A Native RNA Secondary Structure Controls Alternative Splice-site Selection and Generates Two Human Growth Hormone Isoforms*

(Received for publication, January 31, 1992)

Patricia A. Estes§§, Nancy E. Cooke§§, and Stephen A. Liebhaber‡‡

From the ¶Howard Hughes Medical Institute, Departments of §Genetics, ‡Biology, and ‡‡Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Consensus sequences at the splice donor, splice acceptor, and lariat branch point regions are necessary but insufficient determinants of splice-site selection in nuclear precursor mRNAs. Sequences outside of these regions can have a significant effect on the utilization of splice sites. Although the mode of action of such sequences is undefined in most cases, higher order RNA structures have been suggested as a potential contributor to splice-site selection. During a detailed analysis of the splicing patterns of the human growth hormone transcript, we located 2 bases in the vicinity of the exon 3 major splice-acceptor site (B) which facilitate the utilization of a competing downstream acceptor (B'). The effects of a series of site-specific mutations on the splicing pattern demonstrate that these 2 bases function by stabilizing a specific stem-loop structure in the native transcript. This defined secondary structure selectively encompasses the upstream B splice-acceptor site together with its lariat branch point region. Increasing the predicted stability of this stem by point mutations results in a corresponding shift in splicing towards the alternative B' splice-acceptor site. These results indicate that a specific secondary structure within the native human growth hormone transcript controls the relative utilization of two competing splice-acceptor sites with the consequent generation of two functionally distinct hormone isoforms.

Alternative splicing is a powerful post-transcriptional mechanism for regulating gene expression. By utilizing different splice sites within the same precursor RNA, multiple transcripts and the protein products they encode can be generated from a single gene (for review, see Ref. 1). Prior to splicing, nuclear mRNA precursors must progress through an ordered pathway in which splicing factors assemble onto the transcript and form a complex called the spliceosome (for reviews, see Refs. 2–5). Subsequently, splicing proceeds through two transesterification steps. In the first step, the lariat branch point A located upstream of the 3' splice acceptor site forms a 2',5'-phosphodiester linkage (lariat) with the G residue at the 5' terminus of the intron. The second transesterification reaction is between the free 3' end of the upstream exon and the 3' acceptor site, resulting in release of the intron and ligation of the two exons. Primary structural determinants of splice-site selection (consensus sequences) are found at the splice donor, splice acceptor, and lariat branch point. However, only a fraction of the sites that conform to these primary consensus sequences are actually selected during splicing and sequences remote from splice sites can have a major effect on their utilization (6–11).

Several authors have suggested that one way in which sequences other than the primary consensus sequences might influence splicing patterns is to alter the secondary structure within the precursor RNA and thereby alter the utilization of a splice site (12–16). The potential of secondary structure to affect splice-site selection was first demonstrated with artificially constructed transcripts. Using the adenovirus tripartite leader, the utilization of an entire exon or a single donor or acceptor splice site when sequestered within a stem-loop structure was shown to be inversely proportional to the stability of the stem in vitro (16). However, corresponding inhibition of splice-site selection within the same transcript in vivo was only observed when the stem length was greater than 50 nucleotides, a size exceeding most duplexes found in native mRNA transcripts (16). This difference between the in vitro and in vivo results may reflect a difference in the ability of mRNA to form secondary structure in the two systems. In vitro, the entire transcript is synthesized and subsequently added to a splicing extract, whereas in vivo, splicing factors as well as other proteins may bind the transcript as it is made. The latter in vivo situation may strongly favor the formation and stabilization of local versus long distance secondary structures. This speculation was confirmed by studies in which a donor site was sequestered in a stem of fixed size (13). In vivo utilization of the sequestered splice site was inhibited only if the connecting loop was below a certain minimum size (50 bases). These results suggest that local rather than long distance secondary structures are more likely to affect the splicing pattern of precursor RNAs in vivo. A role for specific secondary structures in the regulation of splice-site selection has now been documented for the chicken β-tropomyosin transcript (17, 18), the E1A transcript of adenovirus 2 (19), the E3 region of adenovirus 2 (20), and the immunoglobulin μ heavy chain pre-mRNA (21).

