Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 6253–6257, August 1990
Biochemistry

Mechanism for cryptic splice site activation during pre-mRNA splicing

(thalassemia/U1 small nuclear ribonucleoprotein particle/5′ splice site selection)

KRISTIN K. NELSON AND MICHAEL R. GREEN†

Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138

Communicated by Raymond L. Erikson, May 30, 1990 (received for review March 30, 1990)

ABSTRACT **The 5′ splice site of a pre-mRNA is recognized by U1 small nuclear ribonucleoprotein particles (snRNP) through base pairing with the 5′ end of U1 small nuclear RNA (snRNA). Single-base substitutions within a 9-nucleotide 5′-splice-site sequence can abolish or attenuate use of that site and, in higher eukaryotes, can also activate nearby "cryptic" 5′ splice sites. Here we show that the effects of single-base substitutions within a 5′ splice site can be completely or partially suppressed by cis mutations that improve the overall complementarity of the site to U1 snRNA. We further show that in the presence of the normal 5′ splice site, a cryptic 5′ splice site can be activated by increasing its complementarity to U1 snRNA. U1 snRNP binding experiments confirm that cryptic 5′ splice sites are activated when their affinity for U1 snRNP approaches that of the authentic 5′ splice site. Based upon these results, we propose a spliceosome competition model for 5′-splice-site selection and cryptic 5′-splice-site activation. We discuss our results with regard to the factors involved in 5′-splice-site recognition.**

During pre-mRNA splicing, intron sequences are removed from the original RNA transcript (for review, see refs. 1–4). Three sequence elements that direct this RNA processing have been identified: the two sequence elements at the intron borders and the branch site upstream of the 3′ splice site. The 5′ splice site is a 9-nucleotide (nt) sequence. The consensus, 5′-CAG/GUAAGU-3′, is complementary to the first 9 nt of the 5′ end of U1 small nuclear RNA (snRNA) and base pairing of these two RNAs is a necessary step in pre-mRNA splicing (5). Natural 5′ splice sites match the consensus to various extents, the vast majority deviating at two or three positions (6–8). The most invariant positions are the almost universal GU dinucleotide at the cleavage site and the guanosine at position +5, which is present in 85% of the sequences analyzed.

Splicing occurs in a large complex, the spliceosome, which is composed of U1, U2, U5, and U4/6 small nuclear ribonucleoprotein particles (snRNP) and multiple proteins (for review, see refs. 1–4). It is believed that during spliceosome assembly the splice sites are selected. An initial step in assembly is binding of U1 snRNP to the 5′ splice site (9–13). The U1 snRNP binding reaction does not require ATP and can occur on RNA substrates that contain only a 5′ splice site (11). Binding of U2 snRNP to the branch site is another early event. This binding reaction requires ATP hydrolysis and can occur on substrates that lack the 5′ splice site (9, 11, 14, 15). Thus, in higher eukaryotes, the two halves of the intron appear to be independently assembled into partial complexes, which are then brought into proximity.

Human β-globin thalassemia is a disease that results from a deficiency of β-globin polypeptide. At the molecular level the basis of the disease is diverse (see ref. 16 and references within). One class of thalassemic lesions is single-base substitutions within the 5′ splice site of the human β-globin first intron (IVS*1*). These mutations reduce or eliminate use of the authentic 5′ splice site, while activating three "cryptic" 5′ splice sites (Fig. 1*A*). The three cryptic 5′ splice sites are located 13–38 nt from the authentic 5′ splice site and each resembles the 5′-splice-site consensus.

At least two possibilities can be offered to explain the β-globin thalassemic splicing phenotype. First, the decreased use of the 5′ splice site and concomitant activation of cryptic 5′ splice sites may be due only to the reduced affinity of the mutant site for U1 snRNP. This reduced affinity is the result of decreased complementarity to U1 snRNA. Alternatively, the effects of the mutations could result from the presence of nonconsensus bases at specific critical positions, perhaps those that show the least variability. Mutations at such critical positions may, for example, prevent binding of a splicing factor(s) other than U1 snRNP.

