Volume 13 Number 13 1985 Nucleic Acids Research

Intron splicing: a conserved internal signal in introns of *Drosophila* pre-mRNAs

Elizabeth B.Keller and William A.Noon

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA

Received 16 April 1985; Revised and Accepted 5 June 1985

ABSTRACT

The introns of Drosophila pre-mRNAs have been analysed for conserved internal sequence elements near the 3' intron boundary similar to the T-A-C-T-A-A-C in yeast introns and the C/T-T-A/G-A-C/T in introns of other organisms. Such conserved internal elements are the 3' splice signals recognized in intron splicing. In the lariat splicing mechanism, the G at the 5' end of an intron joins covalently to the last A of a 3' splice signal to form a branch point in a splicing intermediate. Analysis of 39 published sequences of Drosophila introns reveals that potential 3' splice signals with the consensus C/T-T-A/G-A-C/T are present in 18 cases. In 17 of the remaining cases signals are present which vary from this consensus just in the middle or last position. In Drosophila introns the 3' splice signal is usually located in a discrete region between 18 and 35 nucleotides upstream from the 3' splice point. We note that the Drosophila small nuclear U2-RNA has sequences complementary to C-T-G-A-T, one variant of the signal, and to C-A-G, one variant of the 3' terminus of an intron. We also note that the absence of any A-G between -3 and -19 from the 3' splice point may be an essential feature of a strong 3' boundary.

INTRODUCTION

The well-known conserved sequences important for splicing of pre-mRNA introns are the 5' splice site with the consensus $\frac{C}{A}$-A-G↑G-T-$\frac{A}{G}$-A-G-T and the 3' splice site, a pyrimidine-rich region followed by $\frac{C}{T}$-A-G↑, where arrows indicate the two splice points (1,2). An additional conserved sequence was discovered in 1983 by Langford and Gallwitz (3) and Pikielny et al. (4) in an internal position in yeast introns. This sequence, required for splicing in yeast, is an invariant T-A-C-T-A-A-C located in all yeast introns just upstream from the 3' splice site (5). We call this type of conserved internal 3' sequence a 3' splice signal (6). Closely related sequences were noted by Woudt et al. (7) and Kinnaird and Fincham (8) in introns of Neurospora, and are present in the introns of other filamentous fungi (9-11). These sequences are probably the counterpart of the yeast TACTAAC, though in many cases Gs replace the first 2 As.

Recently it was discovered by Ruskin et al. (12) and Padgett et al.

(13) that mammalian introns are spliced *in vitro* by a mechanism in which the 5' terminal G of the intron first forms a branch with an A upstream from the 3' splice site resulting in a lariat structure. The intron is then spliced out in the form of a lariat. The sequence around the branch point in several globin introns has the consensus C-T-$\frac{G}{A}$-A-$\frac{C}{T}$ (12) which is identical to the last 5 nucleotides of the conserved 3' sequence in filamentous fungi. *In vivo* experiments on globin intron splicing confirm the lariat intermediate (14). *In vitro* studies on splicing of an adenovirus intron have indicated a branch at the A in the sequence C-T-T-A-T (15). The recent finding that the essential 3' splice signal T-A-C-T-A-A-C of yeast is a branch point sequence in a lariat intermediate (16, 17) suggests that all eukaryotes employ the lariat mechanism for intron splicing.

We have found that potential 3' splice signals like those above are present near the 3' boundary of introns from a wide variety of animals (6). We have proposed that selection of a true 3' intron boundary depends on two distinct kinds of recognition. One kind is sequence-specific recognition of the two conserved elements, the 3' splice signal and the 3' terminal $\frac{C}{T}$AG↑. The other is a different kind of recognition of the pyrimidine-rich region of variable sequence which lies between the two sequence-specific elements. We noted that the sequence of rat small nuclear U2-RNA is complementary to both of the sequence-specific elements of rat introns, and we therefore proposed that U2-ribonucleoprotein (U2-RNP) may be required for recognition of a 3' intron boundary.

This paper identifies potential 3' splice signals in *Drosophila* introns and proposes a role for *Drosophila* U2-RNP in splicing.

RESULTS AND DISCUSSION

Selection of Potential 3' Splice Signals in Drosophilia Introns.

In Table 1 we present the 3' sequences from 39 introns in *Drosophila* genes. Fifty nucleotides are given in each case. Two sets of criteria were used in selecting potential 3' splice signals in these introns. The first prescribes the sequences allowed and the second the distance from the 3' splice point.

