View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Cell, Vol. 91, 875–879, December 26, 1997, Copyright ©1997 by Cell Press

Classification of Introns: U2-Type or U12-Type

Minireview

Phillip A. Sharp and Christopher B. Burge
Center for Cancer Research
and Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139-4307

The highly conserved dinucleotides GT and AG at intron 5′ and 3′ boundaries, respectively, were recognized with the sequencing of the first introns. These consensus dinucleotides were almost universal, suggesting that introns were of a common origin and excised from pre-mRNA by a common apparatus, the spliceosome, which is formed through the interactions of U1, U2, U4/U6, and U5 snRNPs. Some 20 years later, exceptions to this major class of consensus dinucleotides identified a subclass having AT and AC at the 5′ and 3′ boundaries of introns, respectively (Jackson, 1991; Hall and Padgett, 1994). We recently learned that some members of this subclass of introns, initially named AT–AC, are excised by a novel spliceosome composed of snRNPs U11, U12, U4atac/U6atac, and U5 (Hall and Padgett, 1996; Tarn and Steitz, 1996a, 1996b; Kolossova and Padgett, 1997). Thus, it appeared that spliceosomes and introns were of two types, GT–AG and AT–AC, as defined by the terminal dinucleotides of the intron, suggesting a binary division of introns. However, these simple Newtonian-type rules must now be revised since some introns with GT and AG boundary sequences are spliced by the U12-type spliceosome, and conversely, some introns with AT and AC boundary sequences are spliced by the U2-type spliceosome (Dietrich et al., 1997 [the December issue of *Molecular Cell*]). Happily, hybrid introns that hypothetically would be excised by a chimeric spliceosome composed of a mixture of snRNPs from the two snRNP families have not been identified. Hence, as will be discussed in more detail below, any single intron is of one class or the other dependent upon the type of spliceosome formed. This strongly indicates that introns should be identified by the type of spliceosome, either U2-type or U12-type, and not on the basis of their terminal sequences (Dietrich et al., 1997).

Examples of a subset of introns with boundary sequences AT–AC that are probably spliced by the U12 system are listed in Table 1 (adapted from Tarn and Steitz, 1997) and occur in the total population of introns at a frequency of roughly 1/5,000 to 1/10,000. This class of introns is found in plants, Drosophila, and vertebrate genes, and thus the U12 spliceosome responsible for their excision has coexisted with the predominant U2 spliceosome process for at least a billion years. Furthermore, these AT–AC introns coexist within genes with introns excised by the major U2 spliceosome. The AT–AC introns that are processed by the U12 spliceosome all contain a strongly conserved consensus sequence at their 5′ splice site and branch site (see Table 2). In fact, these sequences are almost invariant among the known examples, suggesting that the processing of this subclass of introns is largely defined by recognition of these sequences by U11 and U12 snRNAs. The structure of the 3′ splice site is also distinct for this class of introns. There is typically no polypyrimidine tract and a relatively small number of nucleotides, 10 to 16, separating the branch site from the 3′ splice site. Interestingly, as has been noted by others (Tarn and Steitz, 1997), this intron 3′ end structure is quite similar to that of self-splicing group II introns, which commonly have AC at their 3′ boundary and 7 to 8 nucleotides separating the 3′ splice site from the branch site. This similarity might indicate that the U12-type spliceosome may be more closely related to group II introns than the more common U2-type spliceosome.

The original classification of the two types of introns based on their terminal dinucleotide sequences has recently been shown to be ambiguous. A more natural classification is based on the type of spliceosome active in excision of the introns, since not all introns that are excised by the U12-type spliceosome have the characteristic AT–AC sequence at their boundaries (Dietrich et al., 1997). For example, the human calcium channel genes CACNL1A2 and CACNL1A3 have GT–AG sequences at their boundaries but otherwise have the conserved internal sequences typical of recognition by U12-type spliceosomes at their 5′ splice site and branch site (see Wu and Krainer, 1997; Table 2). Three other introns of this type have previously been identified in the genes for: SmE protein, a component of the core proteins of snRNP particles; ADPRP, the ADP-ribose polymerase that modifies nuclear proteins; and human c-Raf-1, a second messenger signaling kinase. As might be expected from the presence of the internal conserved sequences, experiments in vitro showed that the splicing of the ADPRP intron was blocked by addition of antisense oligonucleotides to U12 snRNA but not by similar antisense oligonucleotides to U2 snRNA. How common are introns with GT–AG terminal dinucleotides that are processed by the U12 spliceosome as compared to the U2 spliceosome? Though this question will require further study, Dietrich et al. (1997) observed three such introns in a dataset of about 1800 introns, giving a frequency of about 1/600 introns, approximately 10-fold higher than the frequency of U12-type introns with AT–AC termini. Consistent with this relatively high estimated frequency, we identified dozens of additional examples by a straightforward computer search of the GenBank sequence database (Release 103, 1997) for introns matching the 5′ splice site and branch site consensus patterns expected of a U12-type (with at most one or two discrepancies). Twenty-three such sequences (in addition to the five identified previously) are shown in Table 2. One of the most interesting examples found is the last intron (number 66) of the Huntington's disease gene, Huntingtin.