A prediction of the first possibility is that the effects of the thalassemic mutations can be compensated for by additional substitutions in cis that increase complementarity to U1 snRNA. Here we test this prediction and discuss our results with regard to the mechanisms involved in 5′-splice-site selection.

MATERIALS AND METHODS

RNA Substrates. The wild-type β-globin (pSP64HβΔ6), the thalassemics (P1 and P5; ref. 17), and 5′ → C (18) have been described. CP1 and CP5 were generated from 5′ → C, and −38* was from pSP64HβΔ6 by using the Amersham *in vitro* site-directed mutagenesis kit. RNA was transcribed *in vitro* from *Bam*HI-linearized templates with SP6 RNA polymerase.

COS Cell Transfections. The *Nco* I–*Bam*HI fragments of the above globin constructions were moved into the vector β^+-E1a (19). Each DNA (10 μg) was transfected onto 60% confluent COS cells as a calcium phosphate precipitate. The medium was changed after 24 hr, and the cells were harvested 44 hr after transfection. Cytoplasmic RNA was harvested as described (20).

RESULTS

Effects of Compensatory 5′-Splice-Site Mutations. We used two characterized thalassemic mutations, designated P1 and P5. P1 changes the sequence of the authentic β-globin IVS1 5′ splice site at position +1 of the intron, G → A (CAG/

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: snRNP, small nuclear ribonucleoprotein particle(s); snRNA, small nuclear RNA; nt, nucleotide(s); IVS*1*, first intron of human β-globin gene.

†Present address: Program of Molecular Medicine, University of Massachusetts Medical Center, 373 Plantation Street, Worcester, MA 01605.