With regard to allowed sequences, we started with the last 5 nucleotides of the conserved *Neurospora* consensus C-T-$\frac{G}{A}$-A-C (8). We assumed that the A at position 4 is invariant, as this would be the splicing branch point (see Table 2 for the numbering system). We have also considered as invariant the T at position 2, and have been able to find signals in 35 cases with a T at this

position. We assumed that a wobble-type switch of T and C, which occurs at the 3' end of the signals in human globin introns (12), can also occur at the 5' end, thus allowing $\frac{C}{T}$-T-$\frac{G}{A}$-A-$\frac{C}{T}$. Signals agreeing with this consensus can be found in many Drosophila introns, but not all. For the remaining cases, we allowed deviations only at positions 3 and 5. Deviations at these 2 positions have been reported in mammalian branch point sequences (12, 15, and M. Green, personal communication).

Signals conforming to these criteria can be found in 35 of the 39 introns given in Table 1. The resulting predominant consensus for Drosophila was found to be $\frac{C}{T}$-T-$\frac{A}{G}$-A-$\frac{T}{C}$ (Table 1, c). By analogy with other DNA and RNA signals, we consider the signals agreeing with this predominant consensus as strong signals and the signals deviating from it as weaker ones. The complete consensus structure (Table 1, b) shows the number of these weaker signals. In four cases we could not find a signal to match the above criteria. Branch point analysis will be required in those cases to locate the signal used.

With regard to the acceptable distance between signal and 3' splice point, we note that most of the signals fall in the region between -18 and -35. This discrete region appears to be the optimum locus for 3' splice signals in Drosophila. We can find a signal within this locus in 30 introns, and we can find one just upstream from this locus in 5 introns. In 3 of these latter cases there is an alternative weaker signal lying within the optimum locus which could be the preferred signal if distance is the critical factor in determining splicing efficiency in Drosophila.

In cases where alternative signals can be seen, branch point analysis will be needed to determine which signal is actually employed in splicing.

Of the 39 introns shown in Table 1, 24 have signals which match the consensus $\frac{C}{T}$-T-$\frac{A}{G}$-A-$\frac{T}{C}$ and are located from 15 to 40 bases upstream of the 3' splice point. We estimate the probability of this occurring in random DNA to be one in 3.1×10^4. In 10 additional cases, signals can be found that differ from this consensus in only one position.

Organization of the 3' splice site of Drosophila introns.

A 3' splice site has generally been taken to be $\frac{C}{T}$-A-G↑ preceded by a pyrimidine rich region, $Y_{11}N$, free of A-Gs (1, 2, 34). From inspection of the Drosophila introns in Table 1, it can be seen that in quite a number of cases the 3' splice site is not pyrimidine-rich. We note, however, that in 36 of the introns, A-G is absent from positions -3 to -19 (numbering upstream from the splice point). We have overlined all A-Gs, except the 3' terminal A-G, in order to emphasize that the region upstream from -19 shows the expected random

Table 1. 3' Intron Boundaries from Drosophilia Genes.

