

Hydroxyl radical "footprinting" of RNA: Application to pre-mRNA splicing complexes

(spliceosome/small nuclear ribonucleoprotein)

XIAODONG WANG AND RICHARD A. PADGETT*

Department of Biochemistry, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75235

Communicated by Phillip A. Sharp, July 20, 1989

ABSTRACT We present an adaptation of the hydroxyl radical DNA "footprinting" technique that permits high-resolution mapping of protected regions of RNA. Hydroxyl radical cleaves RNA independently of base sequence and secondary structure of the RNAs examined and allows resolution of protected regions at the single nucleotide level. By using this technique, we show that several regions of the 3' splice site of mRNA precursors are protected during the formation of splicing-specific ribonucleoprotein complexes in an *in vitro* splicing system. These regions include the 3' intron/exon junction and a portion of the adjacent exon, the polypyrimidine tract, and the site of branch formation. These protections appear to be due to splicing specific complexes since their formation is sensitive to point mutations at crucial residues and requires ATP and incubation. The formation of these protected regions is independent of the presence of a 5' splice site.

The ability to examine protein-double-stranded DNA interactions has been facilitated by the techniques of detecting DNA protections using the DNase I "footprinting" assay originally described by Galas and Schmitz (1). Since then, various chemical cleavage strategies have been developed to probe these interactions in greater detail (2-5). The hydroxyl radical technique of Tullius and Dombroski (3) uses a different strategy in which hydroxyl radicals are produced upon reduction of H₂O₂ by an EDTA-iron complex that is not bound to the DNA. The hydroxyl radical cleaves DNA at positions that are accessible to solvent and does not cleave at positions involved in protein-DNA interactions. Due to the small size and neutral charge of the hydroxyl radical, it serves as a high-resolution, sequence-independent probe of protein-DNA interactions and DNA structure (2).

In contrast to double-stranded DNA, the techniques for high-resolution mapping of RNA-protein interactions are less well developed. A major difficulty encountered with the use of enzymes to detect specific interactions is the effect of RNA secondary structure on the cleavage pattern. In addition, chemical techniques that depend on intercalation of reagents into helical structures (4, 5) are obviously unsuitable for single-stranded RNA. In light of these problems, the hydroxyl radical technique offers some major benefits. (i) The cleaving species is very small and uncharged and would be expected to be able to penetrate loosely structured regions. (ii) It is not specific to cleavage at particular bases. (iii) The cleavage chemistry is compatible with most buffer components and is active over a wide range of pH, salt, and temperature conditions.

Here we demonstrate that hydroxyl radical cleavage of RNA shares with DNA cleavage the properties of sequence and structure independence and that it can be used to define RNA-protein interactions at high resolution. The system we

have investigated is the formation of mRNA splicing-specific complexes at the 3' splice site regions of synthetic mRNA precursors. The regions of protection have been mapped with much greater precision than has been previously possible using ribonucleases and demonstrate that the 3' splice site sequences shown to be important for splicing by mutational analysis are involved in specific interactions during the earliest part of the reaction.

MATERIALS AND METHODS

Transcription and Labeling of RNA. Construction pSPAL-5 was derived from pBSAL-4 (6) and consists of the *Bam*HI to *Eco*RI fragment of the rabbit β -globin gene inserted into the plasmid pSP64. The *Eco*RI site was converted to a *Sca* I site using commercial linkers. The 426-nucleotide (nt) transcript consists of a 53-nt upstream exon, a 319-nt intron, and a 54-nt downstream exon.

Construction pRSP-7 (7) contains the 3' 50 nts of the first intron and the second exon of the adenovirus 2 major late transcription unit. Pre-mRNA transcribed from this template contains only 3' splice site signals. The 131-nt transcript contains 93 nts of vector and intron sequence followed by 38 nts of exon 2 sequence.

The transcription templates were cleaved with *Sca* I, and RNA was transcribed by SP6 RNA polymerase in the presence of unlabeled nucleoside triphosphates and GpppG to give 5' capped pre-mRNA as described (8). The RNA was then labeled at the 3' end in a 15- μ l volume containing 50 mM Hepes buffer (pH 7.6), 20 mM MgCl₂, 10 μ g of bovine serum albumin per ml, 6 μ M ATP, 3 mM dithiothreitol, 10% dimethyl sulfoxide, 50 μ Ci of 3',5'-[5'-³²P]cytidine bisphosphate (3000 Ci/mmol; 1 Ci = 37 GBq; DuPont/NEN), and 6 units of T4 RNA ligase (New England Biolabs) for 2 hr at 4°C. The RNA was then purified by preparative polyacrylamide gel electrophoresis and eluted as described (9).