The converse, introns with AT–AC boundaries that are processed by the U2-dependent spliceosome (Wu and Krainer, 1996, 1997; Dietrich et al., 1997), has also been found in some genes (Table 2). However, it has been known for a decade that the U2-type spliceosome would excise introns with AT–AC termini. Intron RNAs with

Table 1. Sample of Genes Containing U12-Type AT–AC Introns

Gene	Organism(s)	Description
ArLIM15	Arabidopsis	RecA-like protein
AtG5	Arabidopsis	Putative transmembrane protein (SAC1-like)
BAP	Human, mouse	B-cell receptor associated protein
CACNL1A4	Human	P/Q type calcium channel, α1 subunit
CDK5	Mouse	Cyclin-dependent kinase
CMP	Human, chicken	Cartilage matrix protein
E2F family	Human, mouse	Family of cell cycle regulatory proteins
GT335	Human	Unknown function
HPS	Human, mouse	Hermansky-Pudlak syndrome protein
P120	Mammals	Proliferating cell nucleolar protein
Prospero	Drosophila	Homeobox protein involved in neurogenesis
REP-3	Mouse	Similar to DNA mismatch-repair proteins
SCN4A	Human	Skeletal muscle sodium channel α subunit
SCN5A	Human	Cardiac muscle sodium channel α subunit
SCN8A	Human, mouse	Brain sodium channel α subunit
TFIIS.oA	Xenopus	Transcription elongation factor
XPG	Mouse	DNA excision repair gene (defective in XP)

A sample of genes containing introns that match the U12-type 5′ splice site and branch site consensus patterns and have AT–AC termini is shown. Previous compilations of such introns were shown in Jackson (1991), Hall and Padgett (1994), Tarn and Steitz (1997), and Wu and Krainer (1997). References for all genes can be found in Wu and Krainer (1997, see Table 1), except BAP. The human and mouse homologs of this gene are described in Ansari-Lari et al. (1997) and M. A. Ansari-Lari et al., (unpublished data), respectively.

the /GU sequence at the 5′ splice site mutated to /AU typically arrest in splicing by the U2-type spliceosome at the lariat intermediate. Mutations of the 3′ splice site of the same intron from -AG/ to -AC/ allow active splicing by the U2-type spliceosome. This observation that the U2 spliceosome has good activity on AT–AC termini as contrasted to other nonconsensus pairs has been canonized in the hypothesis that there is direct base–base recognition between the 5′ splice site and 3′ splice site. In this scheme, alternative base-pairing schemes between G-G at the termini of the wild-type intron and A-C at the termini of the mutant intron have interchangeable functions in the splicing process (Parker and Siliciano, 1993). This conclusion has been challenged on the basis that available data do not strongly support a direct base–base pairing recognition (Luukkonen and Seraphin, 1997). An alternative explanation for the high activity of introns containing AT–AC termini in splicing by the U2-type spliceosome is the potential benefit of some overlapping splicing activity by both types of spliceosomes. Hypothetically, introns with AT–AC boundaries could be spliced by either the U2- or U12-type spliceosomes assuming sufficiently active internal consensus sequences for recognition by both systems. This explanation for the high activity of the AT–AC terminal sequences by the U2 spliceosome is not necessarily inconsistent with the terminal base–base pairing scheme mentioned above but weakens the case by providing another reason for the high splicing activity of these sequences.

