Intron splicing: A conserved internal signal in introns of animal pre-mRNAs

(3' splice signal/3' splice site/computer search/U2 small nuclear RNA)

ELIZABETH B. KELLER AND WILLIAM A. NOON

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

Communicated by Gordon G. Hammes, August 17, 1984

ABSTRACT Splicing of introns of yeast pre-mRNAs requires an internal conserved sequence T-A-C-T-A-A-C that is located 20–55 nucleotides from the 3' intron boundary. Sequences differing only in certain positions from this yeast signal have now been identified in the corresponding internal region of pre-mRNA introns of a variety of animal genes. A computer program that searches for homologues to a consensus structure and calculates the accuracy of match of each homologue is used to locate these sequences. We list here the signals found by this search in introns of sea urchin, mouse, rat, and human genes and give the consensus for each species. We also give the consensus found for *Drosophila* and chicken and duck signals. We then discuss the accumulating evidence that these internal signals are required for splicing in animals. It is also noted that a single-stranded region of small nuclear RNA U2 contains sequences complementary both to the proposed mammalian internal signal and to the neighboring $\frac{C}{T}$ -A-G at the 3' intron boundary. A role for U2 ribonucleoprotein in intron splicing is thus suggested.

In eukaryotes many of the nuclear genes coding for proteins are interrupted by one or more introns that must be precisely spliced out of the initial transcript of the gene before the RNA is transported to the cytoplasm to be translated. Introns are bounded at the 5' end by a conserved sequence of 9 nucleotides called a 5' splice site. They are bounded at the 3' end by a 3' splice site that is a pyrimidine-rich region of about 11 nucleotides followed by N-Y-A-G, where N is any nucleotide and Y is a pyrimidine (1, 2). Langford and Gallwitz (3) and Pikielny *et al.* (4) discovered that in yeast introns there is an additional signal, T-A-C-T-A-A-C, which lies within the intron near the 3' splice site and which is required for splicing. Apparently in yeast the information for specifying a 3' boundary is provided by the coupling of two neighboring signals T-A-C-T-A-A-C and (Y)₁₁-N-Y-A-G, which are separated by a variable distance. In this paper we will call this new type of internal signal by the descriptive name "3' splice signal" or "3' signal" while retaining the term "3' splice site" for the conserved 3' boundary sequence.

The yeast 3' splice signal is very highly conserved; in 16 examples compiled by Langford *et al.* (5) there is not a single deviation from T-A-C-T-A-A-C. If the first or last A is changed to C, splicing is abolished (5). However, the yeast splicing machinery can recognize variations in certain positions of this sequence, because splicing can occur when the signal is A-A-C-T-A-A-C (6) or T-A-T-T-A-A-C (5).

Potential 3' splice signals similar to the yeast signals have been noted in 9 introns of *Neurospora* genes (7, 8). The consensus of the presumed *Neurospora* 3' splice signal, C-T- $\frac{C}{T}$ -

A-C, has only five conserved positions with either G or A in the third position. Nevertheless, these signals in *Neurospora* appear to correspond to the T-A-C-T-A-A-C of yeast.

The full yeast signal T-A-C-T-A-A-C is only rarely seen in metazoan introns, and metazoan introns lacking the T-A-C-T-A-A-C are not spliced out in yeast (9). One interpretation of these findings has been that yeast (and perhaps fungi in general) has a splicing mechanism that requires the 3' signal, whereas metazoans have a different mechanism that does not require this signal (3, 4, 10). An alternative interpretation, which we adopted, is that all eukaryotes share a common basic mechanism for intron splicing that requires a 3' splice signal like the one discovered in yeast, but that the exact sequence of this signal has been modified in different organisms in the course of evolution. With the idea that a 3' splice signal is essential for providing enough information to correctly specify a 3' boundary, we have for some time been compiling lists of 3' splice signals related to the yeast and *Neurospora* signals in the introns of various organisms. In this paper we describe the computer programs used in this search and present some of the lists of potential 3' signals we have located.