Sca I-cleaved DNA was used because SP6 RNA polymerase produces runoff transcripts with heterogeneous 3' ends from DNA cleaved by most other restriction endonucleases. This is seen for enzymes producing both blunt and staggered ends (data not shown). We have found, however, that DNA cleaved with *Sca* I produces homogeneous RNA transcripts.

Splicing Reactions. Splicing complexes were formed on the substrate RNA essentially as described by Konarska and Sharp (10). The final concentrations in the 40- μ l reaction mixtures were 1.5 mM ATP, 5 mM creatine phosphate, 2.25 mM MgCl₂, and 35% HeLa cell nuclear extract prepared as described (11). Titration of RNA and extract concentration was necessary to optimize complex formation. Incubations were carried out at 30°C for 30 min except where indicated. In order to remove glycerol, which interferes with the footprinting chemistry, the splicing reaction mixtures were cen-

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: nt, nucleotide; snRNP, small nuclear ribonucleoprotein.

*To whom reprint requests should be addressed.

trifuged through Bio-Rad P-60 columns [1-ml bed volume in recycled Quick Spin columns (Boehringer Mannheim)] equilibrated with 35 mM KCl, 2.25 mM MgCl₂, and 7 mM Hepes (pH 7.6). Analysis of splicing complexes by gel electrophoresis was as described (10).

Footprinting Reactions. Hydroxyl radical cleavage of the RNA was carried out as described for DNA by Tullius and Dombroski (3). The hydroxyl radical cleavage reagent was prepared by mixing 30 μ l of a freshly prepared solution of 0.4 mM ammonium iron sulfate (Aldrich) and 0.8 mM EDTA, 30 μ l of 20 mM sodium ascorbate, and 30 μ l of 0.6% H₂O₂. The 40- μ l splicing reaction mixture was mixed with 70 μ l of 16 mM Hepes buffer (pH 7.6), 80 mM KCl, 5 mM MgCl₂, 16 μ M EDTA, and 0.4 mM dithiothreitol and added to the 90- μ l cleavage reagent solution. The reaction mixtures were incubated at room temperature for 90 sec and reactions were stopped with the addition of thiourea to 10 mM. The cleaved RNA was then extracted with phenol/chloroform (1:1) and ethanol precipitated, and equal amounts of radioactivity in each lane were loaded onto an 8% polyacrylamide, 0.4-mm-thick sequencing gel containing 8.3 M urea, 45 mM Tris-borate (pH 8.3), and 1 mM EDTA. Control reactions were carried out with the labeled RNA plus 0.7 mg of RNase-free bovine serum albumin (Pharmacia) per ml in the same buffer as the splicing reactions using one-half volume of the cleavage reagent.

The extent of cleavage was controlled by varying the amount of the iron/EDTA/ascorbate/H₂O₂ solution added to the splicing reaction. The optimum extent of cleavage for observing the footprints was much less than one cleavage per RNA chain. Higher levels of cleavage led to a loss of the footprint signal.

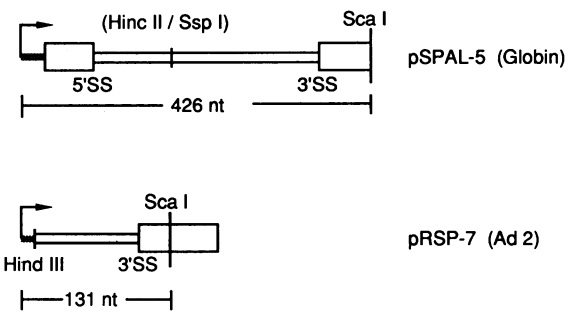


FIG. 1. Structure of transcription templates. The origin and construction of the transcription templates are described in the text. The thick boxes represent exons, the thin boxes represent introns, and the shaded bars represent vector sequences. The sites of initiation and the direction of transcription are denoted by the arrows and the size of the transcript is indicated below each structure. Ad2, adenovirus 2.

