Volume 15 Number 16 1987 Nucleic Acids Research

A new natural hGH variant—17.5 kd—produced by alternative splicing. An additional consensus sequence which might play a role in branchpoint selection

Corine M.Lecomte, André Renard and Joseph A.Martial

Laboratoire Central de Génie Génétique, University of Liège, B-4000 Sart-Tilman, Belgium

Received June 4, 1987; Accepted June 30, 1987

ABSTRACT

From a human pituitary cDNA library, we have cloned 3 distinct human growth hormone (hGH) cDNAs, coding respectively for the 22 K hGH, the 20 K variant, and a yet unknown 17.5 K variant.

S1 mapping analysis using human pituitary RNA confirms the existence of at least four distinct hGH mRNAs originating from alternative acceptor sites at the second intron of the primary transcript.

We have analysed the hGH gene sequence to explain the high frequency of alternative splicings which occur only at this location. In this study we propose CTTGNNPyPyPy as an additional consensus sequence guiding the selection of the branched nucleotide.

INTRODUCTION

Mechanisms of alternative RNA splicing are commonly implicated in the generation of protein diversity. A few genes use different promoters like the α-amylase gene (1), or different polyadenylation signals such as the calcitonin gene (2) or the immunoglobulin μ gene (3). However, there are some genes (γ -fibrinogen (4), α A-crystallin (5), troponin T (6), myelin-basic protein (7), rat fibronectin (8), rat proopiomelanocortin (9)) that yield several different mature messenger RNAs by an alternative processing of the same primary transcript. The human growth hormone (hGH) normal gene seems to belong to this last group.

Human pituitary extracts contain a mixture of GH variants in addition to the predominant 22,000-dalton form. Most of them originate from degradations or post-translational modifications of the major form. Nevertheless, there is a 20,000-dalton

variant that constitutes 5-10% of the total pituitary growth hormone and differs from the 22 K protein by the lack of residues 32 to 46 (10,11,12). Both 22 K and 20 K hormones are encoded by a unique chromosomal gene (13). Their biosynthesis results from alternative splicings of the same primary gene transcript at the acceptor site of the second intron (14,15).

Little is known about the exact mechanisms of splice site selection. According to the current theory, the first step in the RNA splicing process is the selection of the 3' splice site and its polypyrimidine tract, accompanied by the association of factor(s) with an intronic branch point site (16,17). It is thought that this branch point site is mainly determined by its distance to the 3' splice junction (18,19,20). Thus it seems obvious that the choice of the 3' splice site is a key element both in the general mechanism of splicing and in the special case of alternative processing. This is in agreement with the observation that alternative RNA processing more often involves the use of an alternative acceptor site (as it occurs with hGH mRNA) than of an alternative donor site.

In the present communication, we describe the construction of a human pituitary cDNA library in phage lambda. This library was screened with specific DNA probes to select clones corresponding to hGH cDNAs. Restriction mapping and sequence analysis distinguished three kinds of cDNAs; one coding for the 22 K hGH, one coding for the 20 K variant and a third coding for a 17.500-dalton hGH variant.

The cDNA encoding the 22 K growth hormone has already been cloned and expressed in bacteria (21,22) and the biosynthetic 22 K methionyl-hGH purified and compared with its natural counterpart (23,24,25). On the other hand, using oligonucleotide-mediated mutagenesis, Adelman et al. (26) have constructed the nucleotide sequence encoding the 20 K variant, and the deleted gene has been expressed in E.coli. Nevertheless, the cloning of the corresponding "natural" cDNA has not yet been reported in the literature.

Our S1 nuclease analysis provides evidence for the existence of multiple RNAs. We believe that the hGH gene may be a good system to approach some unsolved questions about the accura-

cy of RNA splicing and the regulation of splice site selection in alternative pathways.

EXPERIMENTAL PROCEDURES

mRNA preparation

Total RNA was isolated from human pituitary tumors by the guanidinium thiocyanate method (27,28). Poly(A+)RNA was purified by affinity chromatography on oligo(dT) cellulose (29).

Synthesis and cloning of cDNA

Oligo$(dT)_{12-18}$ was used to prime reverse transcription of poly(A+)RNA. The mRNA : cDNA hybrid molecules were 3'-dCMP-tailed, and the synthesis of the second strand was primed with oligo$(dG)_{12-18}$ (30). Double-stranded cDNAs, after addition of synthetic EcoRI linkers, were inserted into the unique EcoRI restriction site of λ 641DNA, an insertion vector (λ imm^{434}) (31).

The recombinant DNA was packaged *in vitro*, using packaging extracts prepared from BHB 2688 and BHB 2690 strains.

Bacterial strain POP13b was used to select recombinant phages. It carries a mutation (lyc 7) in the gene whose product normally antagonizes to establishment of the lambda repression, allowing an important increase of the frequency of stable lysogenisation (32).

Screening and sequencing

Lambda transformants were transfered onto nitrocellulose filters (33). Hybridizations were performed in 6XSSC, 1X Denhardt's solution and 100 ng/ml of sheared salmon sperm DNA at 65°C (34). Filters were washed three times (20 minutes each) in the same buffer and at the same temperature, and autoradiographed.

