Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 1301–1304, February 1992
Biochemistry

Two members of a conserved family of nuclear phosphoproteins are involved in pre-mRNA splicing

AKILA MAYEDA*, ALAN M. ZAHLER†, ADRIAN R. KRAINER*‡, AND MARK B. ROTH†‡

*Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724-2208; and †Fred Hutchinson Cancer Research Center, Seattle, WA 98104

Communicated by Thomas P. Maniatis, November 13, 1991

ABSTRACT **Monoclonal antibody 104 recognizes a subset of amphibian nuclear granules (B-snurposomes) and active sites of RNA polymerase II transcription in vertebrates and invertebrates. Monoclonal antibody 104 reacts with a set of nuclear serine- and arginine-rich phosphoproteins (SR family) with strikingly conserved apparent molecular masses. The most abundant family members in human (SRp33) and *Drosophila* (SRp55) cell lines can replace one another as essential splicing factors in a human cell-free system. Each of these polypeptides can functionally replace human SF2, an essential splicing factor that also regulates 5′ splice site selection of alternatively spliced pre-mRNAs *in vitro*. *Drosophila* SRp55 also functions as an alternative splicing factor in the human cell-free system. Analysis of cloned cDNAs shows that SRp55 and SF2 are highly related and reveals regions of similarity to genetically defined regulators of alternative splicing in *Drosophila*. These results suggest that the conserved SR family of phosphoproteins, which includes SRp55 and SF2, is involved in constitutive pre-mRNA splicing and in the specificity of alternative splice site selection.**

Alternative pre-mRNA splicing is a widespread mechanism for generating structurally and functionally distinct protein isoforms (1, 2). SF2 is a human protein essential for the first RNA cleavage-ligation step in general pre-mRNA splicing (3–5); the concentration of this factor also determines which 5′ splice site is selected in pre-mRNAs containing alternative sites (5–8). A human 33-kDa protein doublet has SF2 activity and can complement SF2-deficient extracts for the formation or stabilization of specific pre-spliceosome complexes (4). When pre-mRNAs containing alternative 5′ splice sites are spliced *in vitro*, SF2, also known as alternative splicing factor (ASF) (8), is able to influence which splice site is used; high and low concentrations of SF2 result in preferential use of proximal or distal 5′ splice sites, respectively (6, 8). Human cDNAs encoding SF2/ASF were recently isolated using DNA probes derived from the amino acid sequence of the purified protein (5, 7). The full-length cDNAs encode a polypeptide with a predicted molecular mass of 28 kDa, which contains an RNA recognition motif (RRM) (9, 10), and a region rich in arginine-serine dipeptides (RS domain). The anomalous electrophoretic mobility and heterogeneous appearance of SF2 are due to phosphorylation (refs. 5 and 7; A.M. and A.R.K., unpublished data).

Characterization of proteins that interact with active sites of RNA polymerase II transcription in amphibian germinal vesicles led to the identification of the SR family of conserved serine- and arginine-rich nuclear phosphoproteins (11, 12). These studies began with the characterization of monoclonal antibody 104 (mAb 104), raised against *Xenopus laevis* oocyte nuclear proteins (11). mAb 104 cross-reacts widely throughout the animal kingdom and stains active sites of RNA polymerase II transcription on the lateral loops of amphibian lampbrush chromosomes (11) as well as puffs on *Drosophila* polytene chromosomes (12). mAb 104 also stains a subset of nuclear granules (B-snurposomes) in amphibian oocytes (11, 13) and the nucleoplasmic speckled network in vertebrate cells (12, 14). These nuclear substructures contain many splicing components and are thought to be sites of pre-mRNA processing (11–17).

mAb 104 recognizes a family of nuclear phosphoproteins with strikingly conserved apparent molecular masses of 20, 30, 40, 55, and 75 kDa[§] in the somatic tissues of various animal cells (12). These proteins have highly characteristic solubility properties in ammonium sulfate and magnesium chloride solutions (12). The most abundant SR family member in *Drosophila* Kc cells is the 55-kDa polypeptide, termed SRp55 (12). *Drosophila* cDNA clones encoding SRp55 and a closely related variant, known as B52, were recently isolated (12, 18). The SRp55 open reading frame encodes a 39-kDa protein; the difference between apparent and actual molecular mass is largely due to phosphorylation (12). *Drosophila* SRp55, like human SF2, contains an RRM and an RS domain. The presence of similar sequence elements in human SF2 and *Drosophila* SRp55, and their similar fractionation properties, prompted us to examine whether SF2 belongs to the SR family and whether SRp55 is also a splicing factor.