a) Sequences	Dis.	Signal	Gene Source	Ref.
CGTTCGCTTCAGCAGTTGCACTTGTGCTTAATCCTTTGGTGCACTTTCAG	18	T-T-A-A-T	Actin 88F	18
AGCTCCACGGTCCGTAGCACCACCAATGGATTCTATTTCCGCCTCTTCAG			Actin 79B	
GATACGAACCAACATACTCGATCCCTAACGAATGCCTATTTCTCCTTCAG	21	C-T-A-A-C	Cuticle protein 1	19
CAGGACATTTATGAACTCGCCATCTAATTGGTATCATTTCCTCTATCCAG	22	C-T-A-A-T	" " 2	
ATGCCTTGAGGAGCATAGTGACTTCGCAGTCTAATCCTGGATTATCCTAG	15	C-T-A-A-T	" " 3	
TCTGAAGTTTAAAGCCGGACAGTTCAATGAGTAATCCCGGAATATCCTAG	37	T-T-A-A-A	" " 4	
TGAACCCACAGACTATAACTAATCCTAATGATTTTGTAAATCCATTGCAG	21	C-T-A-A-T	HSP83	20
ATTTTTTATGTGACCCAAATCCACTTAGCCATCCGTTCATTCTGACCCAG	22	C-T-T-A-G	68C Glue gene II	21
ATCTTTAATCCACAACCAACTTCAATATCTCGCATCCTCAATATCCCCAG	41	T-T-A-A-T	" " " III	
TTTTCGGTATTTGAGCCACTGATATATTCATCCGTTTGCCTTCTCCACAG	27	C-T-G-A-T	" " " IV	
ACACGAACTGGAAACCAACAACTAACGGAGCCCTCTTCCAATTGAAACAG	24	C-T-A-A-C	ADH D. mel. I 1	22
CCACAGGCTCCATGCAGCGATGGAGGTTAATCTCGTCTATTCAATCCTAG	19	T-T-A-A-T	" " I 2	23
GCAAAGTTTTCAAAAAAAAACAAAACTAATTTGATTTATAACACCTTTAG	20	C-T-A-A-T	" " I 3	
TGATTGCATCCATTGGTTGATTGAGGTTAAATTCGTCTTATCAATCCTAG	29	T-T-G-A-T	" D. ore. I 2	
GACTAAGACCGCAAAGTAAAAAACGCTAATTTCCTCAATGACACCTTTAG	20	C-T-A-A-T	" D. ore. I 3	
ACTGACATTGCGGTTCTTATTGATATTCCATTTATATATACCACTTTCAG	26	T-T-G-A-T	Gene H	24
GTCGAAAGCGTACTACTTAGCCTATTATCTTACTTCACTCTTGCACTTAG	17	C-T-T-A-C	Gene L	
TATAAAAGCTCATTGAGCTAACCCATTTTTTCTTTGCTTATGCTTACAG	28	C-T-A-A-C	ftz	25
AAAAGTTCTACCAATCATGTTATATTTACACGCACTATCCTATCCCGCAG	21	T-T-T-A-C	Yolk protein 2	26
NNNNTTTCTTATCCCCTACCAGATTCATAATAACCTGATATTTTACCTAG	22	T-T-C-A-T	Ras 2	27
AAATCTATTGTGATTAATGTGTATTAACTTGAACTATGAATCGCTTGCAG	22	T-T-A-A-C	Dsrc	28
ACTAAGCAAGAAGCAAGAGGAAATGCCATAAATATGTTTGCTTATTTCAG	44	C-T-A-A-G	Dash	
ACACTCGAACTAATGCCAAACTCTCAAATCTTTTTCGATTCCGTTGACAG	36	C-T-A-A-T	MLC-ALK I 1	29
ACTAACATGAACATGAAGATGTAAATGCTAATCGAGTTCCTTTTGCGCAG	18	C-T-A-A-T	" I 2	
AACATGCATATATCCTGGGCTATCCCTCTAAATCATTGTGTGCCCTATAG	18	C-T-A-A-A	" I 3	
TTGTTTACCCGCCGTTTGTCGCCGATCCACTTGAAATTGAATGAAAACAG	15	T-T-G-A-A	" I 4	
ATTTGGGTAATCCCTTGCTAATATGGCCCTTTTCTTCGTCTACTCCACAG	28	C-T-A-A-T	" I 5	
CAACTTATATGCTTTGAAAGAAATCCTTTTAAAATATCTTTATCAAATAG	17	T-T-A-A-A	GARTrans. I 1	30
TTTTGACCACTTCCCTTAGATCTTCATAATAGCTTGTATCGTACTTTCAG	23	T-T-C-A-T	" I 2	
ACCTATTGATCAGCTACTACTATCTTCTTCCGTATCCTTTCCATACGTAG	40	T-T-G-A-T	" I 3	
ATGACCGAGCTCCTCTGAAATCCATCGATTAATTTGAAATTTCTCTTTAG	17	T-T-A-A-T	" I 4	
TGGTATCTTTTTATATACAAATGCTTTATGCTTTGTCCAATCTCCTCCAG	21	T-T-T-A-T	" I 5	
AAAACTTTCCAAGTACTAACTGCTTTACATATACATACATTACTTTCCAG	30	C-T-A-A-C	G3PD	31
TCGTAGGATACTTCGTTTTGTTCGGGGTTAGATGAGCATAACGCTTGTAG			White locus I 1	32
GATCTGTGTGAAATCTTAATAAAGGGTCCAATTACCAATTTGAAACTCAG	30	T-T-A-A-T	" I 2	
5' splice point-->GTGAGTTCGATGTGTTTATTAAGGGTATCTAG			" I 3	
TGTTGAAGATATGAATATTAATGAGATGCGAGTAACATTTTAATTTGCAG	28	T-T-A-A-T	" I 4	
ATTTGCATATTAATAATTTACTAACTTTCTAATGAATCGATTCGATTTAG	17	C-T-A-A-T	" I 5	
AAAAAAAAAAAAACCTCGAATATTCTATGGAATATATATATCCTTTGTAG			Yolk protein 1	33