Even though a small number of mutations can potentially switch introns as substrates between U2- or U12-dependent spliceosomes (see below), for some introns such switching is very rare. As evident from Table 1, some U12-type AT–AC introns are evolutionarily stable. For example, homologous introns of this class are present in the cartilage matrix protein genes of both humans and chickens, indicating stability over at least 200 million years. Similarly, two homologous introns that are U12-type and have AT–AC sequences are found in the human sodium channel genes expressed in skeletal muscle (SCN4A) or cardiac muscle (SCN5A) (see Table 2). The same two genes also contain homologous introns that are U2-dependent and have the atypical termini of AT–AC (Table 2). Two homologous introns have the same termini in the puffer fish Fugu rubripes (Table 2), and thus this atypical intron structure and spliceosome relationship has been stable for over 400 million years of evolution. However, a homologous intron in the sea slug Aplysia california has the GT–AG boundaries more typical of U2-dependent splicing. This suggests that during the 600 million years of evolution that separates these species there has been a conversion of the terminal sequences (Dietrich et al., 1997).

It is interesting to speculate that the coexistence of the major U2-dependent spliceosome and the rarer U12-dependent spliceosome might represent an old path for the evolution of introns (see below). As mentioned previously, the internal consensus sequences for U12-dependent introns are highly conserved and have distinct features that are similar to those of the self-splicing group II introns, including a frequent AC/3′ boundary, the lack of a polypyrimidine tract upstream of the 3′ splice site, and a short tract of nucleotides between the branch site and the 3′ splice site. Thus, it could be conjectured that at an earlier time the U12-dependent spliceosome might have recognized and excised introns generated by new insertions of group II self-splicing introns. The next step in the evolution of these introns would be the accumulation of mutations that would convert their splicing to the major U2-dependent spliceosome process. This type of evolution has been experimentally directed by introduction of splice-site mutations and analysis of splicing in vivo (Figure 1). The sequence of the P120 intron F (a) is a prototypical U12-type with AU-AC terminal dinucleotides. The predominant splice excising the intron is indicated by the angled