MATERIALS AND METHODS

Protein Purification. To purify *Drosophila* SRp55, whole cell extracts from 10^{10} Kc cells were fractionated by precipitation with ammonium sulfate essentially as described (12). The fraction that precipitated between 65% and 90% of saturation was dialyzed and incubated on ice for 1 hr in 20 mM magnesium chloride, and the precipitated proteins were recovered by centrifugation (12). The pellet was fractionated by preparative SDS/PAGE, and the prominent mAb 104 immunoreactive 55-kDa protein was excised, eluted, and renatured in the presence of guanidinium hydrochloride and carrier acetylated bovine serum albumin, as described (19). The renatured protein was concentrated again by magnesium precipitation (12) and resuspended in splicing extract buffer (3). Human SRp33 was purified from 5×10^8 HeLa cells by the same method. Chromatographic purification of HeLa SF2 by functional complementation was as described (4). Purified polypeptides were analyzed by SDS/PAGE and stained with Coomassie blue (20).

Pre-mRNA Substrates. Radiolabeled pre-mRNA substrates were prepared by transcription with SP6 RNA polymerase in the presence of cap analogue, as described (3). Wild-type and thalassemic β-globin pre-mRNAs were made from pSP64-HβΔ6 and pSP64-HβΔ6IVS1-1A DNA templates (21) linear-

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: mAb, monoclonal antibody; RRM, RNA recognition motif; ASF, alternative splicing factor; snRNP, small nuclear ribonucleoprotein particle.

‡To whom reprint requests should be addressed.

§Approximate molecular masses to the nearest 5 kDa, as given in ref. 12; a value of 33 kDa is used here instead of 30 kDa.

ized with *Bam*HI. SP6 RNA polymerase, restriction enzyme, and cap analogue were from New England Biolabs.

Extract Preparation and Assay Conditions. The preparation of nuclear and cytoplasmic S100 extracts, splicing reaction conditions, RNA extraction, and urea/PAGE analysis were according to published procedures (3–5, 21).

RESULTS

Comparison of SRp55 and SF2 Amino Acid Sequences. Comparison of the predicted protein sequences of human SF2 (28 kDa) and *Drosophila* SRp55 (39 kDa) reveals identities in 123 of 248 amino acids in SF2 (50%) and conservative substitutions in 18 additional amino acids (Fig. 1). Human SF2 and *Drosophila* SRp55 contain an N-terminal RRM and a C-terminal RS domain. Similar sequence motifs are found singly or in combination in the genetically defined *Drosophila* alternative splicing regulators transformer (23), transformer-2 (24, 25), Sex-lethal (26), and suppressor-of-white-apricot (27) and in the 70-kDa polypeptide of the essential splicing factor U1 small nuclear ribonucleoprotein particle (snRNP) (9, 28, 29). The sequence similarities of these proteins to SF2 and to SRp55 are limited to the degenerate RRM and/or RS domains and have been previously noted (5, 7, 12). In contrast, the similarity between SF2 and SRp55 is much more extensive, both within and beyond the RRM and RS domains.