In a detailed comparison of splicing patterns used by two highly similar transcripts encoding the pituitary and placental growth hormones, hGH-N and hGH-V, respectively, we have located two divergent bases critical to the selective activation of an alternative splice-acceptor site in the hGH-N transcript.

*This work was supported in part by National Institutes of Health Grants HD 25147 and MOD 1-1015. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

**An Investigator in the Howard Hughes Medical Institute. To whom all correspondence should be addressed: Dept. of Genetics, Room 437A, 422 Curie Blvd, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6149. Tel.: 215-898-7834; Fax: 215-898-1257.

1The abbreviations used are: hGH-N human growth hormone; hGH-V, human growth hormone variant; bp, base pair; PCR, polymerase chain reaction; RT, reverse transcriptase; BPV, bovine papilloma virus.
(22). In the present study we provide evidence that this dinucleotide activates the alternative splicing pathway by stabilizing a specific stem-loop structure encompassing the major splice-acceptor site. These results indicate that a specific structure within the native hGH-N transcript controls the relative utilization of two competing splice-acceptor sites.

MATERIALS AND METHODS

Site-directed Mutagenesis—The two 294-base pair SacI-SmaI fragments containing the third exons and surrounding intron sequences of hGH-N and hGH-V were separately subcloned into SacI-SmaI-digested M13 mp19 RF and the single-stranded forms were used as templates for site-directed mutagenesis (23) with modifications described in Ref. 22. The positions of the two subcloning sites, SacI and SmaI in introns 2 and 3, respectively, are shown in Fig. 1. Oligonucleotide primers used for site-specific mutagenesis are listed in Table I. To generate the V-47+24 and V-18+24 mutations, V+24 M13 mp19 was used as template for the V-47 and V-18 primers. The mutated SacI-SmaI fragments of the hGH-N or hGH-V genes were reintroduced into the hGH-N gene using the native restriction sites and the recomconstituted hGH-N or hGH-NV3 genes were then transfected to the pBPV-MTX expression vector. The mutated SacI-SmaI fragments of the hGH-V gene were cassetted into a hGH-N background rather than the hGH-V transcript in order to eliminate the effects of more distal sequences (22) and allow us to focus on the effects of sequences proximal to exon 3. All mutations were confirmed by dideoxy sequencing (24). The pBPV-MTX vector (described in detail in Ref. 6, a gift from D. Hamer, National Institutes of Health) consists of the bovine papilloma virus genome, the mouse metallothionein gene, and the 2.6-kilobase “poison-minus” sequences of pBR322.

Cell Culture and RNA Isolation—Thirty μg of the pBPV-MTX-hGH recombinant supercoiled plasmid, together with 3 μg of a RSVneo supercoiled plasmid were introduced into a mouse mammary tumor cell line, C127, by calcium phosphate co-precipitation (25, 26). After 3–4 d of selection minimal essential media containing 400 μg/ml G418 (27), total RNA was isolated from a single P-100 Petri dish of stably transformed foci and used without further purification in the RT/PCR assay (detailed in Ref. 22).

Nuclear RNA was isolated (28) with the following modifications. Pelleted nuclei were homogenized in a 6 M guanidine HCl, 10 μm diithiothreitol solution and the liberated nuclear RNA was sedimented through a 5.7 M CsCl cushion (29), ethanol precipitated, and used in the nuclear RT/PCR assay described below.

RT/PCR Assays—Reverse transcription of total cellular RNA was primed with a synthetic oligonucleotide (V14) which anneals to the 5′ end of exon 3 of both the hGH-N and hGH-V transcripts. The cDNA products, purified by phenol extraction, were added to a PCR reaction containing Taq polymerase and amplified for 25 cycles between the V14 oligonucleotide and a 3′-end labeled 5′ primer corresponding to the 5′ end of the second exon of hGH-N (V13). The products of the PCR reaction were analyzed on a 5% polyacrylamide, 8 M urea gel and quantified with a Phosphor Imager (Molecular Dynamics, Sunnyvale, CA). Under the conditions used, the amplifications of the two splicing products are linear and accurately reflect input levels of mRNA (22).