Table 2. Consensus Sequences of U12-type and U2-type Introns

<table>
<tr><th>Gene</th><th>Intron</th><th>5′ Splice Site</th><th>Putative Branch Site</th><th>3′ Splice Site</th><th>Dist.</th></tr>
<tr><td>U12-Type AT–AC Introns[a]</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Arabidopsis AtG5</td><td>7</td><td>-aa|atatcctt-----</td><td>----atattaaccaa</td><td>attaaccaaggcttaa|gt-</td><td>12</td></tr>
<tr><td>Arabidopsis ArLIM15</td><td>14</td><td>gcc|atatccttttaac</td><td>tttttccttaactga</td><td>taactgaagaaatcac|agt</td><td>14</td></tr>
<tr><td>Drosophila Prospero</td><td>2</td><td>-ct|atatcctt-----</td><td>--aatccttgactcc</td><td>cttgactcctttgcac|tc-</td><td>12</td></tr>
<tr><td>Xenopus TFIIS.oA</td><td>6</td><td>aag|atatcctttttat</td><td>atattccttaaatcc</td><td>ttaaatccctttgcac|cg-</td><td>13</td></tr>
<tr><td>Chicken CMP</td><td>7</td><td>aat|atatcctttggaa</td><td>tttctccttaactct</td><td>cttaactctcactcac|tgg</td><td>12</td></tr>
<tr><td>Mouse BAP</td><td>6</td><td>atg|atatccttgtgtt</td><td>agaacccttaaccct</td><td>aaccctgtctgctcac|ctt</td><td>15</td></tr>
<tr><td>Mouse HPS</td><td>15</td><td>gtc|atatcctctgcct</td><td>aagccccttaactca</td><td>taactcaggcctgcac|cta</td><td>14</td></tr>
<tr><td>Mouse REP-3</td><td>6</td><td>aag|atatccttttagg</td><td>tttttctttaatcat</td><td>ctttaatcattactac|atg</td><td>11</td></tr>
<tr><td>Mouse SCN8A</td><td>2</td><td>ttc|atatccttttctg</td><td>tcgccccttaactcc</td><td>cttaactcctctctac|agt</td><td>12</td></tr>
<tr><td>Mouse XPG</td><td>13</td><td>aga|atatcctttcttt</td><td>ataaaccttaactgc</td><td>ccttaactgccagcac|at-</td><td>11</td></tr>
<tr><td>Human BAP</td><td>6</td><td>atg|atatccttctgct</td><td>gtgtaccttaaccct</td><td>ccttaaccctcctcac|ctt</td><td>11</td></tr>
<tr><td>Human CACNL1A4</td><td>1</td><td>gcc|atatcctttttgcc</td><td>tttttccttaattcc</td><td>cttaattccccaatac|tcc</td><td>12</td></tr>
<tr><td>Human CMP</td><td>7</td><td>aac|atatcccttccca</td><td>cctctccttaactct</td><td>ttaactctgagtccac|tgg</td><td>13</td></tr>
<tr><td>Human E2F1</td><td>4</td><td>gcg|atatccttggatt</td><td>gtggtccttgactct</td><td>ccttgactctgccaac|cct</td><td>11</td></tr>
<tr><td>Human E2F4</td><td>3</td><td>-tg|atatcctc-----</td><td>--tctccttaacccc</td><td>tccttaaccccacac|ct-</td><td>10</td></tr>
<tr><td>Human GT335</td><td>6</td><td>gtc|atatccttcctgg</td><td>tttttccgtaactct</td><td>ccgtaactctgcacac|gaa</td><td>11</td></tr>
<tr><td>Human P120</td><td>6</td><td>agg|atatccttgcagg</td><td>cagttccttaacagc</td><td>tccttaacaggcccac|atg</td><td>10</td></tr>
<tr><td>Human SCN4A</td><td>2</td><td>tgc|atatcctgcccaa</td><td>actttccttgaccct</td><td>cttgaccctgccccac|gct</td><td>12</td></tr>
<tr><td>Human SCN5A</td><td>3</td><td>ctc|atatcctttgcag</td><td>gctttccttgacccc</td><td>ccttgaccccacgcac|gct</td><td>11</td></tr>
<tr><td>Human SCN8A</td><td>2</td><td>---|atatcctt-----</td><td>----cccttgactct</td><td>cttgactcttcyctac|---</td><td>12</td></tr>
<tr><td>Consensus</td><td>—</td><td>ooo|ATATCCTTToooo</td><td>oooTTCCTTRACYCY</td><td>oooooooooooooYAC|ooo</td><td>10–15</td></tr>
<tr><td>Conservation</td><td>—</td><td>---|+++++++$^{95}{}_{85}{}^{69}$----</td><td>---$^{50}{}_{65}{}^{95}{}_{90}{}^{95}$+++$^{85}{}_{95}{}^{75}{}_{85}$</td><td>-------------$_{95}$+$_{95}$|---</td><td>—</td></tr>
<tr><td>U2-Type AT–AC Introns[b]</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>A. nidulans xlnC</td><td>8</td><td>agg|ataagtttttcc</td><td>cgtcacactgacaac</td><td>actgacaaccccaaac|agc</td><td>n/a</td></tr>
<tr><td>A. tubingensis xlnC</td><td>7</td><td>-gg|ataagtac-----</td><td>--gatagctgacagc</td><td>gctgacagccccttac|ag-</td><td>n/a</td></tr>
<tr><td>P. chrysogenum xylP</td><td>7</td><td>-gg|ataagtac-----</td><td>--actagctaacagc</td><td>gctaacagcctctcac|ag-</td><td>n/a</td></tr>
<tr><td>Fugu-1 SCN</td><td>n/a</td><td>agg|ataagacc-----</td><td>---------------</td><td>gttctgggctgtttac|tta</td><td>n/a</td></tr>
<tr><td>Fugu-2 SCN</td><td>n/a</td><td>aag|ataagaac-----</td><td>---------------</td><td>gcgcatgtccctttac|ttt</td><td>n/a</td></tr>
<tr><td>Human SCN4A</td><td>21</td><td>aag|atgagtat-----</td><td>--tcaacctgac---</td><td>ccctatcccactatac|tta</td><td>n/a</td></tr>
<tr><td>Human SCN5A</td><td>25</td><td>aag|atacgtag-----</td><td>--ggcctctgag---</td><td>tgccttctctttgcac|tta</td><td>n/a</td></tr>
<tr><td>Consensus</td><td>—</td><td>ARG|ATAAGToooooooo</td><td>unknown</td><td>oooooooooooooYAC|ooo</td><td>n/a</td></tr>
<tr><td>Conservation</td><td>—</td><td>+++|++$^{86}{}_{86}$+$_{71}$-------</td><td>unknown</td><td>-------------$_{86}$++|---</td><td>n/a</td></tr>