Table 1. (cont.)

b) Consensus Structure

A	-	-	24	35	4
G	-	-	5	-	2
C	18	-	2	-	7
T	17	35	4	-	22

c) $\frac{C}{T}$ T $\frac{A}{G}$ A $\frac{T}{C}$

a) The last 50 bp of introns from various Drosophilia genes. Multiple introns are labeled I1, etc. The proposed primary signals are underscored with solid lines and aligned under "Signal". Distance is the number of base pairs between the end of the primary signal and the 3' splice point. Alternative weaker signals are underscored with a dotted line. AG's other than the final AG are overscored. The line between -19 and -20 delimitsthe predominantly AG-free region. In 4 cases, no signal agreeing with the consensus structure was found.
b) Consensus structure of the primary signals indicating the number of occurences of each base in each position.
c) A predominant consensus for 3' splice signals in Drosophilia .

distribution of A-Gs, whereas the region starting at -19 and extending downstream to the $\frac{C}{T}$-A-G↑ is A-G-free. We conclude that the absence of an A-G from -3 to -19 is a more consistent feature of this region than a predominance of pyrimidines. We therefore propose to define a 3' splice site in Drosophila as $\frac{C}{T}$-A-G↑ preceded by a 16 nucleotide region which is free of A-Gs and which is usually, but not necessarily, pyrimidine-rich.

We have found that this definition of a 3' splice site in Drosophila also holds for introns of other organisms, for example, for a series of 50 introns from human genes (Keller and Noon, unpublished). It is also noteworthy that Wieringa et al. (35), working with mutant globin introns, found inhibition of splicing when an A-G was inserted into the region downstream from -20; the closer the A-G to the 3' splice point, the greater the inhibition.

Species Specificity of the 3' Splice Signal.

We have identified potential 3' splice signals in the introns of a wide range of organisms. From these studies an overall pattern of evolutionary change has emerged (Table 2). The yeast signal has 7 nucleotides (3-5), while the signals of other fungi have either 6 or 7 nucleotides (7-11), and those of higher organisms are limited to 5 nucleotides. The 5 nucleotides of higher organisms correspond to the last 5 nucleotides of the fungal signals.

Signals of fungi are invariant or of very limited variability, but those of higher organisms vary considerably at certain positions. It is likely that

Table 2: Species difference in 3' splice signal consensus[a].

Species	Position[b] 1-2-3-4-5	Ref.
Yeast	T-A-C-T-A-A-C	3,4,5
Filamentous fungi	$\begin{matrix}T & G\\A & A\end{matrix}$-C-T-$\begin{matrix}G\\A\end{matrix}$-A-C	7-11
Plant	$\begin{matrix}C\\T\end{matrix}$-T-$\begin{matrix}A\\G\end{matrix}$-A-$\begin{matrix}T\\C\\A\end{matrix}$	c
Nematode	$\begin{matrix}C\\T\end{matrix}$-T-$\begin{matrix}G\\A\\C\end{matrix}$-A-$\begin{matrix}T\\C\\A\end{matrix}$	36,37
Sea Urchin	A-$\begin{matrix}T\\C\end{matrix}$-T-A-A-T	6
Drosophila	$\begin{matrix}C\\T\end{matrix}$-T-$\begin{matrix}A\\G\end{matrix}$-A-$\begin{matrix}T\\C\end{matrix}$	this paper
Chicken and Duck	$\begin{matrix}C\\T\end{matrix}$-T-$\begin{matrix}G\\A\\C\end{matrix}$-A-$\begin{matrix}C\\T\end{matrix}$	6
Rat and Mouse	C-T-$\begin{matrix}G\\A\end{matrix}$-A-$\begin{matrix}C\\T\end{matrix}$	6
Human	$\begin{matrix}C\\T\end{matrix}$-T-$\begin{matrix}A\\G\\C\end{matrix}$-A-$\begin{matrix}T\\C\\G\end{matrix}$	6,12,14

[a]The consensus given is the predominant one. It excludes nucleotides which occur less than 10-15% of the time.
[b]The position numbers refer only to the 5 place signals.
[c]To be published.

the A in position 4 is invariant in all organisms, since the 2' branch in the lariat structure has been mapped to this A. From our analysis of introns from many different species, we suggest that the T at position 2 may be largely invariant in all organisms. In positions 1, 3, and 5, the signals apparently become more and more variable as one proceeds up the evolutionary ladder. We find that the signals in a series of 50 introns from human genes diverge considerably more than the Drosophila signals (Keller and Noon, unpublished).