In order to assay partially spliced nuclear RNA, an oligonucleotide that hybridizes to intron 3 was used to prime cDNA synthesis with the nuclear RNA preparations (22, 31) and in transfected mouse C127 fibroblasts (32). We have previously identified three sequence differences within the base of exon 3 which is designated as +1.

RESULTS AND DISCUSSION

The hGH-N Transcript Contains Two Alternative Splice-acceptor Sites within Exon 3—The transcripts of the hGH-N and hGH-V genes differ in their splicing patterns despite 92% identity (6). The hGH-N transcript contains two active splice-acceptor sites in exon 3; the major site (B) at the 5′ end of the exon and an alternative site (B′) 45 bases within the exon (6, 31; see Fig. 1). These two acceptors, B and B′, are used at a ratio of 9:1 in the hGH-N transcript both in vivo (31) and in transfected mouse C127 fibroblasts (6), whereas the B′ site in the hGH-V transcript is inactive (30). At a corresponding 9:1 ratio, the hGH-N gene encodes the 22-kDa growth hormone protein and a 20-kDa isoform which displays a selective reduction in insulin-like properties (for review, see Ref. 32). We have previously identified three sequence differences within exon 3 of the hGH-N and hGH-V transcripts.
that account for the selective activity of the alternative (B') acceptor site in the hGH-N transcript (Fig. 1) (22): 1) an A at position +24, and 2) the CA dinucleotide at positions +17+18. The A at position +24 of the hGH-N transcript appears to function as an essential lariat branch point for the B' acceptor. Changing this A to the G found at the corresponding position in the hGH-V transcript completely eliminates B' utilization in the hGH-N transcript (mutation N+24 in Fig. 1). Changing the sequences around this A to create a lariat branch point region with a perfect match to the yeast and mammalian consensus UACUAAC (mutation N+19 in Fig. 1) causes splicing to take place exclusively at the B' acceptor (studies detailed in Ref. 22). We have confirmed that these mutations act directly at the level of nuclear RNA splicing, and have no significant effect on the nuclear stability, transport, or cytoplasmic stability of the B and B' spliced mRNAs by comparing the pattern of splice-site selection in cytoplasmic and in partially-processed nuclear RNAs (Fig. 2).

The ratios of splice-site selection measured in the nuclear and cytoplasmic RNAs are identical: hGH-N uses both B and B' at a ratio of 9:1, mutation N+24 causes all transcripts to use the B site, and mutation N+19 results in the production of transcripts that splice exclusively to B' (Fig. 2).

The results described above confirm the importance of the A at position +24 to the activation of the B' splice-site in the hGH-N transcript. However, to fully activate the B' site in hGH-V, the G at position +24 must be replaced by A, and in addition, the UG at positions +17+18 must also be replaced by the CA found at the corresponding positions in the hGH-N transcript (V+17+24, Fig. 1). The reciprocal substitution of CA with UG in the hGH-N transcript (N+17, Fig. 1) fully suppresses the B' acceptor (22). These data demonstrate that the +17+18 dinucleotide exerts a controlling effect on the level of B' splice-site selection.

**Stem I Influences B' Splice-site Selection**—The two bases at +17+18 are located outside of the identified splicing consensus regions. They could alter splicing through an unexpected effect on primary structure, or by effects on the secondary or higher order structure of the transcript. The area surrounding the B and B' acceptor sites was inspected for possible secondary structures with the aid of the computer program RNA FOLD (33). To search for a possible secondary structure that controls splicing in this system, we set several criteria: 1) the secondary structure should include the B acceptor site, but not the B' acceptor site; 2) the structure should be selectively stabilized by the CA at positions +17+18 in the hGH-N sequence and destabilized by the presence of a UG at this position as is found in the hGH-V sequence; and 3) the structure should not include a loop size of greater than 50 nucleotides (as discussed in the Introduction) (13).