<tr><td>U12-type GT–AG Introns[c]</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Arabidopsis ArLIM15</td><td>9</td><td>cct|gtatccttaatat</td><td>tttttccttaacccc</td><td>cccctaaattttgtag|ccg</td><td>17</td></tr>
<tr><td>Arabidopsis Luminidp.</td><td>10</td><td>cgt|gtatcctttcgtt</td><td>agtttccttagcttt</td><td>ccttagctttcagcag|ata</td><td>12</td></tr>
<tr><td>Arabidopsis Villin</td><td>1</td><td>acc|gtatcctttcagc</td><td>gtgaaccttaacttt</td><td>taacttttgttattag|agg</td><td>14</td></tr>
<tr><td>Arabidopsis VP1</td><td>4</td><td>aaa|gtatcctttattt</td><td>gtttatcttaacaaa</td><td>tcttaacaaattctag|ttt</td><td>11</td></tr>
<tr><td>Bombyx XDH</td><td>2</td><td>tat|gtatcctttactt</td><td>tactttcttaacatt</td><td>cttaacattttaccag|tca</td><td>12</td></tr>
<tr><td>Fugu Huntingtin</td><td>66</td><td>tct|gtatccttatgag</td><td>ctcttccttaactcc</td><td>ccccgttctgaaccag|gct</td><td>20</td></tr>
<tr><td>Xenopus RPL1a</td><td>3</td><td>aac|gtatcctttggg</td><td>attttccttaatatt</td><td>ttaatatttcatgtag|atg</td><td>13</td></tr>
<tr><td>Hamster PLCD1</td><td>3</td><td>aca|gtatccttttggg</td><td>tgttttcctaactgc</td><td>ctaactgcttgcccag|ctg</td><td>13</td></tr>
<tr><td>Mouse ERK2</td><td>2</td><td>tgt|gtatcctttttcta</td><td>tttgaccttaacctc</td><td>acctcctgtcacacag|ata</td><td>16</td></tr>
<tr><td>Mouse PBGD</td><td>7</td><td>att|gtatcctttcaga</td><td>cttctccttagcaac</td><td>tagcaacgctccacag|cgg</td><td>15</td></tr>
<tr><td>Mouse TCP1</td><td>5</td><td>aat|gtatccttgaatg</td><td>gctatccttaactgt</td><td>ttaactgttgtctcag|aaa</td><td>13</td></tr>
<tr><td>Mouse MHC P35B</td><td>8</td><td>act|gtatcctttggct</td><td>tatggtcttgaccag</td><td>ttgaccaggcatccag|ttt</td><td>13</td></tr>
<tr><td>Human ADPRP</td><td>n/a</td><td>cga|gtatcctttcttg</td><td>tggctccttaacaag</td><td>acaagcttcccctcag|gta</td><td>16</td></tr>
<tr><td>Human AOX1</td><td>2</td><td>agc|gtatcctttcttt</td><td>atgatccttaactat</td><td>actatacctcttccag|ttc</td><td>16</td></tr>
<tr><td>Human CACNL1A1</td><td>1</td><td>gaa|gtatccttc----</td><td>aacttccttgactcc</td><td>tgactcccttttctcag|acc</td><td>14</td></tr>
<tr><td>Human CACNL1A2</td><td>1</td><td>gaa|gtatcctttttttt</td><td>atataccttaacaca</td><td>taacacatttttcag|acc</td><td>14</td></tr>
<tr><td>Human CACNL1A3</td><td>1</td><td>gaa|gtatccttcaggg</td><td>gctctccttaaccct</td><td>cttaaccctgctccag|gcc</td><td>12</td></tr>
<tr><td>Human CANP</td><td>6</td><td>gat|gtatccttggggg</td><td>actcttcttaacacc</td><td>aacaccctcccaccag|agc</td><td>15</td></tr>
<tr><td>Human CLCN6</td><td>5</td><td>cat|gtatcctttttcac</td><td>agtttccttaaccca</td><td>aacccagtttccgcag|cgg</td><td>15</td></tr>
<tr><td>Human c-Raf-1</td><td>1</td><td>aca|gtatcct------</td><td>-------ttaacaag</td><td>-ttaacaagcattgag|ata</td><td>12</td></tr>
<tr><td>Human FHIT</td><td>5</td><td>gac|gtatcctttcttt</td><td>atgggccttaacaga</td><td>acagaccttgctacag|atg</td><td>16</td></tr>
<tr><td>Human PBGD</td><td>8</td><td>att|gtatcctttttaga</td><td>gtggtccttagcaac</td><td>tagcaactctccacag|cgg</td><td>15</td></tr>
<tr><td>Human Huntingtin</td><td>66</td><td>gat|gtatcctctctgg</td><td>tttttccttaactcc</td><td>ttaactcctgcaccag|cct</td><td>13</td></tr>
<tr><td>Human INSIG1</td><td>2</td><td>gct|gtatccttaattt</td><td>cttttccttaacttt</td><td>acttttatatcaccag|aaa</td><td>16</td></tr>
<tr><td>Human Myosin HC</td><td>5</td><td>cct|gtatcctttccct</td><td>agaatccctaactcc</td><td>actccctccaaaccag|atc</td><td>16</td></tr>
<tr><td>Human PLCB3</td><td>16</td><td>gga|gtatcctttgaag</td><td>gtggcccctgaccac</td><td>tgaccaccaatctcag|ata</td><td>14</td></tr>
<tr><td>Human PTEN</td><td>1</td><td>cct|gtatccatttctg</td><td>tctttccttaactaa</td><td>taactaaagtactcag|ata</td><td>14</td></tr>
<tr><td>Human Sm E</td><td>1</td><td>atc|gtatcctacgcag</td><td>tttttccttagccac</td><td>agccactgtggtgcag|aac</td><td>16</td></tr>
<tr><td>Consensus</td><td>—</td><td>ooT|GTATCCTTToooo</td><td>oooTTCCTTAACYoY</td><td>oooooooooooooYAG|Aoo</td><td>11–20</td></tr>
<tr><td>Conservation</td><td>—</td><td>--$_{54}$|+++++++$^{96}{}_{93}{}^{70}$----</td><td>---$^{52}{}_{74}{}^{81}{}_{96}{}^{89}$+$^{89}{}_{86}{}^{96}{}_{68}$-^{71}</td><td>-------------$_{96}$++|$_{54}$--</td><td>—</td></tr>
<tr><td>U2-Type GT–AG Introns[d]</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Consensus</td><td>—</td><td>KAG|GTRAGToooooooo</td><td>oooooooCTRAYooo</td><td>YYYYYYYYYYYYoYAG|Goo</td><td>10–50+</td></tr>
<tr><td>Conservation</td><td>—</td><td>$^{70}{}_{60}{}^{81}$|++$^{94}{}_{71}{}^{84}{}_{46}$-------</td><td>-------$^{76}{}_{91}{}^{81}{}_{99}{}^{83}$---</td><td>$^{71}{}_{76}{}^{78}{}_{79}{}^{80}{}_{80}{}^{80}{}_{80}{}^{80}{}_{82}{}^{88}{}_{87}$-$_{96}$++|$_{52}$--</td><td>—</td></tr>
</table>