Distinguishing True from Cryptic 3' Boundaries.

A 3' boundary has been generally defined by the consensus Y_{11}-N-$\begin{matrix}C\\T\end{matrix}$-A-G↑ with A-G being absent from the Y_{11} region (2, 34). But sequences agreeing with this consensus are seen to occur frequently in the interior of introns,

and in some cases these false sites agree more closely with this consensus than the true 3' splice site does (38). Recognition of the true 3' boundary must therefore depend upon additional criteria. We are proposing two important additional criteria: first, the presence of a 3' splice signal at an appropriate distance from the 3' splice point, and second, the absence of an A-G between -3 and -19.

Using the new criteria, we searched for potential 3' intron boundaries in a number of long introns and found that in many cases these new criteria were sufficient to eliminate the false boundaries. However, in a few cases, a potential 3' boundary fulfilling the new criteria could still be found within the intron. The signals in these false boundaries may prove to be weaker than the signals in the true boundaries.

The Possible Role of U2-RNP in Intron Splicing in Drosophila.

In 1980 Lerner et al. (39) noted that a sequence exactly complementary to the consensus of the 5' intron boundary is present in the small nuclear U1-RNA. It is now established that the ribonucleoprotein particle containing U1-RNA (U1-RNP) is required for intron splicing in vertebrate systems (40,41). Since present evidence supports a common lariat mechanism for intron splicing, it is probable that a U1-RNP or an analogue of it functions in splicing in all eukaryotes.

We previously noted that there is a similar complementarity between sequences in the small nuclear U2-RNA and the two conserved parts of the 3' intron boundary: the 3' splice signal and the terminal $\frac{C}{U}$-A-G. We therefore proposed that the U2-RNP recognizes the 3' intron boundary just as the U1-RNP recognizes the 5' boundary (6). We first noted this complementarity (6) in the case of the rat U2-RNA (42). The complementary region in this U2-RNA lies between nucleotides 41 and 53, and includes C-U-G and A-U-C-A-G which could hybridize to the 3' signals $\frac{C}{U}$-U-G-A-U and $\frac{C}{U}$-A-G if G-U pairs are allowed at helix ends (43). For other variants of the rat 3' splice signal, hybrids would necessarily be weaker, just as they would be for variants of the 5' splice site with U1-RNA. This model was supported by the finding of Pederson's group that U2-RNA binds to heterogeneous nuclear RNA in HeLa cells in vivo (44) in the same way as U1-RNA (45).

We now propose to extend this model to splicing in Drosophila. The Drosophila U2-RNA sequence (Fig. 1, ref. 46, 47) is identical to the rat U2-RNA sequence in the pertinent region, and the consensus structure of the 3' splice signal is nearly the same in the two organisms (Table 2). Figure 1 shows how the 3' splice signal $\frac{C}{U}$-U-G-A-U and the 3' terminal $\frac{C}{U}$-A-G could

