Arabidopsis Intron Splice-Site Exceptions

This page gives information on exceptions to Arabidopsis splice-site
consensus sequences and was kindly supplied by
Gordon Simpson,
The John Innes Centre, Norwich, United Kingdom.

Gordon is **collating and updating** Arabidopsis exceptions to the GT-AG rule
and would be grateful if you could inform him of any that you find.

UPDATED ON 9 OCTOBER, 1997

The splice site sequences of the vast majority of introns in
Arabidopsis mRNA genes fit the canonical GT-AG
- U2-dependent consensus borders (where GT refers to the terminal
residues of the intron at the 5' splice site and AG refers to the
terminal residues of the intron at the 3' splice site). However,
there are rare examples of introns possessing distinct non-canonical
splice site sequences. If an Arabidopsis gene you have sequenced does
not fit the U2-dependent
GT-AG class you may have found
something novel and important - the following information is a pointer
to exceptions.

Two intron classes processed by two classes of spliceosome

Introns are excised from pre-mRNA by a dynamic multicomponent complex
called the spliceosome. It has recently emerged that there are at
least two compositionally and functionally distinct classes of
spliceosome: the **major class** processes **U2-dependent
introns,** while the recently identified **minor class**
processes non-canonical **U12-dependent introns**. This
terminology derives from UsnRNA components (U2 and U12 snRNA) that are
specific components of each spliceosome class.

Background

The importance of the identity of the terminal intron residues

Until very recently almost all research into pre-mRNA splicing had
centred on U2-dependent introns possessing the canonical GT-AG splice
sites - this is not surprising, since these are far and away the most
frequently occuring intron class. The terminal G residues of such
introns play a key role in the splicing reaction. When either G is
mutated, splicing to these sites is blocked. However, splicing can be
restored to some extent if particular mutations are made at both
terminal positions: A-C and A-A terminal residues work the best.
This suppression of mutant defect is indicative of either a direct
non-Watson-Crick interaction between the terminal residues taking

place within the spliceosome or the action of an unknown spliceosomal
component(s) sensitive to the identity of both terminal residues.

*Previous reference to **AT-AC class** introns and unified view*
of two mechanistic classes

Recent work has led to the rapid characterisation of the processing of
a class of introns with non-canonical splice sites - widely referred
to as **AT-AC** class - that possess AT-AC borders in place of the
more typical GT-AG and which, in addition, possess sequence
distinctions in the wider splice site and branch site consensus
sequences. These introns are processed by the **U12-containing
minor class spliceosome,** but the overall reaction mechanism
resembles that of the major class. Further characterisation has now
revealed that the AT-AC terminology is misleading - rare, but
naturally occuring introns with divergent terminal residues have been
identified. The specificity in their processing is governed by the
wider splice site and branch site consensus sequences and not by the
identity of the terminal residues.

Table 1: Consensus sequences for Intron classes

	5' splice site	branch site	3' splice site
Mammalian U2-dependent	GURAGU	UNCURAC	(Yn) YAG
	GC		
	A		C
U12-dependent	GUAUCCUU	UCCUUAAC	YCCAG
	A		C
	A		G

Table 1. The consensus sequences derived for splice site and branch
  sites in U2 and U12-dependent introns. Alternative terminal residues
  that have been found in naturally occuring animal genes are included.

Analysis of the small number of U12-dependent introns described
indicates that the branch site may normally be closer (9-19nt) to the
3' splice site than in U2-dependent introns (typically
18-40nt), but exceptions may exist.

Non canonical Arabidopsis introns identified so far

Most introns in Arabidopsis are
U2-dependent and possess GT-AG

borders.   Some U2-dependent introns possess the sequence GC-AG -
this is the most frequently identified non-canonical site in
Arabidopsis. Introns bearing the properties of the non-canonical U12-dependent
class, that possess a variety of terminal residues, are also
present in Arabidopsis.

Table 2: Candidate Arabidopsis U12-dependent introns

Termini	Intron	Accession
A-A	intron 7	[Wu *et al.* (1996) *Nat Genet* **14**, 383-384]
A-C	intron 6	U53856
	intron 14	D45415
	intron 14	U76670
G-G	intron 9	U76670
	intron 9	D45415

Further non-canonical splice sites?

While U2-dependent and U12-dependent introns are processed in distinct
spliceosomes, the overall reaction mechanism is similar. In contrast,
the processing of an intron with non-canonical splice sites in the
HAC1 gene of yeast proceeds via a novel mechanism. HAC1 is involved
in the Unfolded Protein Response (UPR) of yeast. Splicing is
initiated upon activation of Ire1p, a transmembrane kinase that lies
in the ER and/or inner nuclear membrane. Ire1p is also a site specific
endoribonuclease that cleaves HAC1 mRNA specifically at both splice
junctions. Cleaved ends can be ligated by tRNA ligase (Sidrauski
C, Walter (1997) Cell 90: 1031-1039). The distribution of this
kind of pre-mRNA processing in nature is presently unknown. We are
currently investigating the alternative splicing of an Arabidopsis
intron that does not appear to conform either to the U2 or
U12-dependent classes (Macknight, R et al (1997) *Cell*
89, 737-745).

Take-Home Message

The splice sites of the introns in Arabidopsis mRNA genes typically
conform to the GT-AG consensus of U2-dependent introns. However, on
rare occasions they may not and currently, a GC at the 5' splice site
is the most common exception. If an intron you are characterising
does not fit this consensus you should check for the consensus
sequences of the U12-dependent class: the 5'splice site sequence
appears to be highly conserved. If the sequence of the intron differs
from the properties of both these classes you may have something novel
and interesting. I intend to collate and update Arabidopsis
exceptions to the GT-AG rule and would be grateful if you would inform
me of any that you find: gordon.simpson@bbsrc.ac.uk

Further Information

Pierre Rouzéand Søren
Brunak, authors of NetPlantGene , a program that
identifies Arabidopsis introns, are also collecting information on
introns with non-canonical splice sites in order to develop a program
that will identify them in Arabidopsis gene sequences as well.

Up to date information of the U2 and U12-dependent classes of introns
is available in Dietrich, RC., Incorvaia, R and Padgett RA (1997)
Molecular Cell **in press**.

I have recently reviewed splicing in higher plants: Simpson GG and
Filipowicz W (1996) Splicing of precursors to mRNA in higher plants:
mechanism, regulation and sub-nuclear organization of the spliceosomal
machinery.
Plant Mol Biol, **32**, 1-41. I am currently
studying the role of RNA processing in regulating Arabidopsis
flowering time in
Caroline Dean's lab at the
John Innes Centre.

Acknowledgments