(*Table 2 continued on next page*)

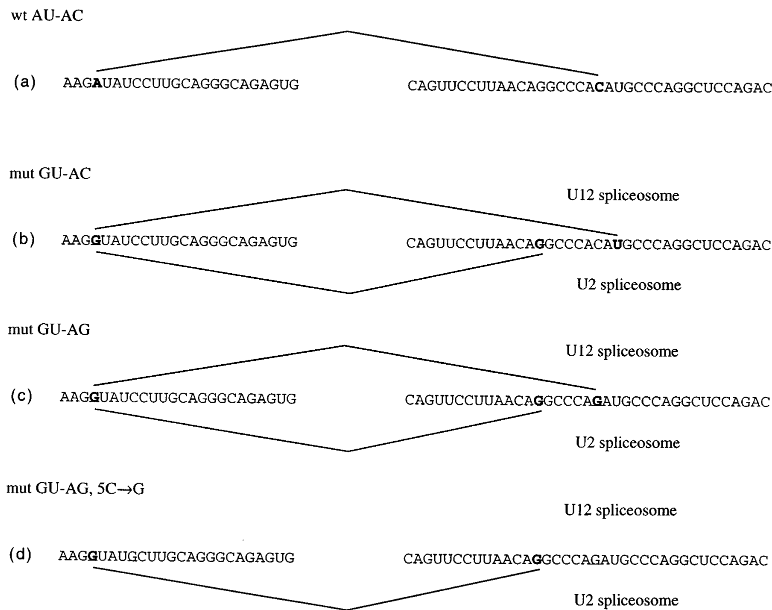

Figure 1. Mutations in the Terminal Nucleotide Sequences of Human P120 Intron F Activate Splicing by the U2 Spliceosome