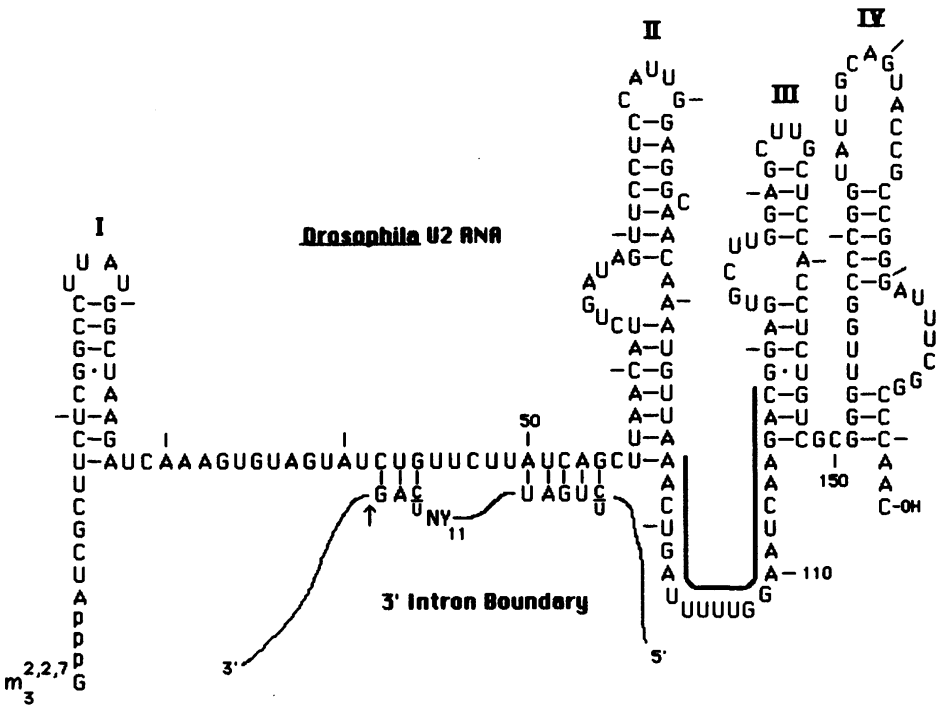

Fig. 1. Secondary structure model of Drosophila U2-RNA when present in U2-RNP, showing the proposed binding of a 3' splice signal C/U-U-G-A-U and 3' intron terminus C/U-A-G. The 3' half of the RNA is from RNA sequencing (46) with a typical U-RNA cap added, and the 5' half is from sequencing of the U2-RNA gene (47). Hairpins II, III, and IV are from references 46 and 48. The region around A-U-U-U-U-U-G which is protected by protein binding in human U2-RNP (49, 50) is overlined. The calculated stability (51) of hairpin I is ΔG = -8.4 kcal, much greater than the stability, ΔG = -2.8 kcal, of the alternative hairpin in two published models for U2-RNA (42, 50). In the human complex of U2-RNP with hnRNP, the region coinciding with the 5' terminus and hairpin I is protected by protein binding (52). The nuclease cleavage sites of Reddy et al. (42) are consistent with the above model.

hybridize to the region between nucleotides 42 and 54 in the Drosophila U2-RNA. Note that we are proposing here that in two cases a terminal G-U pair may form as an alternative to a G-C pair in the protein-binding environment of the U2-RNP. One of these wobble pairings could explain the $\frac{C}{U}$ at position 1 of the 3' splice signal and the other the $\frac{C}{U}$ preceding the A-G at the 3' terminus.

Recently Sass and Pederson (53) published evidence supporting a possible role for U2-RNA in splicing in Drosophila. They showed that U2- and U1-RNP's

are both localized at transcriptionally active sites on insect polytene chromosomes. There could be, however, a problem with recognition of the 3' splice signal by U2-RNA in the case of Drosophila because the predominant nucleotide in the central position of the Drosophila 3' splice signal is A and not G (Table 1). Perhaps as more Drosophila introns are sequenced, G will appear in this position more frequently. There also could be protein-DNA interactions in the U2-RNP which would strengthen the binding in the absence of a standard base pair.

If U2-RNA is to recognize the 3' intron boundary, the pertinent region of the U2-RNA must be single stranded in the U2-RNP. In our secondary structure model in Figure 1, the critical region is single stranded. In addition, the region must not be blocked by protein. Sri-Widada et al. (52) have demonstrated that this region is not blocked by protein in the intact human U2-RNP complexed with hnRNP, since the region is readily digested by micrococcal nuclease. In contrast, two other regions of the U2-RNA are blocked by protein and resistant to micrococcal nuclease in this complex. One blocked region extends from the cap to nucleotide 32 and the other includes the distinctive A-U-U-U-U-U-G (overlined in Fig. 1).

It is of interest that Sri-Widada et al. (53) found that U2-RNA in the U2-RNP-hnRNP complex has a very exposed single-stranded region that is susceptible to light treatment with pancreatic RNase. The exposed region could be the portion of the long single-stranded sequence which precedes the proposed binding sites.