We used two probes isolated from chGH800/pBR322 (21). The first one is a nick-translated fragment containing the 200 bp of hGH cDNA located between the PstI and BamHI restriction sites. The second probe is 50 nucleotides long and was isolated by digestion with BstNI and PstI, and labeled with ^{32}P using T4-polynucleotide kinase (Figure 1, A and B).

DNA sequence analysis was performed according to the methods of Maxam and Gilbert (35), and Sanger, Nicklen and Coulson (36) after cloning in M13 mp8 and mp9.

Table 1. Summary of the hybridization and cell-free translation experiments on RNAs isolated from the pituitary tumors used in this study.

Tumors	Clinical diagnostic	Hybridization with hGH probe	In vitro translation	Hybridization with hPRL probe
803	GH secreting	+	+++	-
824	GH secreting	++	+++	-
847	GH secreting	+++	+++	-
865	GH secreting	-	nd	-
875	GH secreting	++	+++	-
876	non fonctional	-	nd	-
821	non fonctional	-	nd	-
826	GH secreting	+++	+++	-
848	nd	+	nd	-
878	GH secreting	+++	+++	++
810	nd	-	nd	-
872	GH secreting	+++	-	-
929	GH secreting	+++	-	-
931	GH secreting	++	++	-
893	non fonctional	-	nd	-
904	nd	-	nd	-
930	Prl secreting	-	nd	++
938	nd	-	nd	-
943	Acromegaly	+	-	-
955	Acromegaly	++	++	-
J1	GH secreting	++	nd	nd
J2	GH secreting	-	nd	nd
J3	GH secreting	++	nd	nd
J4	GH secreting	++	nd	nd
J5	GH secreting	++	nd	nd
J6	GH secreting	+	nd	nd
J7	nd	-	nd	nd
J8	nd	+++	nd	nd
J9	non fonctional	+	nd	nd
J10	Prl secreting	-	nd	nd
J11	Prl secreting	-	nd	nd

S1 mapping

We used a modification of the Berk-Sharp procedure (37). The RNA and the cDNA probe were heat denatured then hybridized in 80% formamide, 40 mM PIPES pH 6.4, 0.4 M NaCl, 1 mM EDTA for 14 hrs at 57°C. The mixture was diluted ten-fold in S1 buffer containing 20 ug/ml denatured salmon sperm DNA, and digestion with 100 U/ml S1 nuclease was carried out for 30 minutes at 30°C. The precipitated samples were electrophoresed on a 5% polyacrylamide-50% urea gel.

Computer analysis

Computer analysis was performed on a Data General MV4000. Programs were written in C language (38).

DNA sequences were taken from the EMBL data bank (release 6).

RESULTS

Isolation of human pituitary mRNA

Total RNA was isolated from human acromegalic tumors removed by transphenoidal hypophysectomy. The integrity and the relative abundance of growth hormone messenger RNA in each sample were analysed by cell-free translation in the rabbit reticulocyte lysate system, and by Northern blotting and hybridization with a hGH cDNA probe (21). 19 from the 32 extracted tumors yielded RNA containing variable relative amounts of intact hGH mRNA (Table 1).

The six samples most enriched in hGH mRNA were pooled and their polyadenylated RNAs were purified.

Molecular cloning of cDNA

The poly(A+)RNAs were reverse transcribed into complementary double-stranded DNA using the 3'-dCMP-tailing/oligo(dG) priming system for the second strand synthesis (30). The cDNAs were size fractionated onto Sepharose 4B and synthetic EcoRI linkers were added to the cDNA molecules larger than 500 nucleotides.

The resulting cDNAs were inserted into the EcoRI site of the bacteriophage λ 641 DNA. The recombinant DNAs were packaged *in vitro* and the resulting phages were used to infect *E.coli* POP13b bacteria. 120,000 transformants were obtained starting with 10 ng of purified ds cDNA.

Screening and sequencing of the cDNA library

Two DNA probes were isolated from chGH800/pBR322 (21). The first one is a 200 bp fragment which corresponds to the 5'-end of hGH mRNA (Figure 1, probe A). It will only hybridize to full or nearly full-length cDNA clones. The second probe contains the 50 nucleotides located between the BstNI and PstI restriction sites (Figure 1, probe B), and corresponds to the region deleted in the natural 20 K variant cDNA.

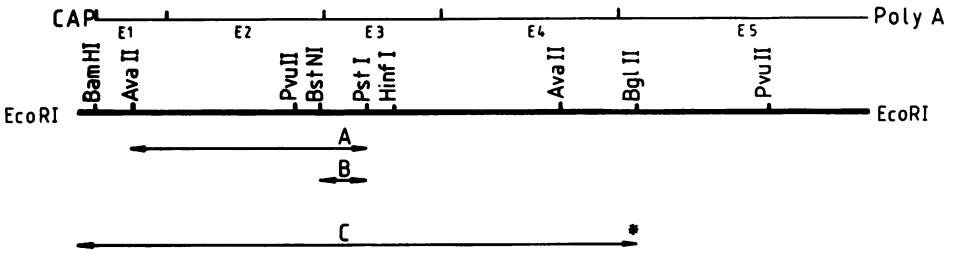

Figure 1. Localization of the probes used for the library screening (A, B) and for the S1 mapping analysis (C).
Arrows show the positions and relative sizes of the probes with regard to the restriction map of the hGH 22 K cDNA, aligned with the above representation of the exons junctions.