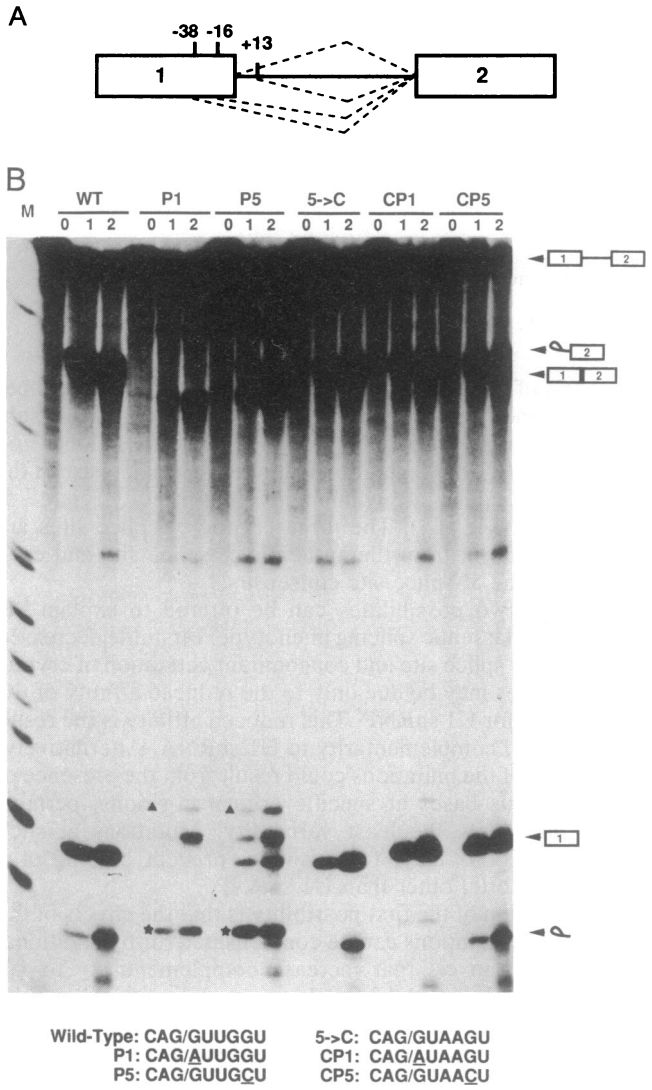

FIG. 1. (*A*) Diagram of the human β-globin IVS*1* and the cryptic 5′ splice sites activated by thalassemic mutations. Boxes represent exon sequences; introns are designated by lines. The carets designating use of the cryptic sites are shown below, and the caret designating use of the authentic site is shown above the diagram of the pre-mRNA. (*B*) Suppression of thalassemic mutations analyzed by *in vitro* splicing. Uniformly ^{32}P-labeled transcripts were incubated with nuclear extract under splicing conditions (17) for the time indicated. The products were separated on a 5% denaturing polyacrylamide gel. First exon species generated by use of the −16 cryptic 5′ splice site are designated with a star, those generated by use of the +13 cryptic are designated by a triangle. Species generated by use of the authentic site are diagrammed to the right. The 5′-splice-site sequences are listed below. CP1 and CP5 first exon species are larger than that of the wild-type due to use of a different vector that contains an additional 2 nt between the start of SP6 transcription and the site of insertion.

GUUGGU to CAG/AUUGGU); P5 alters position +5 of the intron, G → C (CAG/GUUGGU to CAG/GUUGCU). *In vivo*, both mutations activate three cryptic 5′ splice sites (Fig. 1*A*) at positions −38 and −16 (within the exon) and at position +13 (within the intron) (16). Activation of the −16 and +13 cryptic sites is readily detectable *in vitro* (17). Both *in vivo* and *in vitro*, the P1 mutation completely inactivates the authentic 5′ splice site, whereas the P5 mutation significantly reduces its activity.

In the β-globin derivative 5′ → C, the normal β-globin IVS*1* 5′ splice site is altered to be perfectly complementary to U1 snRNA (18). The P1 and P5 mutations were introduced into the 5′ → C background, producing the 5′ splice sites CAG/AUAAGU (CP1) and CAG/GUAACU (CP5), respectively. These derivatives thus contain a thalassemic point mutation within a 5′ splice site that is otherwise perfectly complementary to U1 snRNA.

Uniformly labeled transcripts derived from CP1 and CP5 were spliced *in vitro* (Fig. 1*B*). Splicing of wild-type human β-globin and the thalassemic mutations P1 and P5 were compared in parallel. For the P1 and P5 substrates, use of the authentic 5′ splice site was decreased and at least two cryptic sites were activated, as expected from previous results (17). In contrast, the substrates containing the thalassemic mutations in the consensus background (CP1 and CP5) were efficiently spliced using the authentic 5′ splice site, and cryptic 5′ splice sites were not activated. Note that, in the CP1 substrate, 5′ cleavage occurs at an AU dinucleotide instead of the canonical GU dinucleotide (confirmed by primer-extension analysis of the lariat intermediate; data not shown). Consistent with previous studies (21, 22), this mutant substrate is blocked after the first step of splicing, and thus accurately spliced mRNA is not produced.

Activation of a Cryptic 5′ Splice Site in the Presence of the Normal 5′ Splice Site. The results described above prompted us to test whether a cryptic 5′ splice site could be activated by increasing its complementarity to U1 snRNA. As described above, use of the −38 cryptic site is not detectable *in vitro*. We mutated the −38 site from AAG/GUGACA to AAG/GUGAGU, thereby increasing its complementarity to U1 snRNA at two additional positions. When this −38* substrate was spliced *in vitro*, both the −38* site and the authentic site were used; the −38* site is used in slight preference to the normal 5′ splice site (Fig. 2). Similar experiments with the other cryptic sites (+13, −16) yielded comparable results (data not shown).

Cryptic 5′ Splice Site Activation Correlates with the Levels of U1 snRNP Binding. To verify that our directed substitutions altered U1 snRNP binding affinity in the predicted manner we performed RNase T1 protection/immunoprecipitation assays (23). Bound U1 snRNP protects the pre-mRNA substrate from digestion by RNase T1, and the bound protected fragment is selected by immunoprecipitation with α-Sm antiserum. For the authentic 5′ splice site, a characteristic 15-nt RNase-resistant fragment is generated.