RESULTS

RNA Substrates. The pre-mRNA substrates used in this work are shown in Fig. 1. Construction pSPAL-5 is derived from the rabbit β -globin gene and contains parts of exons two and three and the intron between them. Construction pRSP-7 is derived from the first and second leader exon region of the adenovirus 2 major late transcript and contains the 3' portion of the first intron and the second exon.

Hydroxyl Radical Cleavage of RNA. To assess the uniformity of the hydroxyl radical cleavage reaction on free RNA, 3' end-labeled RNA transcribed from pSPAL-5 was treated

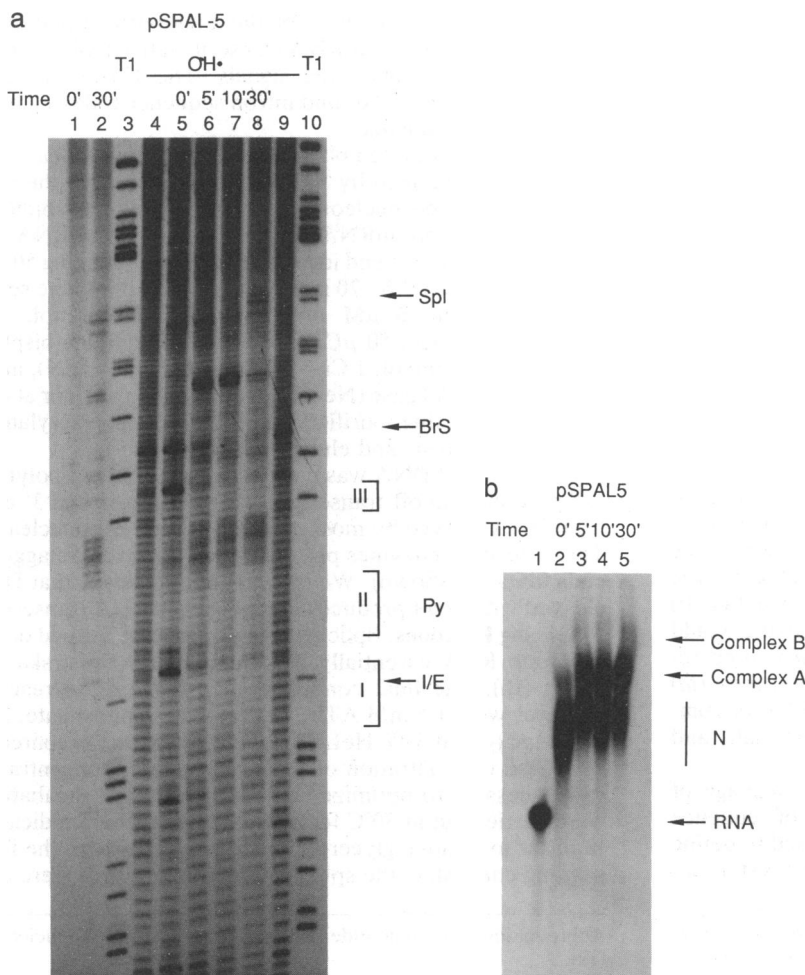


FIG. 2. Hydroxyl radical footprinting of the 3' splice site of rabbit β -globin pre-mRNA. (a) Pre-mRNA transcribed from pSPAL-5 and labeled at the 3' end was incubated with HeLa cell nuclear extract under splicing conditions for the times indicated and subjected to hydroxyl radical footprinting. The RNA was extracted and analyzed on a denaturing 8% polyacrylamide gel. Lanes 1 and 2, RNA extracted from the splicing reaction at 0 and 30 min and analyzed directly; lanes 4 and 9, RNA from control reactions in which RNA was treated with hydroxyl radical in the presence of bovine serum albumin; lanes 3 and 10, RNA partially digested with RNase T1; lanes 5-8, RNA from incubations with HeLa cell nuclear extract under splicing conditions for the times indicated. The regions of protection from hydroxyl radical cleavage are indicated by the numbered brackets. Spl, position of spliced RNA; BrS, branch site; Py, polypyrimidine tract; I/E, 3' intron/exon junction. (b) Gel retardation analysis of the spliceosome formation reaction on pSPAL-5 RNA. The substrate RNA was incubated with HeLa cell nuclear extract under splicing conditions. Samples were removed and treated with heparin, and the complexes were separated on a non-denaturing 4% polyacrylamide gel. Lane 1, RNA by itself; lanes 2-5, from the spliceosome formation reaction sampled at the times indicated. The positions of complexes A and B and free RNA are indicated. N, nonspecific complexes that form initially.

with hydroxyl radical and the products were separated on a denaturing polyacrylamide sequencing gel. As shown in Fig. 2a, lanes 4 and 9, in the absence of binding factors, this RNA is cleaved at every position with only small differences in efficiency. Thus, neither nucleotide sequence nor RNA secondary structure materially affects the cleavage of this RNA by hydroxyl radical.

