicing in unduplicated substrates (Fig. 3).

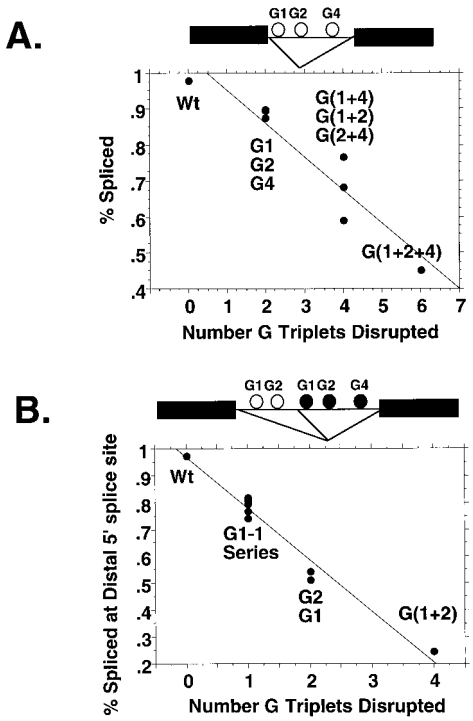

FIG. 11. G triplets function additively to activate the α-globin intron and regulate 5′ splice site selection. (A) Splicing efficiency. Average in vivo splicing efficiencies for the wild-type (Wt) and mutant α-globin intron 2 constructs shown in Fig. 3 were plotted on the *y* axis, and the number of GGG sequences disrupted in each mutant was plotted on the *x* axis. The construct represented by each point is indicated on the plot. (B) 5′ splice site selection. The percentage of in vitro splicing occurring by utilization of the distal 5′ splice site for the wild-type and mutant 5′ splice site duplication constructs shown in Fig. 7 and 9 was plotted on the *y* axis, and the number of GGG sequences disrupted in each mutant was plotted on the *x* axis. The construct represented by each point is indicated on the plot.

Simple intron repeats. The G-rich sequences described here are reminiscent of AU-rich intronic elements described in other systems. It is well established that AU-rich intronic sequences are required for efficient intron recognition in plant nuclei (13). Such sequences are essential for efficient splicing (13) and affect 5′ and 3′ splice site selection in manners indistinguishable from those of the G triplets described here (22, 27, 29). AU-rich elements have also been shown to affect splice site selection in *Drosophila* (28) and *Caenorhabditis elegans*. Interestingly, in *C. elegans*, AU-rich sequences are important for outron recognition during *trans*-splicing (9). Perhaps the G triplets that we have identified in this study are vertebrate counterparts of such AU-rich elements and perform evolutionarily conserved roles in small intron definition.

ACKNOWLEDGMENTS

We thank David Searls for providing the data set of small human introns and Rebecca Sierra for mutant construction and analysis.

This work was supported by grant GM38526 to S.B.