In addition to the plaques that hybridized to both probes, several clones gave a positive signal with the first one (A) without hybridizing to the second (B).

Restriction mapping and sequencing

Restriction mapping of the positive clones allowed us

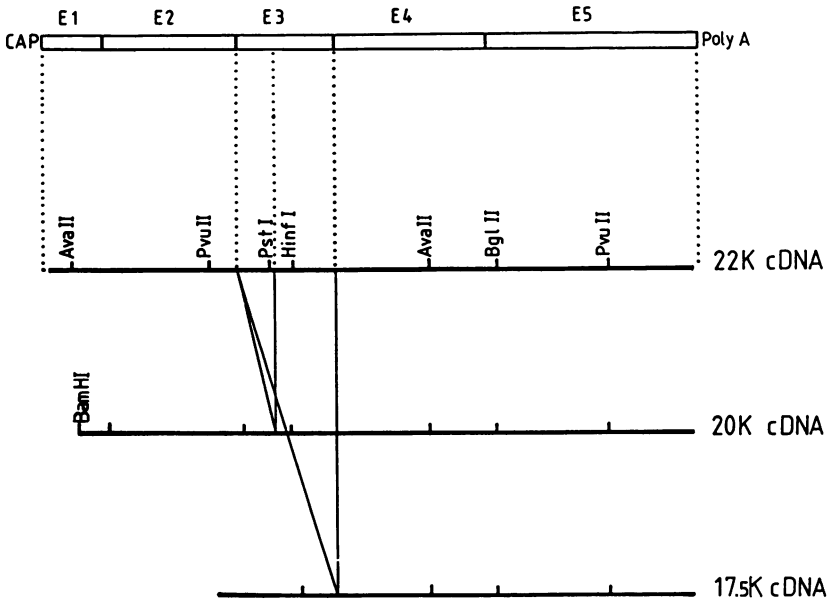

Figure 2. Restriction maps of the three cloned hGH cDNAs. Both the 20 K and 17.5 K deletions are located on the 22 K cDNA map in alignment with the positions of the exons boundaries.

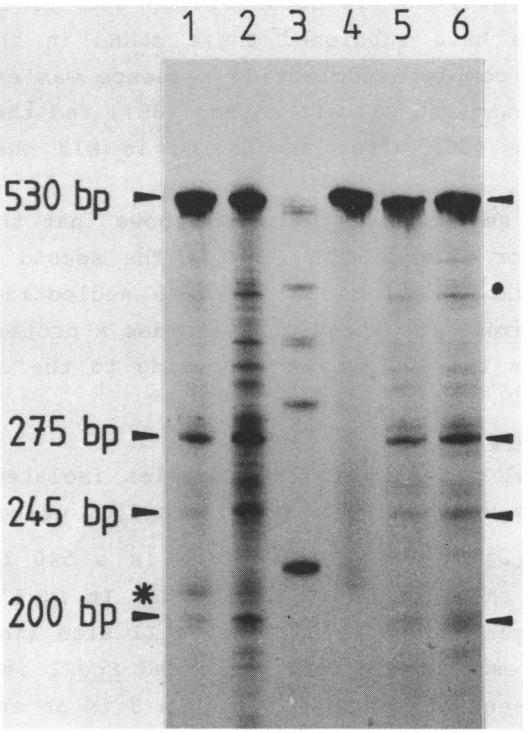

Figure 3. S1 nuclease resistant products.
Lane 3 : HinfI cleaved pBR322 markers, 5'-labeled using T4-polynucleotide kinase. Lane 4 : 540 bp long EcoRI-BglII probe; neither RNA nor S1 digestion. Lanes 1 and 5 : probe hybridized with total pituitary RNA, plus S1 digestion. Lanes 2 and 6 : probe hybridized with poly(A+) pituitary RNA, plus S1 digestion. Lanes 1 and 2 : samples digested 30 minutes at 30°C. Lanes 5 and 6 : samples digested 60 minutes at 30°C. The band indicated by * in lane 1 is also apparent in lane 4 on longer exposure. The band indicated by . has a size of 390 bp and is not present in lane 1, even on longer exposure.

to characterize three types of cDNA (Figure 2). Type I cDNA : this cDNA is similar to the published hGH cDNA sequence (21) except that it contains an AvaII restriction site instead of the BamHI site at its 5'-end. There is a BamHI site 34 nucleotides upstream from this AvaII site. This is consistent with the published sequence of the hGH gene (13,15). Type II cDNA : this cDNA lacks the PstI site located at nucleotide 271 of the first cDNA. Type III cDNA : this cDNA lacks the PstI and HinfI sites at posi-

tions 271 and 304, and is about 120 bp shorter compared with the 22 K cDNA. We have subcloned these cDNAs in the EcoRI site of pBR322. Their complete nucleotide sequence was established according to the Maxam and Gilbert method (35), and the dideoxy sequencing technique (36) after subcloning in M13 phage (mp8 and mp9 vectors).

The sequence of the cDNAs shows that the first cDNA encodes the major form of hGH (22 K). The second cDNA encodes the native 20 K variant since it lacks 45 nucleotides at the 5'-end of the third exon. The third cDNA encodes a protein of 17,500 daltons: it lacks the 120 bp corresponding to the entire third exon (Figure 2).