Detection of RNA-Factor Interactions in Spliceosomes. To demonstrate the ability of the hydroxyl radical cleavage reaction to detect specific RNA-factor interactions, the effect on cleavage of forming splicing-specific complexes on pre-mRNA substrates was examined. Pre-mRNA labeled at the 3' end was incubated with HeLa cell nuclear extract under splicing conditions and cleaved with hydroxyl radical. Fig. 2a shows the pattern of cleavage of pSPAL-5 RNA incubated with the HeLa extract for the times indicated. Only the region encompassing the 3' splice site is shown here. The distance of the protections from the labeled 3' end is determined by comparing the RNase T1 partial digest of the substrate RNA to the sequence of this region (12).

Several regions of the RNA become protected from hydroxyl radical cleavage during incubation in the nuclear extract. Region I covers the 3' intron/exon junction and includes the last 3 nts of the intron and about 5 nts of the downstream exon. Region II includes a portion of the polypyrimidine region of the 3' splice site. Region III is located between the polypyrimidine tract and the branch acceptor site. The nucleotides between regions II and III are reproducibly cleaved during incubation in the *in vitro* splicing reaction in the presence of ATP (see below). Thus, although protection of this region from hydroxyl radical cleavage may be masked by adventitious cleavages in the extract, a more likely explanation is that this region is specifically unprotected in the splicing complex. In addition to the highly protected regions noted above, there is weaker protection of the region around the branch site and about 10 nts 5' of the branch site.

The protections of regions I and III and, to a lesser extent, region II develop during incubation in HeLa nuclear extract. To try to correlate the appearance of these protections to the formation of splicing-specific complexes, the same samples were analyzed on a nondenaturing polyacrylamide gel as shown in Fig. 2b. Three complexes can be detected by this analysis: the splicing-specific complexes A and B (nomenclature of ref. 10) and the nonspecific complex N (which may also contain unresolved splicing-specific complexes). The formation of the hydroxyl radical protections correlates best with the formation of complex A. By 30 min, the majority of the pre-mRNA is in complex A or complex N, whereas very little pre-mRNA is found in complex B and only a small amount of spliced RNA has been formed as shown in Fig. 2a. However, the formation of complex A is not quantitative, whereas most of the pre-mRNA must be protected in order to observe hydroxyl radical protection. Complex A has been shown to contain U2 small nuclear ribonucleoprotein (snRNP) and to show protection of the branch site region from RNase T1 (10). Therefore, the protections of regions I, II, and III may precede the addition of U2 snRNP and the formation of complex A. This may explain the low level of protection of the branch site of the β -globin pre-mRNA.

The formation of the splicing-specific complexes A and B requires the presence of ATP (10). The effect of ATP on the hydroxyl radical protections near the 3' splice site of pSPAL-5 RNA after incubation with HeLa cell nuclear extract for 30 min is shown in Fig. 3. Lanes 1-3 were not treated with hydroxyl radical, whereas lanes 5-7 are the corresponding samples treated with hydroxyl radical. Protections of regions I and III, and to a lesser extent, region II are dependent on the presence of ATP in the reaction. This again suggests a

relationship between the formation of splicing complexes and the hydroxyl radical protections.

To determine if these protections are unique to this pre-mRNA or could be detected at other 3' splice sites, RNA from the adenovirus 2 construction pRSP-7 was subjected to hydroxyl radical footprinting under similar conditions. Fig. 4 shows the protections observed with this RNA in the 3' splice site region. Lane 4 shows cleavage of the RNA alone and again demonstrates that the cleavage is nearly uniform at each nucleotide position. Lane 7 shows a partial base hydrolysis of this RNA for comparison. Lane 5 shows the hydroxyl radical footprint pattern seen after incubation with nuclear extract under splicing conditions. Lane 1 is a control reaction not treated with hydroxyl radicals. The distance of the protections from the labeled 3' end is determined by comparing the RNase T1 partial digest with the sequence of this region (9). Three regions of protection are observed in the presence of ATP that are strikingly similar in position and size to the protections observed with the rabbit β -globin substrate RNA. As with the globin substrate, protection of these three regions is dependent on the presence of ATP and incubation at 30°C (data not shown). Another protected region (marked with a dotted bracket in Fig. 4) is seen 40-50