REFERENCES

1. **Basu, A., B. Dong, A. R. Krainer, and C. Howe.** 1997. The intercisternal A-particle proximal enhancer-binding protein activates transcription and is identical to the RNA- and DNA-binding protein p54^{nrb}/NonO. Mol. Cell. Biol. **17:**677–686.
2. **Berget, S. M.** 1995. Exon recognition in vertebrate splicing. J. Biol. Chem. **270:**2411–2414.
3. **Black, D. L.** 1992. Activation of c-src neuron-specific splicing by an unusual RNA element in vivo and in vitro. Cell **69:**795–807.
4. **Black, D. L.** 1995. Finding splice sites within a wilderness of RNA. RNA **1:**763–771.
5. **Burd, C. G., and G. Dreyfuss.** 1994. RNA binding of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. EMBO J. **13:**1197–1204.
6. **Caputi, M., G. Casari, S. Guenz, R. Tagliabue, A. Sidoli, C. A. Melo, and F. E. Baralle.** 1994. A novel bipartite splicing enhancer modulates the differential processing of the human fibronectin EDA exon. Nucleic Acids Res. **22:**1018–1022.
7. **Carlo, T., D. A. Sterner, and S. M. Berget.** 1996. An intron splicing enhancer containing a G-rich repeat facilitates inclusion of a vertebrate micro-exon. RNA **2:**342–353.
8. **Chabot, B., M. Blanchette, I. Lapierre, and H. La Branche.** 1997. An intron element modulating 5′ splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. Mol. Cell. Biol. **17:**1776–1786.
9. **Conrad, R., K. Lea, and T. Blumenthal.** 1995. SL1 trans-splicing specified by AU-rich synthetic RNA inserted at the 5′ end of *Caenorhabditis elegans* pre-mRNA. RNA **1:**164–170.
10. **Crispino, J. D., B. J. Blencowe, and P. A. Sharp.** 1994. Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1snRNP. Science **265:**1866–1869.
11. **Dirksen, W. P., R. K. Hampson, S. Qiang, and F. M. Rottman.** 1994. A purine-rich exon sequence enhances alternative splicing of bovine growth hormone pre-mRNA. J. Biol. Chem. **269:**6431–6436.
12. **Engelbrecht, J., S. Knudsen, and S. Brunak.** 1992. G+C-rich tract in 5′ end of human introns. J. Mol. Biol. **227:**108–113.
13. **Goodall, G. J., and W. Filipowicz.** 1989. The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. Cell **58:**473–483.
14. **Guo, M., P. C. H. Lo, and S. M. Mount.** 1993. Species-specific signals for the splicing of a short *Drosophila* intron in vitro. Mol. Cell. Biol. **13:**1104–1118.
15. **Hawkins, J. D.** 1988. A survey on intron and exon lengths. Nucleic Acids Res. **16:**9893–9908.
16. **Huh, G. S., and R. O. Hynes.** 1993. Elements regulating an alternatively spliced exon of the rat fibronectin gene. Mol. Cell. Biol. **13:**5301–5314.
17. **Huh, G. S., and R. O. Hynes.** 1994. Regulation of alternative pre-mRNA splicing by a novel repeated hexanucleotide element. Genes Dev. **8:**1561–1574.
18. **Humphrey, M. B., J. Bryan, T. A. Cooper, and S. M. Berget.** 1995. A 32-nucleotide exon-splicing enhancer regulates usage of competing 5′ splice sites in a differential internal exon. Mol. Cell. Biol. **15:**3979–3988.
19. **Kennedy, C. F., and S. M. Berget.** 1997. Pyrimidine tracts between the 5′ splice site and branch point facilitate splicing and recognition of a small *Drosophila* intron. Mol. Cell. Biol. **17:**2774–2780.
20. **Lamond, A. I., M. M. Konarska, and P. A. Sharp.** 1987. A mutational analysis of spliceosome assembly: evidence for splice site collaboration during spliceosome formation. Genes Dev. **1:**532–543.
21. **Lavigueur, A., H. La Branche, A. R. Kornblihtt, and B. Chabot.** 1993. A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. Genes Dev. **7:**2405–2417.
22. **Lou, H., A. J. McCullough, and M. A. Schuler.** 1993. 3′ splice site selection in dicot plant nuclei is position dependent. Mol. Cell. Biol. **13:**4485–4493.
23. **Lou, H., Y. Yang, G. J. Cote, S. M. Berget, and R. F. Gagel.** 1995. An intron enhancer containing a 5′ splice site sequence in the human calcitonin/calcitonin gene-related peptide gene. Mol. Cell. Biol. **15:**7135–7142.
24. **Lou, H., R. F. Gagel, and S. M. Berget.** 1996. An intron enhancer recognized by splicing factors activates polyadenylation. Genes Dev. **10:**208–219.
25. **MacMillan, A. M., P. S. McCaw, J. D. Crispino, and P. A. Sharp.** 1997. SC35-mediated reconstitution of splicing in U2AF-depleted nuclear extract. Proc. Natl. Acad. Sci. USA **94:**133–136.
26. **Matunis, M. J., J. Xing, and G. Dreyfuss.** 1994. The hnRNP F protein: unique primary structure, nucleic acid-binding properties, and subcellular localization. Nucleic Acids Res. **22:**1059–1067.