S1 nuclease mapping

Total and poly (A+)RNA samples isolated from human pituitary tumors were analysed by S1 mapping, using a single-stranded 5'-^{32}p-labeled probe. This probe is a 540 bases BglII-EcoRI fragment from the full-length 22 K cDNA. It is 5'- labeled at the BglII end (Figure 1, probe C). The BglII site lies in the 5' part of the fifth exon of the hGH gene, and EcoRI is the cDNA insertion site (5'-end of the mRNA). Figure 3 is an autoradiogram showing the S1-resistant products fractionated on an urea-polyacrylamide gel. Several protected cDNA bands are clearly visible. Four of the bands (530, 275, 245 and 200 nucleotides long) are present in all samples, even after undergoing the most stringent S1 digestion.

The major band corresponds to the fully protected probe. The small size difference observed between the cDNA probe and the protected fragment results from the dC-tail added for the second strand synthesis. This band thus originates from mRNA identical to the cDNA. In contrast, the three other protected fragments reflect mRNA species with upstream sequences that diverge from the cDNA probe at distinct points. The major one corresponds to the 20 K coding mRNA. The smallest band is consistent with the predicted size for the 17.5 K coding cDNA. The intermediate S1-resistant fragment leads to a divergence point within exon 3, downstream from the 20 K splicing site. This last band corresponds to another variant mRNA that we later call GHI.V.

These S1 mapping results support the notion that the

Figure 4. Predictions for acceptor splice sites. Positions within the sequence presenting the highest score (as defined in the text) are drawn on the schematic map of the hGH gene. The length of the bars is proportional to the score .

distinct messenger RNAs observed are generated by distinct choices of the 3' acceptor splice site of intron B. We decided to search the hGH gene for sequences which could explain these unexpected splicing events.

Computer search for potential acceptor sites in the hGH gene

Consensus sequences at both 5' and 3' junctions are known to be essential for in vitro and in vivo splicing (39,16, 40,41). At the 3' site, the consensus sequence can be represented by $(Py)_{11}$ N Py A G/G (39), or more accurately by the base frequency matrix given by Staden (42) and based on the splicing site compilation of Mount (39). We have scanned the hGH gene (13,15) with this matrix to identify potential acceptor sites. The result of this search is represented in figure 4. The score we attributed to a specific match was obtained by multiplying by each other the base frequencies of each position of the matched sequence as read in the matrix. The higher the score of a sequence, the better its match with the consensus sequence based on the compilation of Mount (39).

We can see (Figure 4) that for introns A, C and D, the best predicted site is the one effectively used. However, for intron B, we found two sites within exon 3 better than the one actually used. The first one corresponds to the 20 K hGH variant mRNA, and the second one to the GHI.V variant mRNA. This intermediate mRNA may thus have arisen from the use of an optional acceptor site, located 28 nucleotides downstream from the 5'-end of the 20 K exon 3.

We observe a high multiplicity of intron B potential ac-

```
          -60                 -40                 -20
           '         '         '         '         '         '
A         agggagagtgttggccTCTTGCTCTccggctccctctgttgccctCTGGTTTCTCCCCAG
B-22K     gaaaagtaacaatgggaGCTGGTCTCcagCGTAGACcttggtgggCGGTCCTTCTCCTAG  (1)
          cagaaaagtaaaatgggagcaggtctcagCTCAGACcttggtgggCGGTCCTTCTCCTAG  (2)
B-20K     cggtccttctcctaggaagaagcctatatCCCAAAGgaacagaagTATTCATTCCTGCAG
B-I.V     tcccaaaggaacagaagtattcattcctgcagaaccCCCAGACctCCCTCTGTTTCTCAG
C         gggagacctgtagtcagagcccccgggcagcacAGCCAATgcccgTCCTTGCCCCTGCAG
D         agaaagggagggaacagtacccaagcGCTTGGCCTCtccttctctTCCTTCACTTTGCAG
```

Figure 5. Summary of the search for consensus sequences at the 3'-ends of all the introns of the hGH gene. Underlined uppercase letters indicate the position of the consensus sequences ("CTTG" box, branchpoint site, 3' splicing site). Underlined lowercase letters indicate the position of "CTTG" boxes with a score lower than 5.10^5.
(1) sequence from (15).
(2) sequence from (13).

ceptor sites, which can explain why alternative splicings occur at the 3'-end of the second intron rather than at the other introns of the hGH gene. Nevertheless, there is no relationship between the frequencies of the alternative splice events and the score of the potential acceptor sites.

Computer search for sequences agreeing with the published consensus branchpoint sequence

Concerning the putative branchpoint site used in lariat formation, several investigators have proposed a consensus sequence (PyXPyTPuAPy) located 18 to 40 nucleotides upstream of the 3' splice site (43,44,45).

In contrast with the yeast TACTAAC box (46,47,48), the sequence used for branch formation in higher eukaryotes is poorly conserved. Nevertheless, its location seems to be a stringent constraint : the branched nucleotide, mainly an A residue, lies at a fixed distance from the intron 3' end (18,19,20,49).

From these data, we have found a coherent branch-forming region in the introns B and C of the hGH gene, as shown in the Figure 5. Concerning the first intron of the gene, Hartmuth and Barta (50) have shown that the branched nucleotide is a C residue, which is, after the A residue, the next most frequently found branched nucleotide (49).