Human SF2 and *Drosophila* SRp55 have very similar, though not identical, RRMs that include the two conserved submotifs known as RNP-1 and RNP-2 (10, 22); the spacing between RNP-2 and RNP-1 is 30 amino acids in SF2 and 25 amino acids in SRp55. The two unusual prolines in the RNP-1 element of SF2 correspond to asparagine and glycine in SRp55; most RRM proteins contain arginine or lysine followed by glycine at these positions (9, 10, 22). The octapeptide EFEDPRDA, which overlaps the RNP-1 element of SF2, is identical to an octapeptide of unknown function that precedes the RRM of human and *Xenopus* U1 snRNP 70-kDa polypeptide (9, 28). *Drosophila* SRp55 contains the highly related sequence EFEDYRDA at a position homologous to that in SF2 (Fig. 1). The *Drosophila* 70-kDa homologue contains the less homologous but still related sequence DFEDPKDT preceding its RRM (29).

Both human SF2 and *Drosophila* SRp55 contain RS domains that occur primarily as consecutive RS dipeptides. Most of the length difference between the two proteins is accounted for by a longer basic region that includes the RS domain of SRp55 (Fig. 1). Both proteins are extremely basic in their unphosphorylated state: their theoretical isoelectric points are 10.77 and 11.92, respectively. Like the U1 70-kDa polypeptide (9, 28, 29), they contain glycine-rich stretches that may function as flexible hinges separating the RRM and RS domains. The extensive homology between SF2 and SRp55 also includes additional peptides not found in any other proteins in the current databases—e.g., amino acids 111–125 and 133–151 in the human SF2 sequence (Fig. 1).

Purification and Functional Characterization of SR Proteins. In addition to the extensive sequence homology between human SF2 and *Drosophila* SRp55, HeLa SF2 purified by functional complementation is recognized by mAb 104 in immunoblots, suggesting that SF2 is the 33-kDa SR family member (data not shown). mAb 104 does not recognize recombinant SF2 produced in *Escherichia coli* (data not shown; ref. 5), consistent with the previous finding that this antibody is specific for a phosphoepitope shared among the SR family members (11, 12). A further indication that SF2 belongs to the SR family is the fact that extracts deficient in SF2 activity are prepared by ultracentrifugation in the presence of 4.5 mM magnesium chloride, conditions that were later shown to sediment the entire SR protein family (3, 12). Highly purified SF2 also displays this unusual insolubility in magnesium chloride (data not shown).

The functional conservation of several splicing activities, including SF2, between humans and *Drosophila* has been noted in biochemical complementation experiments with partially purified fractions (ref. 30; C. Siebel and D. Rio, personal communication). To determine if SRp55 is functionally related to SF2, we purified SRp55 from *Drosophila* Kc cells by ammonium sulfate fractionation, followed by magnesium-induced precipitation (12). The prominent SRp55 polypeptide was purified further by preparative SDS/PAGE, renatured, and concentrated (Fig. 2, lane 3). The human mAb 104 immunoreactive SRp33 polypeptide, which may be identical to SF2, was similarly purified from HeLa cells to determine if it has SF2 activity (lane 2). This procedure yields highly purified SR polypeptides, except for the presence of carrier acetylated bovine serum albumin. Gel-purified *Drosophila* SRp55 and human SRp33 and chromatographically purified human SF2 (lane 1) (4) were tested for the ability to complement an SF2-deficient, splicing-inactive S100 extract (3, 4) from HeLa cells (Fig. 3).

All three preparations efficiently complemented the S100 extract (which provides all other essential splicing factors but lacks SF2 activity) to splice a wild-type human β-globin pre-mRNA (Fig. 3, lanes 2–5). Control gel-purified fractions lacking SRp55 or SRp33 but containing carrier acetylated bovine serum albumin had no activity (data not shown). A