---

**Fig. 2. Nuclear RNA analysis.** Top, schematic representation of the RT/PCR analysis of nuclear RNA. The hGH-N gene is shown at the top with the exons indicated by the numbered boxes and the introns represented by the connecting lines. An antisense oligonucleotide (NRT) which hybridizes to intron 3 was used to prime the reverse transcription of nuclear RNA from transfected cell lines. The cDNA generated by this reaction was then amplified between a 32P-labeled sense primer (V13) which anneals to the 5' end of the second exon and a nested 3' antisense primer (NX3) complementary to the exon 3/intron 3 border. The three possible products, 506, 299, and 254 bp are shown. cDNA synthesis was primed from the total RNA preparation with an antisense 3' primer (EX3) and then amplified between this primer and the 32P sense primer (V13). Bottom: left panel, autoradiograph of a RT/PCR analysis of nuclear RNA isolated from hGH-N, N+24, and N+19 transfected cell lines. The products were analyzed on a 5% polyacrylamide, 8 M urea gel. The positions of the three products and their respective sizes are noted to the right of the gel. Right panel, autoradiograph of a RT/PCR analysis of total RNA isolated from hGH-N, N+24, and N+19 transfected cell lines. The products were analyzed on a 5% polyacrylamide, 8 M urea gel. The positions of the two products and their respective sizes are noted to the right of the gel. The control lane contains amplified hGH-N cDNA.
fore, we examined the region extending from 50 nucleotides upstream of the B acceptor site to the polypyrurimidine tract of the B' acceptor site (33 base pairs downstream of the B acceptor site). By varying the window size used in the RNA FOLD program, we found a number of possible structures. Two distinct stem-loop structures were identified which met all of the above criteria. Stem I (bold line in Fig. 1; Fig. 3A) sequesters the B acceptor and its putative branch site; stem II (dashed line in Fig. 1; Fig. 4A) sequesters the B acceptor in the loop but excludes its putative branch site. We designed mutations to test if one or both of these structures modulates the relative activity of the two acceptor sites. Substitution of CA by UG at +17+18 which is known to eliminate B' activity in the hGH-N transcript (22), destabilizes both stems (N+17, Figs. 1, 3R, and 4R). To specifically test whether this loss of R' activity resulted from the destabilization of stem I, we introduced a compensatory CA substitution at positions -47 to -48 in the hGH-NV3 transcript that should selectively reestablish the base pairing in stem I (V-47; Figs. 1 and 3B). When introduced into hGH-NV3 alone, this substitution failed to activate the B' acceptor (0%, Fig. 3C). However, when this mutation was introduced into the V+24 transcript containing the putative lariat branch site, B' was activated to levels approximating native hGH-N (V-47+24, Fig. 3, B and C). The slightly lower activity of B' in V-47+24 than in hGH-N (7% versus 9%) parallels its somewhat higher \(\Delta G\) value (74 versus 10.2 kcal/mol) reflecting a difference in base stacking; the three G-C pairs in succession within the native hGH-N stem are interrupted with an A-U pair in the V-47 compensatory mutation. Therefore, substitution of UG for CA at positions +17+18 suppresses B' activity and a compensatory mutation 64 bases upstream of this site reverses

\[\begin{align*}
\text{A.} \\
\text{B.} \\
\text{C.}
\end{align*}\]

\text{FIG. 3. Stem I controls B versus B' splice-site selection.} A, schematic representation of the position within the hGH-N transcript of the secondary structure referred to as stem I. The line represents the second intron and the numbered boxes represent the adjacent second and third exons of the hGH-N transcript. The region between B and B' functions as either intron or exon sequences depending on which splice-acceptor site is used. The B and B' splice-acceptor sites are labeled and their putative lariat branch sites are shown (A and A', respectively). B, stem I of hGH-N and its derivative transcripts generated by site-directed mutations. The nucleotides paired in each stem are shown. The \(\Delta G\) values were calculated (42) and are shown in the loop of each stem in kcal/mol. The position of the B acceptor site within the loop is indicated. The positions of the putative lariat branch points for the B and B' acceptors are indicated: the A within the loop and the A' to the right of the stem, respectively. Positions of mutations are indicated by the arrows. The CA (hGH-N) or UG (hGH-V) at positions +17/+18 which were previously shown to be critical determinants of B' utilization (22) are boxed at the base of each stem. Mutations N+17 and V+24 are identical within the region of stem I and therefore their \(\Delta G\) values are also identical. C, RT/PCR analysis of splice-site selection. The amplification products specific to B and B' splice-site selection are shown schematically to the right of the gel. The percent B' splice-acceptor site utilization is shown below each lane.
FIG. 4. Stem II does not affect B versus B' splice-site selection. A, schematic representation of the position within the hGH-N transcript of the secondary structure referred to as stem II. All symbols are as described in the legend to Fig. 3. B, schematic representation of stem II in the native hGH-N and derivatives generated by the site-directed mutations. The CA (hGH-N) or UG (hGH-V) at positions +17+18 that are important for B' utilization are boxed within the right side of the stems. The predicted ΔG values in kcal/mol are shown above each stem. C, RT/PCR analysis of splice-site selection in the native hGH-N transcript and the hGH-N and hGH-NV3 transcripts containing the site-directed mutations. See Fig. 3C for details of the analysis.