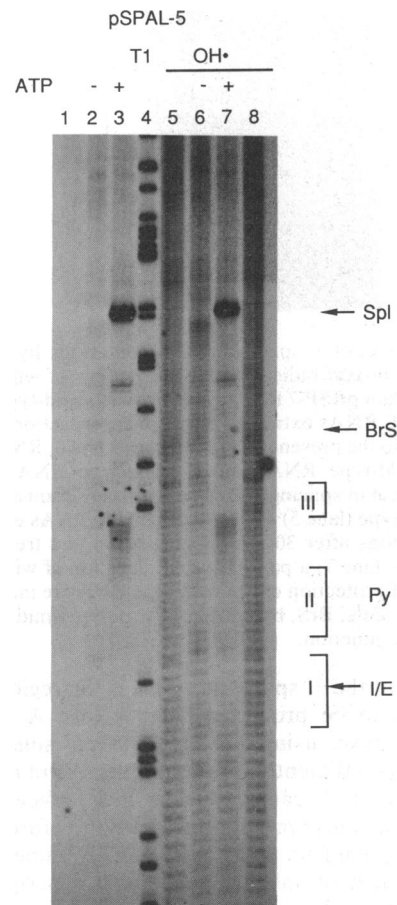


FIG. 3. ATP dependence of hydroxyl radical footprints on pSPAL-5 RNA. Lane 1, RNA before incubation; lane 2, RNA extracted from a 30-min splicing reaction without ATP; lane 3, RNA extracted from a 30-min splicing reaction containing ATP; lane 4, size markers produced by partial digestion of the RNA with RNase T1; lanes 5 and 8, RNA treated with hydroxyl radical in the presence of bovine serum albumin; lane 6, RNA from a 30-min splicing reaction without ATP treated with hydroxyl radical; lane 7, RNA from a 30-min splicing reaction containing ATP treated with hydroxyl radical. The regions of protection of the RNA are indicated by the numbered brackets. BrS, branch site; Py, polypyrimidine tract; I/E, 3' intron/exon junction.

taining the same adenovirus 3' splice site region used here have been shown to form secondary structures (16). That we do not detect protected regions in the absence of splicing extract suggests that the secondary structures at 3' splice sites are sufficiently loose to permit access by hydroxyl radical. (ii) RNAs without splice site sequences (data not shown) or with mutations that drastically reduce splicing (Fig. 4) do not show protection even though they are associated with proteins in nonspecific complexes in the nuclear extract as detected by gel mobility shift analysis (10). (iii) Some of these protections depend on the conditions of the assay, such as the presence of ATP and incubation at 30°C.

The evidence that the protections we observe are due to splicing-specific factors is based on the following observations. (i) The 3' splice site regions from two unrelated genes show very similar patterns of protection. (ii) A specific base change known to drastically affect splicing causes the loss or attenuation of all protections seen on the wild-type RNA. (iii) Some of the protections are dependent on ATP and incubation at 30°C as is spliceosome formation. (iv) The regions of the 3' splice site that are protected in the course of the splicing reaction correlate well with regions of sequence conservation as well as regions shown to be required for splicing by mutational analysis (24).

Hydroxyl Radical Footprints Correlate with Functional Regions of the 3' Splice Site. The 3' splice site region of mRNA precursors includes the conserved CAG/G sequence at the intron/exon junction, the pyrimidine-rich region upstream of this, and the site of branch formation. Mutations or deletions of each of these elements cause reduced splicing efficiency and often the activation of cryptic 3' splice sites. These elements, therefore, probably represent either sites of interaction between the pre-mRNA substrate and the cellular splicing factors or sequences involved in forming specific RNA structures required for splicing.