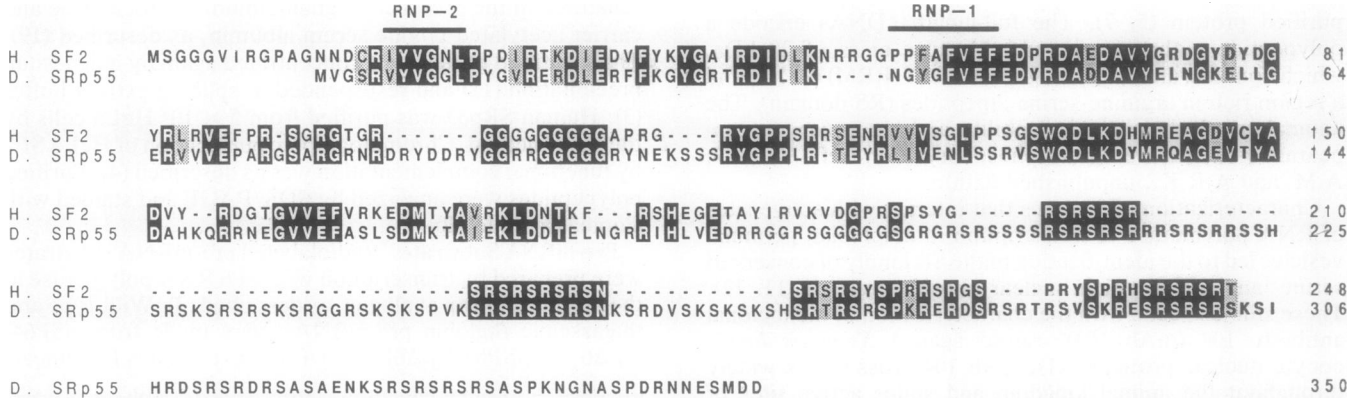

FIG. 1. Primary sequence homology between human SF2 and *Drosophila* SRp55. The complete amino acid sequences of human SF2 (H. SF2) (5, 7) and *Drosophila* SRp55 (D. SRp55) (12, 18) are shown in the one-letter code. Amino acid identities are indicated in reverse type; amino acid similarities are shown in stippled boxes; dashes denote gaps introduced to maximize homology. Amino acid positions in each protein are shown on the right. The conserved RNP-1 and RNP-2 elements of the RRM (9, 10, 22) are indicated by solid bars. The alignment shown preserves the RRM homology and minimizes the number of gaps (nine in H. SF2, two in D. SRp55).

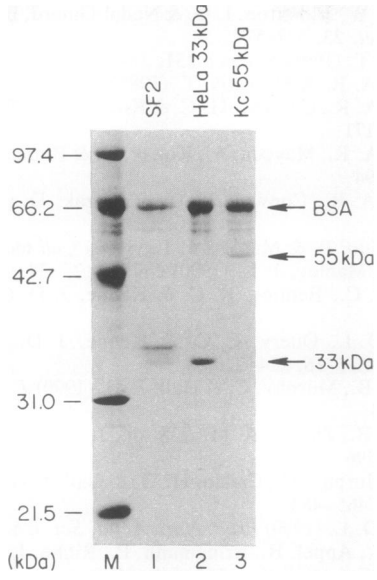

FIG. 2. Purification of human and *Drosophila* SR and SF2 polypeptides. Lane M, molecular mass markers (in kDa); lane 1, 28 μl of chromatographically purified human SF2 preparation; lane 2, 4 μl of gel-purified and renatured human SRp33; lane 3, 4 μl of gel-purified and renatured *Drosophila* SRp55. The molecular masses of the markers and the positions of the 33-kDa and 55-kDa mAb 104 antigens and of carrier bovine serum albumin (BSA) are indicated. Chromatographically purified HeLa SF2 appears heterogeneous because of differences in phosphorylation and because of the presence of a 32-kDa polypeptide unrelated in sequence (4, 5).