In the study of Dietrich et al. (1997), the splicing of mutants of the P120 intron was analyzed after transfection into CHO cells. Intron sequences, excised by the U12 spliceosome and U2 spliceosome, are shown above and below the sequences, respectively. Several alternative splices of both types for each set of intron sequences were characterized but are not illustrated here. Mutant sites are underlined and the specific bases utilized for splicing are in bold. Figure modified from Dietrich et al., 1997.

line above the sequence. Mutation of the A at the 5′ splice site to a G (b) activated splicing through the U2-spliceosome to an AG sequence upstream of the wild-type 3′ terminus AC. At the same time, the intron remained active for splicing by the U12-spliceosome using a 3′ A**U** terminus two nucleotides downstream of the normal AC. This situation changed slightly when the double mutant was tested where the termini had been mutated from AU–AC to GU–AG (c). The predominant U2-dependent splice remained unchanged, but the 3′ boundary of the U12-dependent splice became the mutant AG site. Further mutation of position +5 C of the 5′ splice site to G (d) suppressed the U12-dependent reaction, with the excision of intron sequences totally dependent upon the U2-spliceosome utilizing the cryptic 3′ AG. Note that in this experimentally observed scheme, splicing by the U2- and the U12-dependent spliceosomes was simultaneously active for two of the intermediate mutants suggesting that some functional protein might be produced, perhaps permitting the viability of an organism with these mutations. Furthermore, almost certainly by coincidence, the cryptic 3′ splice site utilized by the U2-dependent spliceosome is in triplet phase with the wild-type 3′ splice site indicating that the mRNA generated by this reaction would retain an open reading frame and generate a protein with an insertion of two additional amino acids. Obviously, such a protein could retain some activity permitting the complete conversion of a U12-intron to a U2-intron.

The potential conversion of introns between the U12

(*Table 2 continued from previous page*)

The symbol (|) indicates a splice junction, (-) indicates missing sequence data. The second column lists the number of the intron shown in the given gene (n/a = not available). The last column lists the distance from the presumed branch site to the 3′ splice junction in nucleotides (n/a = not available): putative branch site adenosines are in bold for U12-type introns in the preceding two columns. Consensus nucleotides for each subset are indicated in capital letters: R indicates either purine (A or G), Y indicates either pyrimidine (C or T), K indicates A or C, and (○) indicates that no clear consensus is present. The degree of conservation of each consensus is indicated immediately below the corresponding letter: (+) indicates conservation in all known sites (exception: a few U2-type GC–AG introns are known), numbers indicate that the consensus is matched in the given percentage of available sites, (-) indicates the absence of strong conservation.

[a] Sequence data for genes AtG5, Prospero, TFIIS, and E2F4 are from Wu and Krainer, 1997. The sequence for human SCN8A is from Dietrich et al., 1997. All other sequence data are from GenBank Release 103 (1997). GenBank accession numbers are given in parentheses following the gene name: ArLIM15 (D45415), chicken CMP (X12352, X12353), mouse BAP (AC002397), HPS (U78954), REP-3 (L10300, L10301), mouse SCN8A (U59963, U59964), XPG (U3982, U3983), human BAP (U47924), CACNL1A4 (Z80114), human CMP (M55675), E2F1 (U47677), GT335 (U53004), P120 (M33132), SCN4A (L04216), SCN5A (AF007781).

[b] Sequence data for A. nidulans xlnC from GenBank Release 103 (1997) accession no. Z49894. Sequence data for A. tubingensis xlnC and P. chrysogenum xylP from Wu and Krainer, 1997. All other sequences from Dietrich et al., 1997.

[c] Sequence data for ADPRP, CACNL1A1, and Sm E from Dietrich et al., 1997. All other sequence data from GenBank Release 103, 1997. GenBank accession numbers are given in parentheses following the gene name: ArLIM15 (D45415), Luminidependens (AC002330), Villin (AC002339), VP1 (U93215), XDH (AB005911), Fugu Huntingtin (X82939), RPL1a = Ribosomal Protein L1a (X06552), PLCD1 = Chinese hamster Phospholipase C-δ-1 (U50565, U50566), ERK2 (D87265, D87266), mouse PBGD (M28664, M28665), TCP1 (D10606), MHC P35B (M30128), AOX1 (AF009442, AF009443), CACNL1A2 (D43706, D43707), CACNL1A3 (U30666, U30667, U30668), CANP (M3150617), CLCN6 (AF00925011), c-Raf-1 (M11376), FHIT (U76267, U76268), human PBGD (M95623), human Huntingtin (L27415, L27416), INSIG1 (U96876), Myosin heavy chain (AF001548), PLCB3 = Phospholipase C-β-3 (U26425), PTEN (AF000726), Sm E (M21253).