Using *in vitro* intron splicing systems, Grabowski *et al.* (11), Ruskin *et al.* (12), and Padgett *et al.* (13) have recently discovered that there is an intermediate in the splicing reaction that is in the form of a lariat. In this intermediate, the 5' end of the intron has been released from the first exon and the 5' terminal G is joined in a branched structure to a point in the intron near the 3' splice site. Ruskin *et al.* (12) report that the branch occurs at an A and that the sequence around this A is C-T-G-A-C, C-T-A-A-T, C-T-G-A-T, C-T-A-A-C, and C-T-C-A-C in the five cases examined. It is of great interest that the branch points occur in the potential 3' splice signals that were selected by our computer search. Thus, the branch point analysis provides evidence for a role for the proposed 3' splice signal in the splicing reaction. The fact that we can find 3' splice signals in the introns of various organisms indicates that the lariat mechanism is probably the general mechanism for pre-mRNA splicing in eukaryotes.

We also note that there are sequences present in the rat U2 small nuclear RNA (snRNA) (14) that are complementary to a predominant mammalian 3' splice signal and also to the $\frac{C}{T}$ -A-G at the 3' intron boundary. These sequences are in a part of U2 snRNA that is predicted to be single stranded (15) and that is precisely conserved in various organisms (14). The possible role of U2 small nuclear ribonucleoprotein (snRNP) in splicing will be discussed.

METHODS

Searching DNA Sequences for Homologues to a Known DNA Signal. Where authentic examples of a DNA signal have been reported, it is possible to write a consensus struc-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: snRNA, small nuclear RNA; snRNP, small nuclear ribonucleoprotein; I or IVS, intron; NC, noncoding region.

ture or frequency table that shows the frequency of occurrence of each base at each position in the signal. A search for additional examples of the signal in other DNA sequences can be made most effectively if the search employs such a consensus structure rather than the simple consensus derived from it. The search program used in this study employs such a consensus structure. The program, which was developed by J. T. Lis and J. Schuster, is called Makecons and is written in Pascal for use on an Apple microcomputer. The program searches a DNA sequence for a string of bases that matches a given consensus structure to a specified level of stringency. The program calculates and presents for each acceptable string the percent match to the consensus structure. The percent match is calculated as the sum of the frequency of occurrence of each base in the string divided by the sum of highest frequencies. This calculation is based on the assumption that in nature the strength of recognition of a signal is proportional to the sum of the strengths of the individual interactions. This simple evaluation system is useful for initial screening for a signal. Other criteria can then be applied in the next stage of the analysis.

Searching for Internal Signals in Introns from Different Organisms. It was necessary to develop a new search procedure to locate in the introns of a given organism a new signal related to, but not identical with, the authentic 3' splice signal of yeast. This search requires operation of the Makecons program through three stages. In the first stage the C-T-A-A-C of the yeast signal (3-5) or the consensus structure of the potential *Neurospora* signal (8) is employed at high stringency to search the introns of organism A from -10 to -60 from the 3' splice point. This first stage shows what proportion of the introns of organism A have a 3' splice signal closely related to the yeast signal or the *Neurospora* consensus structure. From this list of closely related signals a first consensus structure is derived for what is presumed to be the strongest form of the signal in organism A.

In the second stage, the first consensus structure is used at low stringency to find signals in all of the remaining introns. When an intron shows multiple choices at low stringency, selection rules are applied to locate the best match to the yeast and *Neurospora* signals. Any substitution of a pyrimidine for a purine or vice versa is considered a bad match. Wherever possible a signal with T at position 2 and A at position 4 is selected. When more than one good signal is found in an intron, the one closest to the 3' splice site is taken as the primary signal. When at least one signal has been found for each intron, a list of primary signals is made up, and a second consensus structure is developed from this list.

In the third stage, all of the primary signals are aligned in a table together with their flanking sequences to see whether the homology extends into the flanking regions.

RESULTS AND DISCUSSION

Potential 3' Splice Signals in Introns of Sea Urchin Genes. Two distinct actin genes from the sea urchin *Strongylocentrotus purpuratus* have been sequenced, and they contain a total of 5 introns (16). These introns were searched by using the C-T-A-A-C of the yeast 3' signal for the first stage. A potential signal A-T-T-A-A-T was located in each of the introns (Table 1). The distance of the signals from the 3' splice point ranges from 24 to 33.

The consensus structure of the flanking regions of these signals shows that the sea urchin consensus could be A-T-T-A-A-T (Table 1). The initial A extends the resemblance to the yeast signal to six positions.