1	2	3	4	5	6	7	8	9	10	
1	1	1	1	1	1	1	1	1	1	A
1	1	1	1	10	3	1	1	1	1	G
1	10	1	1	1	1	1	5	5	10	C
2	1	10	10	1	3	1	5	5	10	T

Table 2. "CTTG" matrix. This matrix does not directly reflect the box frequencies in the population of sequences fitting the consensus sequence. It rather takes into account the importance of some positions and of some bases compared to others (see the text for comments).

Downstream from the intron B acceptor site, we also found branch acceptor regions suitable for each of the alternative splicing sites (Figure 5).

Detection of a new conserved sequence, the "CTTG" box

In many introns, the computer searches have shown a well-conserved region between coordinates -60 and -21 from the 3' splicing cleavage site. The consensus sequence of this region (CTTGNNPyPyPy) can be represented by the matrix given in Table 2 and will be called the "CTTG" box later in this publication. We propose this box as an additional element affecting acceptor splice site recognition and lariat formation.

The real impact of this box was evaluated by looking for its presence in a library of 70 intronic 40 bp-segments (from -60 to -21 from the 3' splice site). This search was done by the Staden method (42), as precedently described in the computer search for potential acceptor splice sites. By this method, using the matrix given in Table 2, we calculated a "score" for each detected "CTTG" box. We filtered the sequence population obtained by fixing the score threshold to the value of 5.10^5. This procedure gave us 46 sequences (in 38 introns) out of the 70 examined.

This frequency is four times higher than the one expected from the calculation based on the average base frequencies. Similar searches in full genes give ratios between one and two. We can thus consider the enrichment of "CTTG" boxes in the 3'-end of introns as highly significant.

Out of the four introns of the hGH gene, three contain a "CTTG" box in the [-60, -21] region upstream from the acceptor splice site (introns A, B and D). However, the "CTTG" box was not

found in intron C and upstream from the 20 K and GHI.V alternative splice sites.

DISCUSSION

We have constructed a cDNA library in lambda phage using human pituitary tumors as a source of messenger RNAs. Three types of GH cDNAs were selected and analysed : the first one encodes the normal pituitary 22 K hormone; the second one, the 20 K variant that represents 5-10% of the total pituitary growth hormone; and the third one, a yet unknown 17,500-dalton protein. The 20 K variant is well characterized (11,51), but a "17.5 K variant", which lacks the 40 amino acid residues coded by exon 3, has never been reported in the literature.

We have further confirmed by S1 nuclease mapping the existence of multiple RNAs corresponding to the distinct cloned cDNAs, either in total or poly(A+) RNA samples extracted from human pituitaries. The lengths of three S1-resistant products, 530 bases, 275 bases and 200 bases, are consistent with the predicted sizes for the 22 K , 20 K and 17.5 K coding mRNAs. Also detectable is an additional band (245 bases) corresponding to another variant mRNA (called GHI.V). Our S1 mapping results are consistent with the idea that these distinct messenger RNAs are generated by a mechanism of alternative splicing at the 3'-end of intron B. The hGH gene may thus be a good model to explore the problem of RNA splicing specificity, both in general and alternative pathways.

First, we have searched for potential acceptor splice sites in the hGH gene sequence, using the frequency matrix given by Staden (42) and based on the splicing site compilation of Mount (39). For introns A,C and D, the best predicted site is the one effectively used. For intron B, we found three potential acceptor sites, including the "normal" one which is seven times less likely than the 20 K one. Such a multiplicity of potential acceptor sites renders splice site selection ambiguous, and that may explain why, in the hGH gene, alternative splicing is observed at the end of intron B rather than at the end of the other introns. It cannot however explain the relative frequencies of the alternative splice events. Additional factors must therefore play

a role in specifying splice sites. Applying the same arguments to the intron B of the bGH (52) and rGH (EMBL bank), we do not predict any alternative splicing in these cases.

Second, we have searched for the published consensus branchpoint sequence (43,44,45) upstream from each acceptor splice site of the hGH gene. For intron B, we found branchpoint sequences agreeing with the 22 K, 20 K and GHI.V splicing sites.

Third, we propose a new consensus sequence ($CTTGNNP_y$ P_yP_y) as an additional element guiding the recognition of the branch-forming region in the lariat formation. More exactly, we propose the matrix given in Table 2, that we called the "CTTG" box. This matrix was deduced from a well-conserved region found in the hGH gene, and its presence was confirmed at the 3'-end of introns of other genes, between coordinates -60 and -21 upstream from the 3' splicing site (Table 3).

Three observations must be stressed as they suggest a role for this "CTTG" box in the splicing mechanism. First, the "CTTG" box is part of larger conserved sequences found in different introns (introns 1, 2, & 3 in the rat actin gene; intron 1 in the α and β -globin genes; intron 1 in the rat and human growth hormone genes). Second, for many introns, we have found a good complementarity between the region containing the "CTTG" box and the 5'-end of the first loop of U_2 snRNA, considering the secondary structure proposed by Black, Chabot and Steitz (53). Finally, in the human β -globin precursor RNA, we observe that the "CTTG" box is part of the intron region that Black, Chabot and Steitz (53) have found associated to U_2 snRNPs in active splicing reactions. In this case, the branchpoint is located 19 nucleotides downstream from the G of "CTTG".