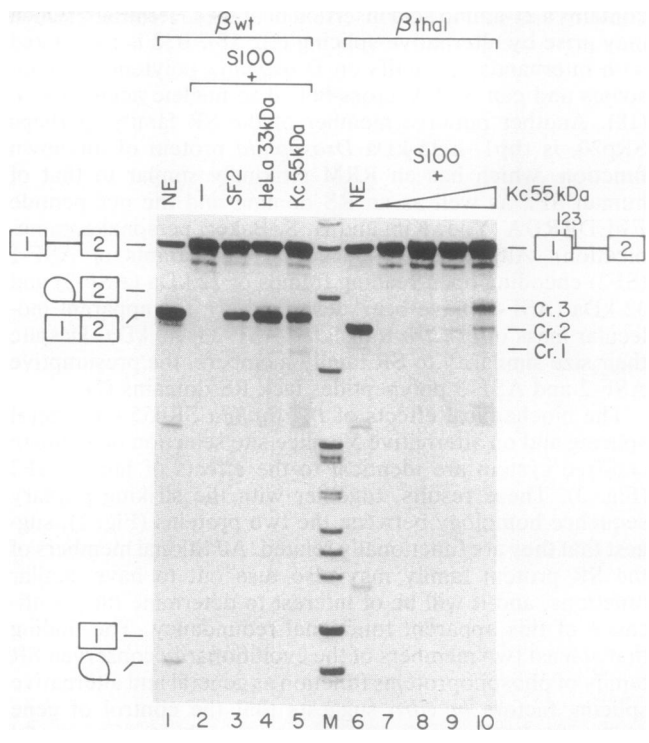

FIG. 3. *Drosophila* SRp55 can replace human SF2 in a human cell-free system. Wild-type (βwt; lanes 1–5) or β-thalassemic (βthal; lanes 6–10) pre-mRNAs containing the first two exons and first intron of human β-globin were spliced *in vitro* with the indicated extracts and fractions in 25-μl reaction mixtures containing 15 μl of extract and splicing extract buffer. Lane 1, 8 μl of HeLa nuclear extract (NE); lane 2, 7 μl of S100; lane 3, 7 μl of S100 plus 8 μl of chromatographically purified HeLa SF2; lane 4, 7 μl of S100 plus 8 μl of gel-purified human SRp33; lane 5, 7 μl of S100 plus 8 μl of gel-purified *Drosophila* SRp55; lane M, *Hpa* II-digested pBR322 molecular mass markers; lane 6, HeLa nuclear extract; lanes 7–10, 7 μl of S100 plus 0, 2, 4, or 8 μl of gel-purified *Drosophila* SRp55, respectively.

comparison of splicing efficiencies (Fig. 3, lanes 4 and 5) and protein concentrations (Fig. 2, lanes 2 and 3) shows that gel-purified *Drosophila* SRp55 has a somewhat higher specific activity than human SRp33. The control chromatographically purified human SF2 has considerably higher specific activity (Fig. 2, lane 1; Fig. 3, lane 3). The apparent stoichiometry of SF2 relative to pre-mRNA has been previously measured (5). The variable specific activities observed here most likely result from incomplete renaturation of the gel-purified proteins and/or from their different phosphorylation states.

To determine if *Drosophila* SRp55 can also influence alternative splicing in the heterologous *in vitro* system in a concentration-dependent manner, a β-thalassemic human pre-mRNA was employed (5, 31). In this mutant substrate, the 5′ splice site is inactive, and three cryptic 5′ splice sites (cr.1, cr.2, and cr.3) are available for alternative splicing to the authentic 3′ splice site. cr.1, cr.2, and cr.3 exist in linear order from 5′ to 3′ direction on the pre-mRNA. In the presence of constant amounts of S100 extract, the efficiency of splicing of the mutant pre-mRNA increased in proportion to the amount of purified SRp55 added (Fig. 3, lanes 7–10). A gradual switch from utilization of cr.2 to cr.3 was observed as the concentration of SRp55 was increased. This is seen as an increase in the ratio of cr.3 to cr.2 spliced mRNA (lane 10) and is comparable to the effect of human SF2 on the same pre-mRNA (5, 6). A similar stimulation of proximal alternative 5′ splice sites upon increasing the concentration of SRp55 was observed with β-globin pre-mRNA derivatives containing either a duplicated authentic first intron 5′ splice site or a proximal simian virus 40 small tumor antigen 5′ splice site (data not shown). We conclude that *Drosophila* SRp55 can functionally replace human SF2 *in vitro* and thus can interact with other human splicing factors to catalyze splicing and to influence alternative 5′ splice site selection.