Potential 3' Splice Signals in Introns of Mouse and Rat Genes. The sequences of 35 introns from 14 different rat and mouse genes were first searched at high stringency for ho-

Table 1. Potential 3' splice signals in introns of sea urchin genes

Intron	Distance	Signal	%	Ref.						
Signals, one for each intron searched										
Actin gene 1 I 121/122	24	A-T-T-A-A-T	100	16						
I 204	26	A-C-T-A-A-T	96							
Actin gene 2 I 5' NC	33	G-T-T-A-A-T	89							
I 121/122	28	A-T-T-A-A-T	100							
I 204	26	A-C-T-A-A-T	96							
Consensus structure of signals with flanking regions										
A	1	3	4	—	—	5	5	—	2	2
G	—	—	1	—	—	—	—	—	—	—
C	—	1	—	2	—	—	—	—	—	—
T	4	1	—	3	5	—	—	5	3	3
<div style="text-align: center;"> $\text{A-}\overset{\text{C}}{\underset{\text{T}}{\text{T}}}\text{-A-A-T}$ </div>										

% gives the match with the consensus structure in the lower portion of table. Distance is the number of nucleotides between the 3' end of the signal and the 3' splice point. I, intron. NC, noncoding region.

mologues of the *Neurospora* consensus structure (8). The signal C-T-G-A-N was found in 54% of the introns. The second stage of the search was carried out at low stringency and located at least one signal in each of the remaining introns (Table 2). For the whole set of 35, a consensus C-T-G-A-C can be written (Table 2). Though positions 3 and 5 are poorly conserved, in the 35 examples there are only 7 signals that do not have a pyrimidine pair followed by a purine pair, and it was possible to find a signal with A in position 4 in all 35 introns. The probability of finding a signal with this degree of variation in a region of 45 nucleotides in each of 35 introns is 7×10^{-10} if the DNA were simply random.

The results of the third stage (Table 2, middle portion) showed that there is no conservation of flanking positions and that the rodent signal, like that of *Neurospora* (8), most likely has only five positions.

Redundant signals of high homology are listed for four cases in Table 2. There are six other examples of the redundant signal C-T-C-A- C-T-T that have not been presented. Variant signals of lower percent match are present in many of the introns and these also have not been listed.

Potential 3' Splice Signals in Introns of Human Globin Genes. The introns of five human globin genes have been searched by the procedures used for the rat and mouse introns (Table 3). The human globin consensus structure (Table 3) is similar to that of the rat and mouse (Table 2). Ruskin *et al.* (12) have analyzed the lariat-type splicing intermediates of some of these introns. In the case of IVS 1 of α - and β -globins, the branch points they found agree with the primary signals selected by our program. However, in the case of the IVS 1 of ϵ -globin, the branch point agrees with our redundant signal C-T-A-A-T at -29 and not with our primary signal C-T-G-A-T at -17.

Potential 3' Splice Signals in Introns of Other Organisms. The introns compiled by Mount (2) have been searched for 3' splice signals with results in most cases similar to those in Tables 1-3. In addition, an analysis has been completed of all available *Drosophila* and chicken and duck introns (unpublished data). For each intron of these organisms, one signal has been selected as the potential primary signal, and a consensus structure has been developed (Table 4). In view of the findings of Ruskin *et al.* (12) and the earlier report by Wallace and Edmonds (36) that branching in total heterogeneous nuclear RNA occurs largely at A residues, it is notable that in most cases a signal can be found with an A in position 4.