The data in Fig. 3 show that although binding of U1 snRNP to the authentic site in the wild-type substrate is readily detectable, we did not detect binding to the authentic or cryptic sites of the P1 and P5 substrates. [We note that others (24) using more sensitive procedures have reported a very low level of binding to a P1 substrate.] However, binding of U1 snRNP to the CP1 and CP5 sites is readily detectable, although the level of binding is significantly less than to the consensus site, 5′ → C. Similarly, binding of U1 snRNP to the −38 cryptic 5′ splice site is not detectable, whereas U1 snRNP efficiently binds to the −38* site resulting in a protected fragment of approximately 12 nt. We conclude that as the overall complementarity between a 5′ splice site and U1 snRNA increases so does the level of U1 snRNP binding. From these combined results, we further conclude that mutations which increase the complementarity of a site to U1 snRNA lead to a corresponding increase in the activity of the site for splicing.

Activity of β-Globin 5′-Splice-Site Derivatives *in Vivo*. Fig. 4 presents the analysis of the splicing pattern of these same β-globin derivatives in transfected COS cells. The structures of the β-globin RNAs were determined by primer-extension analysis. In general, the relative use of 5′ splice sites *in vivo* corresponds closely to that observed *in vitro*. For example, the P1 and P5 mutants give rise to three primer-extension products of sizes consistent with use of the three cryptic sites. The authentic site is also used by the P5 substrate,

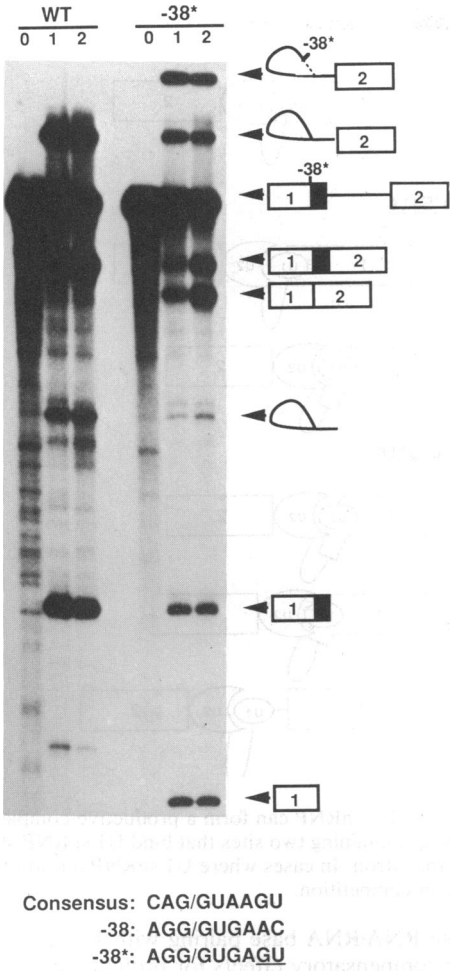

FIG. 2. Splicing time course of the −38* mutant. Splicing reactions were carried out as in Fig. 1*B*. The products were separated on a 10% denaturing polyacrylamide gel and are diagrammed to the right. The solid segment of exon 1 indicates exon sequences removed during splicing to the −38* site; the dotted line represents these same sequences as part of the intron.

consistent with the *in vitro* data (Fig. 1*B*). The 5′ → C and CP5 mutants give rise to a single product, the size of which is identical to that produced from the wild-type substrate. Finally, the −38* site is active *in vivo* (Fig. 4), as well as *in vitro* (Fig. 1*B*). *In vivo*, however, the −38* site is used exclusively, whereas *in vitro* the activities of the normal and the −38* sites are comparable.

The major difference between the *in vivo* and *in vitro* data is with regard to the CP1 mutant. Whereas *in vitro* the cryptic sites in the CP1 substrate are completely suppressed, *in vivo*

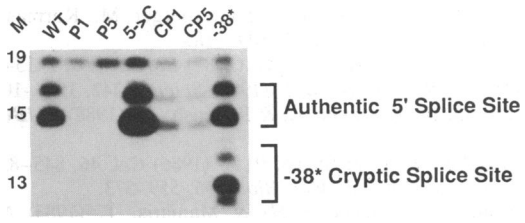

FIG. 3. RNase T1 protection/immunoprecipitations of 5′-splice-site mutants. RNA substrates were incubated under splicing conditions for 30 min and RNase T1 protection/immunoprecipitations were carried out as in Nelson and Green (18). The protected fragments were separated on a 12% denaturing polyacrylamide gel. The identities of the 5′-splice-site fragments are indicated at the right. The sizes are denoted at the left in nt.