A.

B.

C.

this effect. We conclude that the importance of the CA at position +17+18 reflects its effect on secondary structure.

To further test the importance of stem I on the activation of the B' splice acceptor, we selectively increased its stability in a stepwise manner (Figs. 1 and 3B). First, a single A was inserted on the left side of the stem to pair with the bulged U (N−40.5). Second, three additional base substitutions were introduced into this region designed to fully base pair the stem I structure (N−37). These two sets of substitutions sequentially increased the relative activity of the B' acceptor to 19 and 36%, respectively (Fig. 3C). When the calculated free energy values of each of the stem I substitutions were
plotted against the respective proportions of B' utilization, we observed a linear relationship over the range of values tested (Fig. 5).

**Stem II Does Not Affect B' Splice-site Selection**—To test the specificity of the results of stem I, we carried out a similar series of destabilizing and compensatory mutations directed at stem II (dashed underline in Fig. 1; Fig. 4A). Three sets of mutations were introduced to test the controlling effect of stem II (Fig. 4B, panels I-III, respectively). First, to selectively destabilize this stem, the 2 G residues located on the 5' side of stem II were replaced by two C's. This mutation silences the B' acceptor (N-16; Fig. 4C). To determine whether this was the result of destabilization of stem II or reflected the change in primary structure, two compensatory G's were then introduced on the opposite side of the stem (N-16+15). This second site mutation failed to reactivate the B' acceptor, suggesting that the effect of mutation N-16 on splicing was mediated via its primary sequence. Inspection of the sequence in this area reveals a potential explanation for the primary structure effect: introduction of the two additional C's in place of the G's normally found at this position would strengthen the polypyrimidine tract of the B acceptor site (Fig. 1). The length of the polypyrimidine tract can increase the efficiency of splicing to a splice-acceptor site (34). Next, a compensatory substitution was made opposite the UG at positions +17+18 in exon 3 of the V+24 transcript (V-18+24). This mutation, designed to stabilize stem II, had no effect on the splicing pattern of the V+24 transcript (Fig. 4C). As a final test of stem II, 6 changes were introduced into the hGH-N transcript designed to stabilize it considerably (N-15G). A wobble G was used at position -15, rather than A, to avoid introducing an AG upstream of the B site; Seif et al. (35) noted that AG's are never present within 15 nucleotides upstream of an acceptor site. These changes failed to increase the relative utilization of the B' site and in fact decreased the relative utilization of B' to 6% (Fig. 4C). The effects of these 3 sets of mutations are all inconsistent with a relationship between the calculated stability of stem II and the relative activity of the B and B' splice-acceptor sites, despite the higher theoretical stability of stem II than stem I ($\Delta G = -23.6$ versus $-10.2$ in the native hGH-N transcript).

It should be noted that one of the mutations N-16+15, that was specifically introduced to test the presence of stem II, also impacts on stem I (Fig. 1). The C to G substitutions at positions +15 and +16 would destabilize stem I with a predicted decrease in B' utilization while the other two substitutions in this mutation (−16 and −17) would not be expected to have any effect on stem I. Although the loss of B' utilization in the N-16+15 mutation is in fact consistent with the predicted effect of the +15+16 substitutions on stem I, they were only carried out in combination with the two changes made at positions −16 and −17 which by themselves significantly compromise B' utilization. Therefore, the effect of the N-16+15 mutation, although consistent with the predicted effect on stem I, cannot be used to support or dispute its presence.