It is significant that the sequences of the U2-RNAs of rat (42), mouse (54), human (55,56), Xenopus, (57) Drosophila (46, 47) and wheat (58) are invariant in the critical complementary region. This invariance is consistent with the idea that the sequence has an essential function as a single stranded region. In contrast, the folded regions of U2-RNAs of these organisms often show compensatory sequence changes which allow retention of the folded structures of the loops.

Our analysis of the Drosophila U2-RNA contributes further support for our splicing model in which U2-RNP recognizes the 3' boundary of an intron. According to this model U1-RNP would be bound to the 5' boundary and U2-RNP to the 3' boundary and then the two would join to form a large stable complex within which splicing would occur. In such a stable complex the splicing intermediates could be securely held until the various steps of the lariat splicing mechanism are completed.

ACKNOWLEDGEMENTS

We want to express our great indebtedness to the many individuals who obtained the sequence data we employed in this study. We thank M. Green, B. Ruskin, R. Reed, R. Breathnach, A. Efstratiadis, H. Sass, T. Pederson, S. M. Mount, R. A. Padgett, S. Falkenthal, N. Davidson, W. Keller, M. Ares, J. M. Skuzeski, and R. Wu for communicating unpublished information. We also thank K. Kindle and J. Calvo for valuable criticisms of the manuscript. This study was supported by National Institute of Health grant GM10791.

REFERENCES

1. Breathnach, R., and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349-384.
2. Mount, S.M. (1982) Nucleic Acids Res. 10, 459-472.
3. Langford, C.J. and Gallwitz, D. (1983) Cell 33, 519-527.
4. Pikielny, C.W., Teem, J.L., and Rosbash, M. (1983) Cell 34, 395-403.
5. Langford, C.J., Klinz, F., Donath, C., and Gallwitz, D. (1984) Cell 36, 645-653.
6. Keller, E. B. and Noon, W. A. (1984) Proc. Nat. Acad. Sci. USA 81, 7417-7420.
7. Woudt, L.P., Pastink, A., Kempers-Veenstra, A.E., Jansen, A.E.M., Mager, W.H., and Planta, R.J. (1983) Nucl. Acids Res. 11, 5347-5360.
8. Kinnaird, J. H. and Fincham, J. R. S. (1983) Gene 26, 253-260.
9. Shoemaker, S., Schweickart, V., Ladner, M., Gelfand, D., Kwok, S., Myambo, K., and Innis, M. (1983) Biotechnology 1, 691-696.
10. Boel, E., Hansen, M.T., Hjort, I., Hoegh, I., and Fiil, N.P. (1984) EMBO J. 3, 1581-1585.
11. Dons, J.J.M., Mulder, G.H., Rouwendal, G.J.A., Springer, J., Bremer, W., and Wessels, J.G.H. (1984) EMBO J. 3, 2101-2106.
12. Ruskin, B., Krainer, A.R., Maniatis, T., and Green, M.R. (1984) Cell 38, 317-331.
13. Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F., and Sharp, P. A. (1984) Science 225, 898-903.
14. Zeitlin, S. and Efstratiadis, A. (1984) Cell 39, 589-602.
15. Konarska, M. M., Grabowski, P. J., Padgett, R. A., and Sharp, P. A. (1985) Nature 313, 552-557.
16. Domdey, H., Apostol, B., Lin, R.-J., Newman, A., Brody, E., and Abelson, J. (1984) Cell 39, 611-621.
17. Rodriguez, J. R., Pikielny, C. W., and Rosbash, M. (1984) Cell 39, 603-610.
18. Sanchez, F., Tobin, S.L., Rdest, U., Zulauf, E., and McCarthy, B.J. (1983) J. Mol. Biol. 163, 533-551.
19. Snyder, M., Hunkapiller, M., Yuen, D., Silvert, D., Fristrom, J., and Davidson, N. (1982) Cell 29, 1027-1040.
20. Hackett, R.W. and Lis, J.T. (1983) Nucl. Acids Res. 11, 7011-7030.
21. Garfinkel, M. D., Pruitt, R. E., and Meyerowitz, E. M. (1983) J. Mol. Biol. 168, 765-789.
22. Benyajati, C., Spoerel, N., Haymerle, H., and Ashburner, M. (1983) Cell 33, 125-133.
23. Bodmer, M. and Ashburner, M. (1984) Nature 309, 425-430.
24. Snyder, M. and Davidson, N. (1983) J. Mol. Biol. 166, 101-118.
25. Laughon, A. and Scott, M.P. (1984) Nature 310, 25-31.
26. Hung, M.-C. and Wensink, P.C. (1983) J. Mol. Biol. 164, 481-492.