Some rabbit and frog introns have also been searched with

Table 2. Potential 3' splice signals in introns of rat and mouse genes

Intron	Distance	Signal	%	Ref.
Primary signals, one for each intron searched				
Rat				
α -Actin I 5' NC	42	T-T-A-A-C	71	17
I 41/42	14	C-T-G-A-T	92	
I 150	22	C-T-C-A-C	88	
I 204	27	C-T-G-A-G	90	
I 267	37	C-T-G-A-C	100	
I 327/328	15	C-T-G-A-C	100	
β -Actin I 5' NC	26	G-T-G-A-C	79	18
I 41/42	20	C-T-G-A-G	90	
I 121/122	51	C-T-G-A-G	90	
I 267	19	C-T-G-A-T	92	
I 327/328	23	C-T-C-A-C	88	
Insulin I 5' NC	10	C-T-G-A-C	100	19
I 62	16	C-T-G-A-C	100	
γ -Crystallin I 2/3	18	C-T-G-A-C	100	20
I 83/84	32	C-T-G-A-C	100	
Cytochrome c	22	T-T-A-A-C	71	21
Myosin light chain I 1	35	C-T-G-A-C	100	45
I 2	53	C-T-G-A-T	92	
I 3	28	C-C-A-A-C	69	
I 4	19	C-T-C-A-T	79	
I 5	13	C-T-G-A-A	88	
I 6	16	C-C-T-A-C	65	
Mouse				
α A-Crystallin I 62/63	23	C-C-G-A-C	80	22
I 85/86	17	C-C-A-A-A	57	
Ig C μ region I 1:2	25	C-T-G-A-G	90	23
I 2:3	21	C-T-G-A-C	100	
I 3:4	18	C-T-C-A-C	88	
I 4:M1	25	T-T-G-A-C	82	24
I M1:M2	20	T-T-C-A-C	70	
Ig κ MOPC-41 I J:C	51	C-T-G-A-A	88	25
Ig λ -2a I J:C	43	C-T-G-A-G	90	26
Ig κ K2 I L:V	35	T-T-A-A-T	63	27
Ig κ L7 I L:V	19	C-T-A-A-T	81	28
Ig λ -1 I L:V	25	C-T-G-A-T	92	29
Ig H MOPC-141 I L:V	21	C-T-A-A-T	81	30

Consensus structure of primary signals with flanking regions

A	13	11	10	—	—	7	35	3	9	7	3
G	9	6	5	1	—	22	—	5	7	7	5
C	8	11	9	29	4	5	—	19	10	11	15
T	5	7	11	5	31	1	—	8	9	10	12
C-T-G-A-C											

Redundant signals

α -Actin I 5' NC	25	C-T-C-A-C	88	17
Insulin I 62	35	C-T-G-A-T	92	19
Ig λ -1 I L:V	37	C-T-G-A-T	92	29
Ig H MOPC-141 I L:V	33	C-T-A-A-T	81	30

Headings as in Table 1. % gives the match to the consensus structure of all 35 signals (positions 4–8 in the middle position of the table).

similar results. It therefore seems likely that vertebrates in general share common 3' splice signals. This view would be consistent with the finding that various vertebrate introns are spliced out correctly in the frog oocyte (37).

Evidence Supporting a Conserved 3' Splice Signal in Animal Introns. Proof that a 3' splice signal is required for splicing was first obtained in yeast by analysis of deletion mutations (3, 4). Deletions within the corresponding regions of introns from higher eukaryotes have been described (38). None of the deletions tested had any effect on splicing. Inspection of

Table 3. Potential 3' splice signals in introns of human globin genes

Intron	Distance	Signal	%	Ref.
Primary signals, one for each intron searched				
α -2-Globin IVS-1	17	C-T-C-A-C	95	31
IVS-2	16	C-T-G-A-C	100	
β -Globin IVS-1	35	C-T-G-A-C	100	32
IVS-2	31	C-T-A-A-T	80	
δ -Globin IVS-1	33	C-T-G-A-C	100	33
IVS-2	34	T-T-A-A-C	73	
γ -A-Globin IVS-1	28	C-T-G-A-C	100	34
IVS-2	38	C-T-C-A-C	95	
ϵ -Globin IVS-1	17	C-T-G-A-T	88	35
IVS-2	35	C-T-C-A-A	80	

Consensus structure of primary signals with flanking regions

A	—	1	4	—	—	2	10	1	1	4	—	1
G	3	—	3	—	—	5	—	—	—	1	—	5
C	5	5	1	9	—	3	—	7	4	4	—	4
T	2	4	2	1	10	—	—	2	5	1	10	—
C-T-G-A-C												

Redundant signals in 3 of the 10 introns

β -Globin IVS-1	45	C-T-G-A-T	88	32
δ -Globin IVS-1	43	C-T-G-A-T	88	33
ϵ -Globin IVS-1	29	C-T-A-A-T	80	35

Headings as in Tables 1 and 2. IVS, intervening sequence, intron.

the sequences of the deletion mutants shows that in each case an upstream cryptic 3' splice signal had been inadvertently moved into an active position close to the 3' splice site. However, there is in the literature the case of a mutant mouse immunoglobulin gene in which splicing joins the first exon to the third exon, bypassing the 3' splice site of the first intron (39). Inspection of the sequence shows that the bypassed 3' splice site is normal, but the normal 3' splice signal is missing and no alternatives are found. This mutant provides suggestive evidence that a 3' splice signal is required in the mouse.