27. **McCullough, A. J., H. Lou, and M. A. Schuler.** 1993. Factors affecting authentic 5′ splice site selection in plant nuclei. Mol. Cell. Biol. **13:**1323–1331.
28. **McCullough, A. J., and M. A. Schuler.** 1993. AU-rich intronic elements affect pre-mRNA 5′ splice site selection in *Drosophila melanogaster*. Mol. Cell. Biol. **13:**7689–7697.
29. **McCullough, A. J., C. E. Baynton, and M. A. Schuler.** 1996. Interactions across exons can influence splice site recognition in plant nuclei. Plant Cell **8:**2295–2307.
30. **Min, H., R. C. Chan, and D. L. Black.** 1995. The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event. Genes Dev. **9:**2659–2671.
31. **Mount, S. M., C. Burks, G. Hertz, G. D. Stormo, O. White, and C. Fields.** 1992. Splicing signals in *Drosophila*: intron size, information content and consensus sequences. Nucleic Acids Res. **20:**4255–4262.
32. **Nakai, K., and H. Sakamoto.** 1994. Construction of a novel database containing aberrant splicing mutations of mammalian genes. Gene **141:**171–177.
33. **Nussinov, R.** 1988. Conserved quartets near 5′ intron junctions in primate nuclear pre-mRNA. J. Theor. Biol. **133:**73–84.
34. **Nussinov, R.** 1989. Conserved signals around 5′ splice sites in eukaryotic nuclear precursor RNAs: G-runs are frequent in the introns and C in the exons near both 5′ and 3′ splice sites. J. Biomol. Struct. Dyn. **6:**985–1000.
35. **Ramchatesingh, J., A. M. Zahler, K. M. Neugebauer, M. B. Roth, and T. A. Cooper.** 1995. A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. Mol. Cell. Biol. **15:**4898–4907.
36. **Reed, R.** 1995. Initial splice site pairing during pre-mRNA splicing. Curr. Opin. Genet. Dev. **6:**215–229.
37. **Robberson, B. L., G. J. Cote, and S. M. Berget.** 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. **10:**84–94.
38. **Ryan, K. J., and T. A. Cooper.** 1996. Muscle-specific splicing enhancers regulate inclusion of the cardiac troponin T alternative exon in embryonic skeletal muscle. Mol. Cell. Biol. **16:**4014–4023.
39. **Sirand-Pugnet, P., P. Durosay, E. Brody, and J. Marie.** 1995. An intronic (A/U)GGG repeat enhances the splicing of an alternative intron of the chicken β-tropomyosin gene. Nucleic Acids Res. **23:**3501–3507.
40. **Solovyev, V. V., A. A. Salamov, and C. B. Lawrence.** 1994. Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. Nucleic Acids Res. **22:**5156–5163.
41. **Staknis, D., and R. Reed.** 1994. SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. Mol. Cell. Biol. **14:**7670–7682.
42. **Sterner, D. A., T. Carlo, and S. M. Berget.** 1996. Architectural limits on split genes. Proc. Natl. Acad. Sci. USA **93:**15081–15085.
43. **Sun, Q., A. Mayeda, R. K. Hampson, A. R. Krainer, and F. M. Rottman.** 1993. General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. Genes Dev. **7:**2598–2608.
44. **Tacke, R., and J. L. Manley.** 1995. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. EMBO J. **14:**3540–3551.
45. **Tacke, R., and J. L. Manley.** 1997. Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. Proc. Natl. Acad. Sci. USA **94:**1148–1153.
46. **Talerico, M., and S. M. Berget.** 1990. Effect of 5′ splice site mutations on splicing of the preceding intron. Mol. Cell. Biol. **10:**6299–6305.
47. **Talerico, M., and S. M. Berget.** 1993. Intron definition in splicing of small *Drosophila* introns. Mol. Cell. Biol. **14:**3434–3445.
48. **Tanaka, K., A. Watakabe, and Y. Shimura.** 1994. Polypurine sequences within a downstream exon function as a splicing enhancer. Mol. Cell. Biol. **14:**1347–1354.
49. **Tarn, W. Y., and J. A. Steitz.** 1994. SR proteins can compensate for the loss of U1snRNP functions in vitro. Genes Dev. **8:**2707–2717.
50. **Watakabe, A., K. Tanaka, and Y. Shimura.** 1993. The role of exon sequences in splice site selection. Genes Dev. **7:**407–418.
51. **Wise, J. A.** Personal communication.
52. **Xu, R., J. Teng, and T. A. Cooper.** 1993. The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. Mol. Cell. Biol. **13:**3660–3674.
53. **Yeakley, J. M., F. Hedjran, J.-P. Morfin, N. Merillat, M. G. Rosenfeld, and R. B. Emeson.** 1993. Control of calcitonin/calcitonin gene-related peptide pre-mRNA processing by constitutive exon and intron elements. Mol. Cell. Biol. **13:**5999–6011.