DISCUSSION

We have shown that human SF2 and *Drosophila* SRp55 have extensive amino acid sequence homology (Fig. 1), whereas the similarity among either of these proteins, U1 70-kDa, known *Drosophila* splicing regulators, and the large family of RRM proteins, is more limited. However, we do not think that SF2 and SRp55 are true homologues, since human and *Drosophila* cell lines both have 33- and 55-kDa SR family members, albeit with different abundances (12). When human SF2 was first identified as a 33-kDa polypeptide doublet, secondary chromatographic peaks of SF2 activity were also reported, some of which contained a 55-kDa polypeptide (4). A polypeptide of the same size was also detected by immunoprecipitation of ^{35}S-labeled HeLa extracts with anti-SF2 antiserum (4). However, this polypeptide could not be characterized further because of its low abundance. More recently, partial amino acid sequence of human SRp55 isolated by the methods described here shows that it is more closely related to *Drosophila* SRp55 than to human SF2 (A.M.Z., W. Lane, and M.B.R., unpublished data). The slight difference in molecular mass between human SRp33 and SF2 (Fig. 2) is probably due to variable extents of phosphorylation in the two preparations and is consistent with the heterogeneity previously observed upon *in vitro* translation of a cloned SF2 cDNA and with the even greater electrophoretic mobility of unphosphorylated recombinant SF2 (5). However, it is possible that additional SR family members with molecular mass and activity similar to SF2 are also present in the purified SRp33 preparation.

A nearly identical variant of *Drosophila* SRp55, termed B52, has recently been described (18). The two proteins differ at 11 sites of apparent polymorphism, and, in addition, B52

contains a 21-amino acid insertion near the C terminus, which may arise by alternative splicing (12, 18). B52 is associated with interbands and puffs on *Drosophila* polytene chromosomes and can be UV cross-linked to nucleic acids *in vivo* (18). Another putative member of the SR family, perhaps SRp20, is rbp1, a 15-kDa *Drosophila* protein of unknown function, which has an RRM strikingly similar to that of human SF2 as well as an RS domain and the octapeptide EFEDRRDA (Y.-J. Kim and B. S. Baker, personal communication). Alternatively spliced cDNA variants of ASF-1 (SF2) encoding open reading frames of 22 kDa (ASF-3) and 32 kDa (ASF-2) have been described (7); the apparent molecular mass of *in vitro* translated ASF-2 is 40 kDa. Despite their size similarity to SR family members, the presumptive ASF-2 and ASF-3 polypeptides lack RS domains (7).

The biochemical effects of *Drosophila* SRp55 on general splicing and on alternative 5′ splice site selection in a human cell-free system are identical to the effects of human SF2 (Fig. 3). These results, together with the striking primary sequence homology between the two proteins (Fig. 1), suggest that they are functionally related. Additional members of the SR protein family may also turn out to have similar functions, and it will be of interest to determine the significance of this apparent functional redundancy. The finding that at least two members of the evolutionarily conserved SR family of phosphoproteins function as general and alternative splicing factors *in vitro* suggests that the control of gene expression by alternative splicing may involve tissue-specific and developmental regulation of the concentrations or activities of each of these phosphoproteins. Future experiments should address the importance of the phosphorylation state of these proteins, their relative abundances in different cells, and whether each member of the SR family has unique functional properties.

Previous studies have localized one or more SR family members to chromosomal sites of transcription (11, 12, 18) and to snurposomes and speckles (11, 13). Our finding that at least two of these proteins in *Drosophila* and humans are splicing factors is consistent with the views that splicing is a cotranscriptional phenomenon (13, 32), and that snurposomes and speckles are involved in splicing as spliceosome assembly or as storage sites (11–17).

We thank H. Weintraub for comments on the manuscript. This research was supported by grants from the National Institutes of Health (to M.B.R. and to A.R.K.). A.M. was supported by a long-term fellowship from the International Human Frontier Science Program Organization. A.M.Z. is a Burroughs-Wellcome Fund Fellow of the Life Sciences Research Foundation.