It has been noted in many cases that there are multiple examples of good 3' splice sites in the interior of an intron (for example, ref. 40). We have examined a number of such cases and have found that the non-functional 3' splice sites have only remote or poor 3' splice signals, if any at all.

Direct evidence for a role for the 3' splice signal now comes from the demonstration by Ruskin *et al.* (12) that the A at position 4 of the signal accepts the 5' terminal G in the splicing mechanism.

Table 4. Consensus structures of 3' splice signals of *Drosophila* and chicken and duck

<i>Drosophila</i>											
A	5	9	12	—	—	18	24	1	4	4	6
G	5	7	3	—	1	5	—	—	3	4	6
C	9	4	2	17	7	—	—	7	9	4	6
T	5	4	7	7	16	1	—	16	9	12	6
C-T-A-A-T											

Chicken and duck

A	5	15	12	1	—	17	40	4	4	6	7
G	13	5	9	1	—	17	—	4	8	8	11
C	10	8	9	28	—	8	2	18	16	17	7
T	14	14	12	12	42	—	—	16	14	11	17

C-T-G-A-C
A-T

Primary signals, one from each intron searched, were aligned with their flanking regions and the consensus structure obtained.

The fact that the 3' splice signal in yeast is much more conserved than in higher organisms may be related to the fact that only a few yeast genes have introns (3, 4). Perhaps in the evolution of yeast there has been a selective deletion of introns with weak 3' splice signals.

Sequences in U2 snRNA Complementary to the 3' Splice Signal and Site. A 3' splice signal with five conserved bases together with the $\overset{\text{C}}{\text{T}}$ -A-G at the 3' intron boundary could constitute a two-part recognition sequence. Since the 5' splice site is recognized by U1 snRNP (42), we hypothesized that this two-part 3' signal could be recognized by another U snRNP. We found that sequences complementary to the rat 3' signals C-T-G-A-T and C-A-G are indeed present in the rat U2 snRNA (14) in a portion of the molecule that we calculate should be single stranded in the secondary structure (15). (T is used here for simplicity.)

	53												41	
U2 snRNA	- G	A	C	T	A	T	T	C	T	T	G	T	C	-
Intron	- C	T	G	A	T	-	-	-	-	-	C	A	G	-

We propose that the 3' splice signal and the terminal $\overset{\text{C}}{\text{T}}$ -A-G are bound to the U2 snRNA with the aid of the protein of the U2 snRNP and that the variable length between the two binding regions would loop out where we have placed dashes. The variation in binding of the two 3' regions with U2 snRNA would be comparable to that of 5' splice sites with U1 snRNA. In this model, the pyrimidine-rich region of the 3' splice site would not be involved in base-specific binding but could play some other essential role in the recognition process.

Though Padgett *et al.* (42) found that antiserum directed against U2 snRNP did not inhibit splicing *in vitro*, these experiments do not exclude the possibility of U2 snRNP involvement. The fact that the sequence of U2 snRNA between 41 and 53 is highly conserved in various organisms (14) implies a base-specific role for this region. Furthermore, U2 snRNA is found to be hybridized *in vivo* to heterogeneous nuclear RNA (43) in the same way as U1 snRNA (44). U1 snRNP and U2 snRNP may form a complex within which splicing intermediates are securely held.

We express our appreciation to J. T. Lis and J. Schuster for the Makecons computer program, to Lori Silvern for valuable assistance, to G. Nemethy for the statistical analysis, to T. Maniatis, R. A. Padgett, and J. M. Calvo for communicating unpublished work, and to all of the many individuals who obtained the sequence data we employed in this study. This study was supported by National Institutes of Health Grant GM10791.

1. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349-384.
2. Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459-472.
3. Langford, C. J. & Gallwitz, D. (1983) *Cell* **33**, 519-527.
4. Pikielny, C. W., Teem, J. L. & Rosbash, M. (1983) *Cell* **34**, 395-403.
5. Langford, C. J., Klinz, F. J., Donath, C. & Gallwitz, D. (1984) *Cell* **36**, 645-653.
6. Watts, F., Castle, C. & Beggs, J. (1983) *EMBO J.* **2**, 2085-2091.
7. Woudt, L. P., Pastink, A., Kempers-Veenstra, A. E., Jansen, A. E. M., Mager, W. H. & Planta, R. J. (1983) *Nucleic Acids Res.* **11**, 5347-5360.