A number of cellular factors have been shown to interact with these elements during the splicing reaction. Two proteins, the intron binding protein (17, 18) and the heterogeneous nuclear ribonucleoprotein C proteins (19), appear to bind to the polypyrimidine region. Another protein, U2AF, appears to interact with the branch site sequence as well as the polypyrimidine tract (20). The U2 snRNP has been proposed to bind to the branch site (21), whereas U5 may interact with the polypyrimidine tract or the intron/exon junction (22). Which, if any, of these factors is responsible for the protections observed here is not known. However, the combination of biochemical fractionation of the crude splicing extract and the hydroxyl radical footprinting technique will allow the correlation of specific factors and protection patterns.

Much of the information regarding the sites of binding of various factors to splice site regions is derived from RNase protection experiments in which labeled pre-mRNA substrates are incubated in crude *in vitro* splicing reactions and digested with RNase followed by immunoprecipitation with antibodies against snRNPs. The regions of the pre-mRNA protected against RNase digestion and bound to snRNPs are then analyzed. The resolution of this analysis is limited by the location of RNase cleavage sites and the macromolecular nature of the RNase. Quantitation of RNase protection is

difficult, which makes it difficult to determine if the observed interactions are significant. In addition, interactions that depend on largely intact pre-mRNA or small nuclear RNAs or that do not involve immunoprecipitable snRNPs would not be detected by this approach.

The chemical footprinting technique avoids many of these difficulties. It has high resolution and only detects interactions that involve the majority of the RNAs in the reaction. Additional advantages of the technique are that it is rapid and compatible with the conditions of the splicing reaction. Since each substrate RNA molecule is cleaved less than once on average, this technique can potentially detect interactions requiring intact substrate RNA or small nuclear RNAs that would be lost using the RNase protection method cited above.

The excellent technical assistance of S. Burke in cell culture and extract preparation is gratefully acknowledged. This work was supported by grants NP-569 and IN-142 from the American Cancer Society and Grant I-1035 from the Robert A. Welch Foundation. X.W. is supported by the Robert A. Welch Foundation. R.A.P. is a Pew Scholar in the Biomedical Sciences.

- Galas, D. J. & Schmitz, A. (1978) *Nucleic Acids Res.* **5**, 3157-3170.
- Tullius, T. D. (1988) *Nature (London)* **332**, 663-664.
- Tullius, T. D. & Dombroski, B. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5469-5473.
- Spassky, A. & Sigman, D. S. (1985) *Biochemistry* **24**, 8050-8056.
- Hertzberg, R. P. & Dervan, P. B. (1982) *J. Am. Chem. Soc.* **104**, 313-315.
- Lamond, A. I., Konarska, M. M. & Sharp, P. A. (1987) *Genes Dev.* **1**, 532-543.
- Konarska, M. M., Padgett, R. A. & Sharp, P. A. (1985) *Cell* **42**, 165-171.
- Konarska, M. M., Padgett, R. A. & Sharp, P. A. (1984) *Cell* **38**, 731-736.
- Grabowski, P. J., Padgett, R. A. & Sharp, P. A. (1984) *Cell* **37**, 415-427.
- Konarska, M. M. & Sharp, P. A. (1986) *Cell* **46**, 845-855.
- Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. (1983) *Nucleic Acids Res.* **11**, 1475-1489.
- van Ooyen, A., van den Berg, J., Mantei, N. & Weissman, C. (1979) *Science* **206**, 337-344.
- Krämer, A. (1987) *J. Mol. Biol.* **196**, 559-573.
- Keller, E. B. & Noon, W. A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7417-7420.
- Green, M. R. (1986) *Annu. Rev. Genet.* **20**, 671-708.
- Hall, K. B., Green, M. R. & Redfield, A. G. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 704-708.
- Gerke, V. & Steitz, J. A. (1986) *Cell* **47**, 973-984.
- Tazi, J., Alibert, C., Tamsamani, J., Reveillaud, I., Cathala, G., Brunel, C. & Jeanteur, P. (1986) *Cell* **47**, 755-766.
- Choi, Y. D., Grabowski, P. J., Sharp, P. A. & Dreyfuss, G. (1986) *Science* **231**, 1534-1539.
- Ruskin, B., Zamore, P. D. & Green, M. R. (1988) *Cell* **52**, 207-219.
- Black, D. L., Chabot, B. & Steitz, J. A. (1985) *Cell* **42**, 737-750.
- Chabot, B., Black, D. L., LaMaster, D. M. & Steitz, J. A. (1985) *Science* **230**, 1344-1349.
- Latham, J. A. & Cech, T. R. (1989) *Science* **245**, 276-282.
- Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55**, 1119-1150.