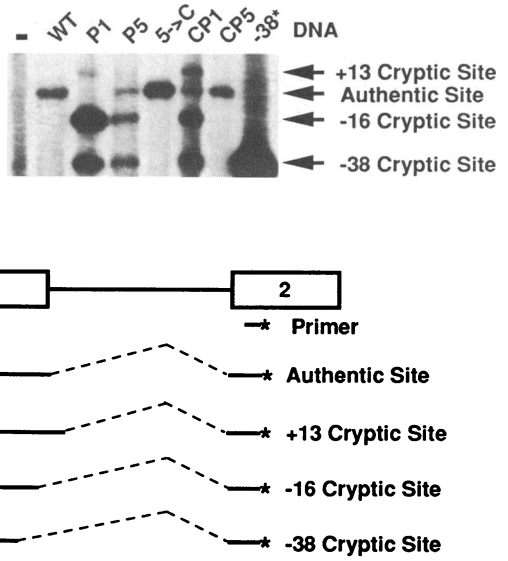

FIG. 4. Primer-extension analysis of mRNAs produced *in vivo* from mutant substrates. Cytoplasmic RNA (40 μg) harvested from COS cells transfected with the DNA listed above each lane was hybridized to a primer complementary to positions 318–337 (25) within exon 2. Primer extension analysis was carried out as in Ruskin *et al.* (26). The products are diagrammed below. Lane − indicates no DNA was transfected.

splicing occurs both to the altered CP1 site and to the cryptic sites. Thus the additional substitutions in CP1 have less of a compensatory effect *in vivo* than *in vitro*. Perhaps this reflects differences in U1 snRNP binding properties *in vivo* and *in vitro*. A second difference is that the G → A substitution results in a block to splicing after the first step *in vitro* (Fig. 1*B*), whereas *in vivo* some accurately spliced mRNA is produced (Fig. 4).

DISCUSSION

The results presented here indicate that the effects of these β-globin 5′-splice-site thalassemic mutations result from the overall decrease in the affinity of the mutant sites for U1 snRNP. In particular, the effects of these mutations are not due solely to the replacement of required nucleotides at critical positions. This conclusion complements that of Aebi *et al.* (21) who suggested that the precise position of 5′ cleavage results from the overall ability of a 5′ splice site to base pair with U1 snRNA.

We further conclude that cryptic 5′ splice sites are activated if their ability to bind U1 snRNP becomes comparable to that of the authentic 5′ splice site. This can result from two different types of mutations. (*i*) Single-base substitutions within the authentic 5′ splice site may decrease the affinity for U1 snRNP to the same approximate level as that of the cryptic sites. (*ii*) A cryptic 5′ splice site, such as the −38 site, can be activated by increasing its ability to base pair with U1 snRNA. In fact, increased complementarity of potential cryptic 5′ splice sites may be the basis of some natural thalassemias (27, 28).

Fig. 5 presents a model for 5′-splice-site selection and cryptic 5′-splice-site activation that incorporates the results presented here within the framework of our current understanding of spliceosome assembly. During this process a partial complex that includes U2 snRNP is assembled at the branch-site/3′-splice-site region of the intron. This partial complex then interacts with U1 snRNP bound at the 5′ splice site (for review, see refs. 1–4). Under standard *in vitro* splicing conditions, U1 snRNP is in excess (18). Thus, a pre-mRNA containing multiple potential 5′ splice sites of