These observations which result from sequence analysis suggest that interactions between the U_2 snRPN and the RNA precursor could be mediated by RNA-RNA base pairing at the level of the "CTTG" box.

Therefore, we propose that the "CTTG" box is, in addition to the branchpoint region and to the polypyrimidine stretch of the acceptor splicing site, a third sequence element guiding 3' splice site recognition and lariat formation.

This element is probably not a determinant for correct

pre-mRNA splicing, since its presence is not observed in all introns. However, its role might be enhanced when there is an ambiguity at the level of the 3' acceptor site.

For instance, in the intron B of the hGH gene, the ambiguity coming from the splice sites multiplicity could be partly "corrected" by the presence of the "CTTG" box upstream from the normal site and by its absence upstream from the alternative 20K and GHI.V acceptor sites (Figure 5).

We observe a similar pattern in the rat fibronectin gene, in which a 5' splice site pairs with three 3' acceptor sites leading to three distinct products. Although it is the second site which fits best with the acceptor consensus sequence, the first site is the most frequently used. This site preference might be explained, as in the hGH gene case, by the fact that only the first acceptor site is preceded by a "CTTG" box.

Table 3. Search for "CTTG" boxes in various introns. 60 bases upstream from 3' splice sites were extracted from sequences taken from the EMBL bank (sequences 1 to 52) or taken from the literature (sequences 53 to 70).

#	EMBL code	intron	Sequence	Pos	Score
1	rnac01	1	TCTTGGCGGC	-47	6.0e+05
2	rnac01	2	CCTAGGGTTT	-56	7.5e+05
			TCTTGGGGGT	-46	6.0e+05
3	rnac01	3	TCTTGTGGCT	-47	3.0e+06
4	rnac01	4			
5	rnac01	5			
6	rnins2	1	ACATGTACCT	-41	7.5e+05
7	rnins2	2	TCTTTGCTTC	-56	1.5e+06
8	rngcry	1	CCTGGGGTCC	-49	7.5e+05
9	rngcry	2			
10	mmcrys	1	CTTTGTCTCT	-57	7.5e+05
			CCTTTGTCTC	-58	7.5e+05
11	mmcrys	2			
12	rnmlc2	1			
13	rnmlc2	2	TCCTGGCTTC	-38	1.5e+06
			CCTTGGGCCA	-50	7.5e+05
14	rnmlc2	3	TCTGGTCCCC	-48	1.5e+06
15	rnmlc2	4			
16	rnmlc2	5	TCATGAACCC	-41	5.0e+05
17	rnmlc2	6			
18	mmigm3	1			
19	mmigm3	2			
20	mmigm3	3	TCTTCCACCC	-37	5.0e+05
21	mmigm1	1	GCTTGACCCT	-32	2.5e+06
			ACTTGGCTTG	-37	7.5e+05
22	mmigm1	2			
23	mmigxx	1			
24	mmigxx	2	CCTTTTCTCC	-46	7.5e+05
25	mmigxx	3			
26	mmigxx	4	CCTTCTGCTT	-46	7.5e+05
27	mmig69	1			
28	mmig71	1			
29	mmigl1	1	TCTTGTCTCT	-49	1.5e+07
30	mmigl4	1			
31	hsegl1	1	TCCTGGCCCC	-49	1.5e+06
32	hsegl1	2	GCTTGGGCCG	-33	7.5e+05
33	hsbgl3	1	TCTTGGGTTT	-60	1.5e+07
34	hsbgl3	2	GCTAGGCCCT	-50	7.5e+05
35	hsdgl1	1	TCTTGGGTTT	-58	1.5e+07
36	hsdgl1	2	TCTTTATTTC	-53	5.0e+05

37	hsggl2	1			
38	hsggl2	2			
39	hsegl1	1	TTTT6CATCT	-44	5.0e+05
40	hsegl1	2	ACTTCTACTC	-47	7.5e+05
41	rngrow2	1	CCTT6CTCCC	-43	2.5e+06
			6ATT66TCCT	-50	7.5e+05
42	rngrow2	2	6CT66TCCCT	-31	7.5e+05
43	rngrow2	3			
44	rngrow2	4	TCTA66TTCT	-55	1.5e+06
45	hsgrow2	1	TCTT6CTCTC	-44	5.0e+06
46	hsgrow2	2	6CT66TCTCC	-43	7.5e+05
47	hsgrow2	3			
48	hsgrow2	4	6CTT66CCTC	-34	7.5e+06
49	btppt3	2			
50	btppt6	5			
51	rnfb65e	1			
52	rnfb65e	2	ACTTATCTCC	-45	7.5e+05
53	rntnt (*)	1			
54	rntnt (*)	3			
55	rntnt (*)	4	CCTT6TCATT	-55	1.5e+06
56	rntnt (*)	5			
57	rntnt (*)	6	CC6T66TTTT	-34	7.5e+05
			TTTT6CCTCT	-30	5.0e+05
58	rntnt (*)	8	TCTTTTTCTT	-40	1.5e+06
			TCTTCCTCTT	-32	5.0e+05
59	rntnt (*)	9	A6TT6TCCCT	-43	7.5e+05
60	rntnt (*)	10			
61	rntnt (*)	11			
62	rntnt (*)	12			
63	rntnt (*)	13	ACT666CCCT	-32	7.5e+05
64	rntnt (*)	14	TCCT666CT6	-59	1.5e+06
65	rntnt (*)	15	CCTT6AA6CC	-51	5.0e+05
66	rntnt (*)	16			
67	rntnt (*)	17			
68	rnfib (**)	1a	6CTTCTTCTT	-50	7.5e+05
			6CTTT6CTTT	-37	7.5e+05
69	rnfib (**)	1b			
70	rnfib (**)	1c			

The matrix of table 2 was overlayed from position -60 to position -21 of all the 70 analysed sequences. Sequences presenting a score greater or equal to 5.10^5 (as defined in the text) were reported in the table.