1. Smith, C. W. J., Patton, J. G. & Nadal-Ginard, B. (1989) *Annu. Rev. Genet.* **23**, 527–577.
2. Maniatis, T. (1991) *Science* **251**, 33–34.
3. Krainer, A. R. & Maniatis, T. (1985) *Cell* **42**, 725–736.
4. Krainer, A. R., Conway, G. C. & Kozak, D. (1990) *Genes Dev.* **4**, 1158–1171.
5. Krainer, A. R., Mayeda, A., Kozak, D. & Binns, G. (1991) *Cell* **66**, 383–394.
6. Krainer, A. R., Conway, G. C. & Kozak, D. (1990) *Cell* **62**, 35–42.
7. Ge, H., Zuo, P. & Manley, J. L. (1991) *Cell* **66**, 373–382.
8. Ge, H. & Manley, J. L. (1990) *Cell* **62**, 25–34.
9. Query, C. C., Bentley, R. C. & Keene, J. D. (1989) *Cell* **57**, 89–101.
10. Kenan, D. J., Query, C. C. & Keene, J. D. (1991) *Trends Biochem. Sci.* **16**, 214–220.
11. Roth, M. B., Murphy, C. & Gall, J. G. (1990) *J. Cell Biol.* **111**, 2217–2223.
12. Roth, M. B., Zahler, A. M. & Stolk, J. A. (1991) *J. Cell Biol.* **115**, 587–596.
13. Wu, Z., Murphy, C., Callan, H. G. & Gall, J. G. (1991) *J. Cell Biol.* **113**, 465–483.
14. Spector, D. L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 147–151.
15. Reuter, R., Appel, B., Bringmann, P., Rinke, J. & Lührmann, R. (1984) *Exp. Cell Res.* **154**, 548–560.
16. Nyman, U., Hallman, H., Hadlaczky, G., Petterson, I., Sharp, G. & Ringertz, N. R. (1986) *J. Cell Biol.* **102**, 137–144.
17. Fu, X.-D. & Maniatis, T. (1990) *Nature (London)* **343**, 437–441.
18. Champlin, D. T., Frasch, M., Saumweber, H. & Lis, J. T. (1991) *Genes Dev.* **5**, 1611–1621.
19. Hager, D. A. & Burgess, R. R. (1980) *Anal. Biochem.* **109**, 76–86.
20. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
21. Krainer, A. R., Maniatis, T., Ruskin, B. & Green, M. R. (1984) *Cell* **36**, 993–1005.
22. Bandziulis, R. J., Swanson, M. S. & Dreyfuss, G. (1989) *Genes Dev.* **3**, 431–437.
23. Boggs, R. T., Gregor, P., Idriss, S., Belote, J. M. & McKeown, M. (1987) *Cell* **50**, 739–747.
24. Goralski, T. J., Edström, J. E. & Baker, B. S. (1989) *Cell* **56**, 1011–1018.
25. Amrein, H., Gorman, M. & Nöthiger, R. (1988) *Cell* **55**, 1025–1035.
26. Bell, L. R., Maine, E. M., Schedl, P. & Cline, T. W. (1988) *Cell* **55**, 1037–1046.
27. Chou, T. B., Zachar, Z. & Bingham, P. M. (1987) *EMBO J.* **6**, 4095–4104.
28. Etzerodt, M., Vignali, R., Ciliberto, G., Scherly, D., Mattaj, I. W. & Philipson, L. (1988) *EMBO J.* **7**, 4311–4321.
29. Mancebo, R., Lo, P. C. H. & Mount, S. M. (1990) *Mol. Cell. Biol.* **10**, 2492–2502.
30. Zamore, P. D. & Green, M. R. (1991) *EMBO J.* **10**, 207–214.
31. Treisman, R., Orkin, S. H. & Maniatis, T. (1983) *Nature (London)* **302**, 591–596.
32. Beyer, A. L. & Osheim, Y. N. (1988) *Genes Dev.* **2**, 754–765.