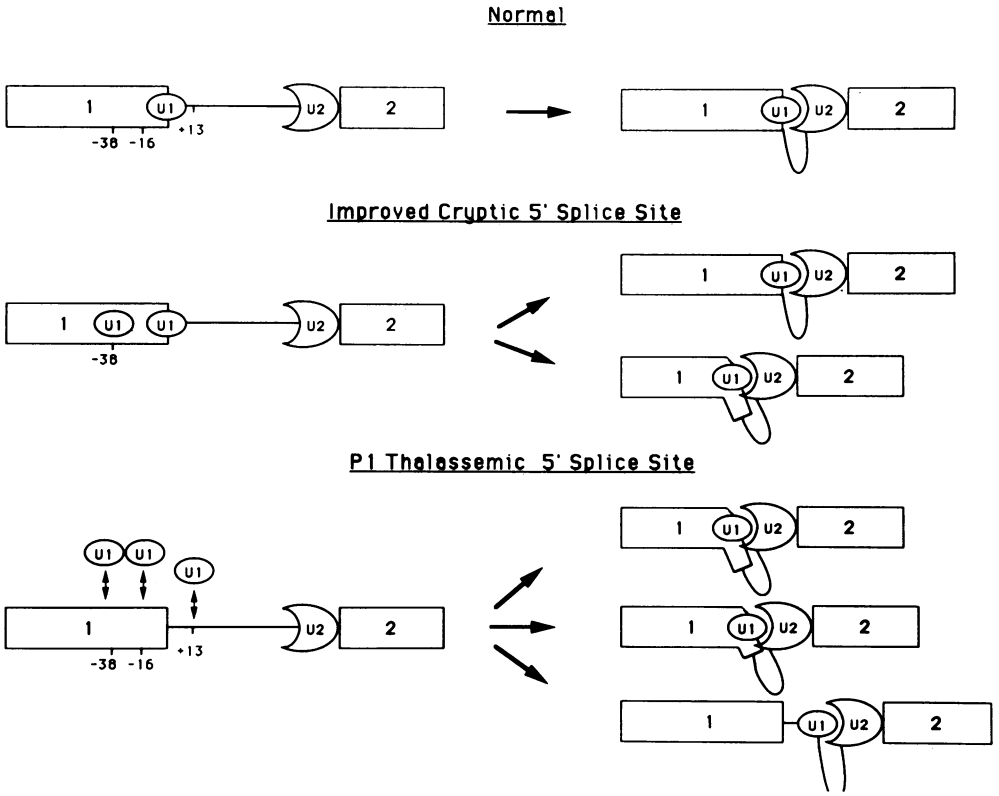

FIG. 5. Spliceosome competition model for 5'-splice-site selection. The stably bound U1 snRNP can form a productive complex with the factors bound to the branch site and polypyrimidine tract/3' splice site. Substrate RNAs containing two sites that bind U1 snRNP stably show splicing to both sites that are in competition for the bound factors at the other end of the intron. In cases where U1 snRNP doesn't bind stably to any site (as indicated by the double-headed arrow), all the weak binding sites are in competition.

approximately equal strength can have a U1 snRNP molecule simultaneously bound to each site (18). Therefore, when the affinity of a cryptic 5' splice site becomes comparable to that of the authentic site, the multiple bound U1 snRNPs can compete with one another for interaction with the single bound U2 snRNP and associated factors. In the wild-type pre-mRNA, the affinity of cryptic sites for U1 snRNP is significantly lower than that of the authentic site. Thus, the vast majority of pre-mRNA molecules contain a single U1 snRNP bound to the authentic site. Although no direct studies have shown that U1 snRNP is also in excess *in vivo*, the similarity of our *in vitro* and *in vivo* results is consistent with this idea.

Although U1 snRNP binding affinity is a critical aspect of 5'-splice-site selection, it is not the only criterion. Some high-affinity U1 snRNP binding sites can be bound by U1 snRNP and yet are inactive for splicing (18). Thus, the context of a U1 snRNP binding site plays an important role in splice site selection. The cryptic sites are all close to the authentic 5' splice site and presumably within a permissive context. This permissive context can be easily disrupted, since a small (9 nt) insertion between the −16 cryptic site and the authentic 5' splice site abolishes all splicing in a P1 mutant (24).