(*) Rat troponin T gene (6)

(**) Rat fibronectin (8,54). Numbers 1a, 1b and 1c correspond to three alternative splice sites for the first intron of this gene.

Influence of the RNA secondary structure on the splice sites selection

Although our previous analysis can explain the frequent processing of 20 K and GHI.V mRNAs, it does not explain the appearance of the 17.5 K mRNA.

Solnick (55) has shown that frequently used splice sites become optional when sequestered in a hairpin loop. He therefore proposed a strong relationship between the RNA secondary structure and the selection of splice sites. We have searched for a local secondary structure in the regions of the alternative splicing sites of the hGH precursor mRNA. We have found a large hairpin structure trapping both the 22 K and the 20 K acceptor sites, and also the branchpoint sequence of the GHI.V site. The

energy model of Papanicolaou, Gouy and Ninio (56) predicts a relatively high stability for this local structure (free energy = -31.2 Kcal). This predicted RNA folding may provide a way to explain the unusual splicing that leads to the 17.5 K coding mRNA.

Indeed, this stable structure, by looping out the 22 K and 20 K splice sites and by interfering with the lariat formation leading to the GHI.V splice site, would favour the jump splice at the 17.5 K alternative acceptor site.

hGHI.V mRNA has two open reading frames, one for exon I and another for exon V of the hGH gene

Our S1 mapping experiments lead us to detect 4 distinct hGH mRNA variants. The 3 first ones keep all along the same reading frame; the encoded proteins only vary by the deletion of a few amino acids. The last transcript switches reading frame at the deletion location, leading to a truncated hGH (48 aa) and a possible reinitiation methionine just ahead of exon 5. The 22 K and 20 K hormones have been detected and well characterized (10, 11,12). No evidence however is currently available to comment on the *in vivo* reality of the proteins encoded by the two smaller mRNA variants.

Acknowledgements.

This work was supported by grants from IRSIA (Univ-Innov.) and SPPS (Actions concertées, belgian government).

We thank Dr Françoise Rentier from our laboratory for helpful discussions and advices throughout the course of this work.

We are grateful to Jean-Claude Le Hégarat (Gif/Yvette, France) for providing the phage vectors and bacterial strains and for helping in the construction of the library.

We also acknowledge Josiane Lejeune and Nicole Bottelbergs for typing the manuscript.

REFERENCES

1. SCHIBLER, U., HAGENBUCHLE, O., WELLAUER, P.K., and PITTET, A.C. (1983) Cell 33, 501-508.
2. AMARA, S.G., EVANS, R.M., and ROSENFELD, M.G. (1984) Mol.Cell. Biol. 4, 2151-2160.
3. EARLY, P., ROGERS, J., DAVIS, M., CALAME, K., BOND, M., WALL, R., and HOOD, L. (1980) Cell 20, 313-319.