8. Kinnaird, J. H. & Fincham, J. R. S. (1983) *Gene* **26**, 253-260.
9. Langford, C. J., Nellen, W., Niessing, J. & Gallwitz, D. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1496-1500.
10. Mount, S. M. (1983) *Nature (London)* **304**, 309-310.
11. Grabowski, P. J., Padgett, R. A. & Sharp, P. A. (1984) *Cell* **37**, 415-427.
12. Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. (1984) *Cell* **38**, 317-331.
13. Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F. & Sharp, P. A. (1984) *Science* **225**, 898-903.
14. Reddy, R. & Busch, H. (1983) *Prog. Nucleic Acid Res. Mol. Biol.* **30**, 127-162.
15. Branlant, C., Krol, A., Ebel, J.-P., Lazar, E., Haendler, B. & Jacob, M. (1982) *EMBO J.* **1**, 1259-1265.
16. Schuler, M. A., McOsker, P. & Keller, E. B. (1983) *Mol. Cell. Biol.* **3**, 448-456.
17. Zakut, R., Shani, M., Givol, D., Neuman, S., Yaffe, D. & Nudel, U. (1982) *Nature (London)* **298**, 857-859.
18. Nudel, U., Zakut, R., Shani, M., Neuman, S., Levy, Z. & Yaffe, D. (1983) *Nucleic Acids Res.* **11**, 1759-1771.
19. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. & Tizard, R. (1979) *Cell* **18**, 545-558.
20. Moorman, R. J. M., den Dunnen, J. T., Mulleners, L., Andreoli, P., Bloemendal, H. & Schoenmakers, J. G. G. (1983) *J. Mol. Biol.* **171**, 353-368.
21. Scarpulla, R. A., Agne, K. M. & Wu, R. (1981) *J. Biol. Chem.* **256**, 6480-6486.
22. King, C. R. & Piatigorsky, J. (1983) *Cell* **32**, 707-712.
23. Kawakami, T., Takahashi, N. & Honjo, T. (1980) *Nucleic Acids Res.* **8**, 3933-3945.
24. Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. & Hodge, L. (1980) *Cell* **20**, 313-319.
25. Max, E. E., Seidman, J. G. & Leder, P. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3450-3454.
26. Yamawaki-Kataoka, Y., Miyata, T. & Honjo, T. (1981) *Nucleic Acids Res.* **9**, 1365-1381.
27. Nishioka, Y. & Leder, P. (1980) *J. Biol. Chem.* **255**, 3691-3694.
28. Pech, M., Hoecht, J., Schnell, H. & Zachau, H. G. (1981) *Nature (London)* **291**, 668-670.
29. Bernard, O., Hozumi, N. & Tonegawa, S. (1978) *Cell* **15**, 1133-1144.
30. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. (1980) *Nature (London)* **286**, 676-683.
31. Liebhafner, S. A., Goossens, M. J. & Kan, Y. W. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7054-7058.
32. Lawn, R. M., Efstratiadis, A., O'Connell, C. & Maniatis, T. (1980) *Cell* **21**, 647-651.
33. Spritz, R. A., DeRiel, J. K., Forget, B. G. & Weissman, S. M. (1980) *Cell* **21**, 639-646.
34. Slightom, J. L., Blechl, A. E. & Smithies, O. (1980) *Cell* **21**, 627-638.
35. Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J. (1980) *Cell* **21**, 621-626.
36. Wallace, J. C. & Edmonds, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 950-954.
37. Wickens, M. P. & Gurdon, J. B. (1983) *J. Mol. Biol.* **163**, 1-26.
38. Thimmappaya, B. & Shenk, T. (1979) *J. Virol.* **30**, 668-673.
39. Seidman, J. G. & Leder, P. (1980) *Nature (London)* **286**, 779-783.
40. Staden, R. (1984) *Nucleic Acids Res.* **12**, 505-519.
41. Mount, S. M., Pettersson, I., Hinterberger, M., Karmas, A. & Steitz, J. (1983) *Cell* **33**, 509-518.
42. Padgett, R. A., Mount, S. M., Steitz, J. A. & Sharp, P. A. (1983) *Cell* **35**, 101-107.
43. Calvet, J. P., Meyer, L. & Pederson, T. (1982) *Science* **217**, 456-458.
44. Calvet, J. P. & Pedersen, T. (1981) *Cell* **26**, 363-370.
45. Nudel, U., Calvo, J. M., Shani, M. & Levy, Z. (1984) *Nucleic Acids Res.* **12**, 7175-7186.