One explanation for context effects is that there are additional factors that bind at or near 5' splice sites and are involved in 5'-splice-site recognition and/or selection. The results presented here demonstrate that the identity of the nucleotide at highly invariant positions +1 and +5 has little effect on either the efficiency or the fidelity of splicing: in each instance the effects of these mutations can either be completely or partially suppressed by appropriate cis substitutions. These results raise the possibility that U1 snRNP is the only factor involved in 5'-splice-site recognition within a given context; it seems unlikely that these cis substitutions, designed to increase RNA·RNA base pairing with U1 snRNA, would also have compensatory effects for protein factors.

We acknowledge the contribution of Gloria Mao for initial work on the project. We thank members of the Green laboratory, Michael Rosbash, and Kathi Hall for critical comments on the manuscript. K.K.N. was supported by a National Science Foundation predoctoral training grant. This work was supported by a grant from the National Institutes of Health to M.R.G.

1. Green, M. R. (1986) *Annu. Rev. Genet.* **20**, 671–708.
2. Padgett, R. A., Grabowski, P. J., Konarska, M. M. & Sharp, P. A. (1987) *Annu. Rev. Biochem.* **55**, 1119–1150.
3. Maniatis, T. & Reed, R. (1987) *Nature (London)* **325**, 673–678.
4. Sharp, P. A. (1987) *Science* **235**, 766–771.
5. Zhuang, Y. & Weiner, A. M. (1986) *Cell* **46**, 827–835.
6. Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472.
7. Ohshima, Y. & Gotoh, Y. (1987) *J. Mol. Biol.* **195**, 247–259.
8. Shapiro, M. B. & Senapathy, P. (1987) *Nucleic Acids Res.* **15**, 7155–7174.
9. Black, D. L., Chabot, B. & Steitz, J. A. (1985) *Cell* **42**, 737–750.
10. Mount, S. M., Pettersson, I., Hinterberger, M., Karmas, A. & Steitz, J. A. (1983) *Cell* **33**, 509–518.
11. Bindereif, A. & Green, M. R. (1987) *EMBO J.* **6**, 2415–2424.
12. Ruby, S. W. & Abelson, J. (1988) *Science* **242**, 1028–1035.
13. Seraphin, B., Kretzner, L. & Rosbash, M. (1988) *EMBO J.* **7**, 2433–2538.
14. Konarska, M. M. & Sharp, P. A. (1986) *Cell* **46**, 845–855.
15. Kramer, A. (1987) *J. Mol. Biol.* **196**, 559–573.
16. Treisman, R., Orkin, S. H. & Maniatis, T. (1983) *Nature (London)* **302**, 591–596.
17. Krainer, A. R., Maniatis, T., Ruskin, B. & Green, M. R. (1984) *Cell* **36**, 993–1005.
18. Nelson, K. K. & Green, M. R. (1988) *Genes Dev.* **2**, 319–329.
19. Lillie, J. W., Green, M. & Green, M. R. (1986) *Cell* **46**, 1043–1051.
20. Favalaro, J. M., Treisman, R. H. & Kamen, R. (1980) *Methods Enzymol.* **65**, 718–749.

21. Aebi, M., Hornig, H. & Weissmann, C. (1987) *Cell* **50**, 237–246.
22. Siliciano, P. & Guthrie, C. (1988) *Genes Dev.* **2**, 319–329.
23. Chabot, B., Black, D. L., LeMaster, D. M. & Steitz, J. A. (1985) *Science* **230**, 1344–1349.
24. Chabot, B. & Steitz, J. A. (1987) *Mol. Cell. Biol.* **7**, 698–707.
25. Lawn, R.M., Efstratiadis, A., O'Connell, C. & Maniatis, T. (1980) *Cell* **21**, 647–651.
26. Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. (1984) *Cell* **38**, 317–331.
27. Goldsmith, G. E., Humphries, R. K., Ley, T., Cline, A., Kantor, J. & Nienhuis, A. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2318–2322.
28. Orkin, S. H., Kazazian, H. H., Antonarakis, S. E., Ostrer, H., Goff, S. C. & Sexton, J. P. (1982) *Nature (London)* **300**, 768–769.