4. CRABTREE, G.R., and KANT, J.A. (1982) Cell 31, 159-166.
5. KING, C.R., and PIATIGORSKY, J. (1983) Cell 32, 707-712.
6. BREITBART, R.E., NGUYEN, H.T., MEDFORD, R.M., DESTREE, A.T., MAHDAVI, V., and NADAL-GINARD, B. (1985) Cell 41, 67-82.
7. de FERRA, F., ENGH, H., HUDSON, L., KAMHOLZ, J., PUCKETT, C., MOLINEAUX, S., and LAZZARINI, R.A. (1985) Cell 43, 721-727.
8. SCHWARZBAUER, J.E., TAMKUN, J.W., LEMISCHKA, I.R. and HYNES, R.O. (1983) Cell 35, 421-431.
9. OATES, E., and HERBERT, E. (1984) J. Biol. Chem. 259, 7421-7425.
10. SINGH, R.N.P., SEAVEY, B.K., and LEWIS, U.J. (1974) Endocrinol. Res. Commun. 1, 449-464.
11. LEWIS, U.J., BONEWALD, L.F., and LEWIS, L.J. (1980) Biochem. Biophys. Res. Commun. 92, 511-516.
12. CHAPMAN, G.E., ROGERS, K.M., BRITTAIN, T., BRADSHAW, R.A., BATES, O.J., TURNER, C., CARY, P.D., and CRANE-ROBINSON, C. (1981) J. Biol. Chem. 256, 2395-2401.
13. SEEBURG, P.H. (1982) DNA 1, 239-249.
14. WALLIS, M. (1980) Nature 284, 512.
15. DE NOTO, F.M., MOORE, D.D., and GOODMAN, H.M. (1981) Nucl. Acids Res. 9, 3719-3730.
16. RUSKIN, B., and GREEN, M.R. (1985) Cell 43, 131-142.
17. FRENDEWEY, D., and KELLER, W. (1985) Cell 42, 355-367.
18. REED, R. and MANIATIS, T. (1985) Cell 41, 95-105.
19. RUSKIN, B., GREENE, J.M., and GREEN, M.R. (1985) Cell 41, 833-844.
20. RUSKIN, B., PIKIELNY, C.W., ROSBACH, M., and GREEN, M.R. (1986) Proc. Natl. Acad. Sci. USA 83, 2022-2026.
21. MARTIAL, J.A., HALLEWELL, R.A., BAXTER, J.D., and GOODMAN, H.M. (1979) Science 205, 602-607.
22. GOEDDEL, D.V., HEYNEKER, H.L., HOZUMI, T., ARENTZEN, R., ITAKURA, K., YANSURA, D.G., ROSS, M.J., MIOZZARI, G., CREA, R., and SEEBURG, P.H. (1979) Nature 281, 544-548.
23. OLSON, K.C., FENNO, J., LIN, N., HARKINS, R.N., SNIDER, C., KOHR, W.H., ROSS, M.J., FODGE, D., PRENDER, G., and STEBBING, N. (1981) Nature 293, 408-411.
24. ROSENFELD, R.G., WILSON, D.M., DOLLAR, L.A., BENNETT, A., and HINTZ, R.L. (1982) J. Clin. Endocrinol. Metab. 54, 1033-1038.
25. GOODMAN, H.M. (1984) Endocrinology 114, 131-135.
26. ADELMAN, J.P., HAYFLICK, J.S., VASSER, M., and SEEBURG, P.H. (1983) DNA 2, 183-193.
27. ULLRICH, A., SHINE, J., CHIRGWIN, J., PICTET, R., TISCHER, E., RUTTER, W.J., and GOODMAN, H.M. (1977) Science 196, 1313-1319.
28. CHIRGWIN, J.M., PRZYBYLA, A.E., MacDONALD, R.J., and RUTTER, W.J. (1979) Biochemistry 18, 5294-5299.
29. AVIV, H., and LEDER, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408-1412.
30. COOKE, N.E., COIT, D., WEINER, R.I., BAXTER, J.D., and MARTIAL, J.A. (1980) J. Biol. Chem. 255, 6502-6510.
31. MURRAY, N.E., BRAMMAR, W.J., and MURRAY, K. (1977) Molec. Gen. Genet. 150, 53-61.
32. LATHE, R., and LECOCQ, J.-P. (1977) Virology 83, 204-206.
33. BENTON, W.D., and DAVIS, R.W. (1972) Science 196, 180-182.
34. THAYER, R.E. (1979) Anal. Biochem. 98, 60-63.

35. MAXAM, A.M., and GILBERT, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
36. SANGER, F., NICKLEN, S., and COULSON, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
37. WEAVER, R.F., and WEISSMANN, C. (1979) Nucl. Acids Res. 7, 1175-1193.
38. KERNIGHAN, B.W., et RITCHIE, D.M. (1985) Masson, Paris.
39. MOUNT, S.M. (1982) Nucl. Acids Res. 10, 459-472.
40. WIERINGA, B., HOFER, E., and WEISSMAN, C. (1984) Cell 37, 915-925.
41. FURDON, P.J., and KOLE, R. (1986) Proc. Natl. Acad. Sci. USA. 83, 927-931.
42. STADEN, R. (1984) Nucl. Acids Res. 12, 505-519.
43. RUSKIN, B., KRAINER, A.R., MANIATIS, T., and GREEN, M.R. (1984) Cell 38, 317-331.
44. ZEITLIN, S., and EFSTRATIADIS, A. (1984) Cell 39, 589-602.
45. KELLER, E.B., and NOON, W.A. (1984) Proc. Natl. Acad. Sci. USA. 81, 7417-7420.
46. LANGFORD, C.J., and GALLWITZ, D. (1983) Cell. 33, 519-527.
47. RODRIGUEZ, J.R., PIKIELNY, C.W., and ROSBASH, M. (1984) Cell 39, 603-610.
48. DOMDEY, H., APOSTOL, B., LIN, R.-J., NEWMAN, A., BRODY, E., and ABELSON, J. (1984) Cell 39, 611-621.
49. HORNIG, H., AEBI, M. and WEISSMAN, C. (1986) Nature 324, 589-591.
50. HARTMUTH, K. and BARTA, A., (1986) CSH, abstract book, RNA Processing Meeting, 21.
51. KOSTYO, J.L., CAMERON, C.M., OLSON, K.C., JONES, A.J.S., PAI, R.C. (1985) Proc. Natl. Acad. Sci. USA 82, 4250-4253.
52. WOYCHIK, R.P., CAMPER, S.A., LYONS, R.H., HOROWITZ, S., GOODWIN, E.C., and ROTTMAN, F.M. (1982) Nucl. Acids Res. 10, 7197-7210.
53. BLACK, D.L., CHABOT, B., and STEITZ, J.A. (1985) Cell 42, 737-750.
54. TAMKUN, J.W., SCWARZBAUER, J.E., and HYNES, R.O. (1984) Proc. Natl. Acad. Sci. USA 81, 5140-5144.
55. SOLNICK, D. (1985) Cell 43, 667-676.
56. PAPANICOLAOU, C., GOUY, M., and NINIO, J. (1984) Nucl. Acids Res. 12, 31-44.