Taken together, the results obtained from testing the effects of stem I and stem II on splice-site selection, suggest the following model. The secondary structure of the native hGH-N transcript is in equilibrium during splice-site selection, with stem I present in 9% of the transcripts. When present, this stem inhibits selection of the B splice site and allows the downstream B' acceptor to be used. By stabilizing stem I (mutations N-40.5 and N-37), the percentage of GH transcripts containing this stem increases with a consequent increase in B' utilization. However, a percentage of transcripts still use the B acceptor and its lariat branch point, suggesting that stem I is not present in this fraction of the GH transcripts. These transcripts most likely have a different structure within this region (stem II or others) that compete with stem I and allow B site utilization.

The relative positions of the B splice-acceptor site and its lariat branch point region differ in stem I and stem II. Previous studies have indicated that if two tandem acceptors are competing for splicing, the splice-acceptor site that is accompanied by the lariat branch point region with the best fit to the consensus will predominate (22, 36, 37). However, in the hGH-N transcript, neither of the two competing acceptors has a clearly superior lariat, creating a delicate balance between the two acceptors. The positioning of the B acceptor site together with its lariat in a specific stem-loop (stem I) inhibits its use so that the downstream B' acceptor site can better compete during spliceosome formation. Although the B acceptor site is sequestered in stem II, its lariat branch point is not included in the stem; this may be permissive for B selection. This suggests that use of the two competing acceptor sites is controlled by the availability of the upstream lariat branch point. Thus, secondary structure as well as primary structure of the lariat branch point regions can alter their ability to participate in the splicing reactions. Stem I may inhibit the recognition of this region of the transcript by blocking binding of the appropriate factors, such as U2AF (38) and the U2 small nuclear ribonucleoprotein (39-41), rendering spliceosome formation at the B lariat and acceptor sites unfavorable. As a result, spliceosome formation at the B' lariat and acceptor sites would be favored, causing utilization of this splice site to increase. Although certain duplexes can be tolerated at a 3' acceptor site, these data indicate that the accessibility of a splice site can be modified through an appropriately positioned local secondary structure.

**REFERENCES**

1. Smith, C. W., Jr., Patton, J. G., and Nadal-Ginard, B. (1989) *Annu. Rev. Genet.* 23, 527-577
2. Green, M. R. (1986) *Annu. Rev. Genet.* 20, 671-708
Secondary Structure and Splice-site Selection

3. Maniatis, T., and Reed, R. (1987) Nature 325, 673-678
4. Padgett, R. A., Grabowski, P. J., Komarska, M. M., Seiler, S., and Sharp, P. A. (1986) Annu. Rev. Biocem. 55, 1119-1150
5. Sharp, P. A. (1987) Science 235, 768-771
6. Cooke, N. E., Ray, J., Watson, M. A., Estes, P. A., Kuo, B. A., and Liefhaber, S. A. (1986) J. Clin. Invest. 77, 270-275
7. Hwang, L.-H. S., Park, J., and Gilhoa, E. (1984) Mol. Cell. Bioi. 4, 2289-2297
8. Mardon, H. J., Sebastiani, G., and Baralle, F. E. (1987) Nucleic Acids Res. 15, 7725-7728
9. Nelson, K. K., and Green, M. R. (1988) Genes & Dev. 2, 319-329
10. Reed, R., and Maniatis, T. (1986) Cell 46, 681-690
11. Somasekhar, M., and Mertz, J. E. (1985) Nucleic Acids Res. 13, 5591-5609
12. Eperon, L. P., Estibeiro, J. P., and Eperon, I. C. (1986) Nature 324, 282-285
13. Meier, S. H. (1988) EMBO J. 7, 2523-2532
14. Solnick, D., and Lee, S. I. (1987) Mol. Cell. 1, 1545-1552
15. Zhuang, Y., and Weiner, A. M. (1989) Genes & Dev. 3, 1545-1552
16. Turner, D. H., Sugimoto, N., and Freier, S. M. (1988) Annu. Rev. Biophys. Biophys. Chem. 17, 167-192