[d] 5′ splice site data for 1254 nonredundant human introns from Burge and Karlin, 1997. Branch site data for plant, mammal, and Drosophila from Senapathy et al., 1990. 3′ splice site data for 1254 nonredundant human introns from Burge, 1997.

and U2 spliceosomes is probably unidirectional. As shown at the bottom of Table 2, the consensus sequences for the highly abundant U2-dependent introns are much less constrained than those of U12-dependent introns, suggesting that conversion from a U2-type to a U12-type pattern would represent an extremely improbable event (Dietrich et al., 1997). This observation also provides a possible explanation for why the vast majority of introns are spliced by the U2-dependent reaction. Furthermore, cells containing one predominant type of intron and splicing pathway would likely have certain advantages over cells containing two or more equally abundant types of introns. In the latter case, nonhomologous DNA recombination events might frequently generate defective introns by, for example, juxtaposing a 5′ splice site of the U12-type from one gene across an intron with a 3′ splice site of the U2-type from another gene. In contrast, if the two splice sites bracketing the new intron created by this recombination event were both predominantly of U2-type, then the intron would often be precisely excised, perhaps creating a novel gene. This would obviously be advantageous for the evolution of new genes and might account for the strong predominance of one type of spliceosome. The U12-type spliceosome is probably retained in most eukaryotic cells to excise the few remaining introns of this class. The yeast Saccharomyces cerevisiae, with its less complex genome and cellular structure, has probably discarded all U12-dependent introns and the corresponding spliceosome components.

Further study of the U12 spliceosome and the biology of introns excised by it could provide insights into many important questions. The phylogenetic distribution of the U12 spliceosome might indicate the origin of the components of both types of spliceosomes and their potential relationship to group II–type self-splicing introns. Although not apparent yet, the splicing of U12-type introns might be regulated as a subclass differently than U2-type introns. Finally, elucidation of the interactions between splicing factors such as the PRP proteins (which are essential for U2 spliceosome formation) and the SR proteins (which facilitate formation of the U2 spliceosome) with components of the U12 splicesome will be very interesting.

Selected Reading

Ansari-Lari, M.A., Shen, Y., Muzny, D.M., Lee, W., and Gibbs, R.A. (1997). Genome Res. *7*, 268–280.

Burge, C.B. (1997). Computational Biology: Machine Learning and Pattern Analysis Methods, In Salzberg, S., Searls, D., and Kasif, S., eds. (London: Reed Elsevier), in press.

Burge, C.B., and Karlin, S. (1997). J. Mol. Biol. *268*, 78–94.

Dietrich, R.C., Incorvaia, R., and Padgett, R.A. (1997). Molecular Cell, in press.

Hall, S.L., and Padgett, R.A. (1994). J. Mol. Biol. *239*, 357–365.

Hall, S.L., and Padgett, R.A. (1996). Science *271*, 1716–1718.

Jackson, I.J. (1991). Nucleic Acids Res. *19*, 3795–3798.

Kolossova, I., and Padgett, R.A. (1997). RNA *3*, 227–233.

Luukkonen, B.G.M., and Seraphin, B. (1997). EMBO J. *16*, 779–792.

Parker, R., and Siliciano, P.G. (1993). Nature *361*, 660–662.

Senapathy, P., Shapiro, M.B., and Harris, N.L. (1990). Meth. Enzymol. *183*, 252–278.

Tarn, W.-Y., and Steitz, J.A. (1996a). Cell *84*, 1–20.

Tarn, W.-Y., and Steitz, J.A. (1996b). Science *273*, 1824–1832.

Tarn, W.-Y., and Steitz, J.A. (1997). Trends Biochem. Sci. *22*, 132–137.

Wu, Q., and Krainer, A.R. (1996). Science *274*, 1005–1008.

Wu, Q., and Krainer, A.R. (1997). RNA *3*, 586–601.