27. Neuman-Silberberg, F.S., Schejter, E., Hoffmann, F.M., and Shilo, B.-Z. (1984) Cell 37, 1027-1033.
28. Hoffmann, F.M., Fresco, L.D., Hoffman-Falk, H., and Shilo, B.-Z. (1983) Cell 35, 393-401.
29. Falkenthal, S., Parker, V.P., and Davidson, N. (1985) Proc. Nat. Acad. Sci. USA 82, 449-453.
30. Henikoff, S., Sloan, J. S., and Kelly, J. D. (1983) Cell 34, 405-414.
31. Tso, J. T., Sun, X.-H., and Wu, R. (1985) Nucl. Acid Res. in press.
32. O'Hare, K, Murphy, C., Levis, R., and Rubin, G. M. (1985) J. Mol. Biol. 180, 437-455.
33. Hovemann, B., Galler, R., Walldorf, U., Kuepper, H., and Bautz, E. K. F. (1981) Nucl. Acid Res. 9, 4721-4734.
34. Seif, I., Khoury, G., and Dhar, R. (1979) Nucl. Acids Res. 6, 3387-3398.
35. Wieringa, B., Hofer, E., and Weissmann, C. (1984) Cell 37, 915-925.
36. Kramer, J. M., Cox, G. N., and Hirsh, D. (1982) Cell 30, 599-606.
37. Karn, J., Brenner, S., and Barnett, L. (1983) Proc. Nat. Acad. Sci. USA 80, 4253-4257.
38. Staden, R. (1984) Nucleic Acids Res. 12, 505-519.
39. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L., and Steitz, J. A. (1980) Nature 283, 220-224.
40. Padgett, R. A., Mount, S. M., Steitz, J. A., and Sharp, P. A. (1983) Cell 35, 101-107.
41. Kramer, A., Keller, W., Appel, B., and Luhrmann, R. (1984) Cell 38, 299-307.
42. Reddy, R., Henning, D., Epstein, P., and Busch, H. (1981) Nucl. Acids Res. 9, 5645-5658.
43. Mizuno, H. and Sundaralingam, M. (1978) Nucl. Acids Res. 5, 4451-4461.
44. Calvet, J.P., Meyer, L.M., and Pederson, T. (1982) Science 217, 456-458.
45. Calvet, J.P., and Pederson, T. (1981) Cell 26, 363-370.
46. Myslinski, E., Branlant, C., Wieben, E. D., and Pederson, T. (1984) J. Mol. Biol. 180, 927-945.
47. Alonso, A., Jorcano, J.L., Beck, E., and Spiess, E. (1983) J. Mol. Biol. 169, 691-705.
48. Branlant, C., Krol, A., Ebel, J.-P., Lazar, E., Haendler, B., and Jacob, M. (1982) EMBO J. 1, 1259-1265.
49. Liautard, J.-P., Sri-Widada, J., Brunel, C., and Jeanteur, P. (1982) J. Mol. Biol. 162, 623-643.
50. Mattaj, I. W. and De Robertis, E. M. (1985) Cell 40, 111-118.
51. Cech, T.R., Tanner, N.K., Tinoco, I.,Jr., Weir, B.R., Zuker, M., and Perlman, P.S. (1983) Proc. Nat. Acad. Sc. USA 80, 3903-3907.
52. Sri-Widada, J., Liautard, J.-P., Brunel, C., and Jeanteur, P. (1983) Nucl. Acids Res. 11, 6631-6646.
53. Sass, H. and Pederson, T. (1984) J. Mol. Biol. 180, 911-926.
54. Nojima, H. and Kornberg, R. D. (1983) J. Biol. Chem. 258, 8151-8155.
55. Van Arsdell, S. W. and Weiner, A. M. (1984) Mol. Cell. Biol. 4, 492-499.
56. Westin, G., Zabielski, J., Hammarstrom, K., Monstein, H.-J., Bark, C., and Pettersen, U. (1984) Proc. Nat. Acad. Sc. USA 81, 3811-3815.
57. Mattaj, I. W. and Zeller, R. (1983) EBMO J. 2, 1883-1891.
58. Reddy, R. and Busch, H., (1983) Prog. Nuc. Acid Res. Mol. Biol 30, 127-162.