

Molecular cloning of a mouse homologue for the *Drosophila* splicing regulator Tra2

Fernando Segade**, Belén Hurlé, Estefanía Claudio, Sofía Ramos, Pedro S. Lazo*

Departamento de Bioquímica y Biología Molecular, Universidad de Oviedo, 33006 Oviedo, Spain

Received 7 March 1996; revised version received 29 April 1996

Abstract We report the identification of a mouse cDNA, SIG41, encoding a protein of 288 amino acids that is 45% identical (58% similar) to the *Drosophila* splicing regulator Tra2. SIG41 cDNA contains four polyadenylation signals whose alternative use gives rise to four types of transcripts (2.1, 2.0, 1.5, and 1.4 kb) in mouse cells. Northern analysis and RT-PCR assays showed that SIG41 mRNA is present in virtually all the cell lines and tissues studied, with remarkable levels of expression in uterus and brain tissues. Differential stability of the SIG41 mRNAs was detected in mouse macrophage cells.

Key words: RNA processing; RNA-binding domain; Arg/Ser domains

1. Introduction

The processing of precursor mRNAs represents an important mechanism in the regulation of many eukaryotic genes. Interactions between particular regions of pre-mRNAs and the multicomponent splicing machinery, or spliceosome, are mediated by a number of *trans*-acting polypeptides that recognize specific RNA sequences or RNA molecules (see [1] for a review). Several of the known splicing factors have been identified as members of the family of the RBD class of RNA-binding proteins. All RBD-class proteins include one or more copies of a region of approximately 80 amino acids termed the RNA-binding domain (RBD), also named RNA-recognition motif (RRM), RNP-CS type RNA binding domain, RNP-80, or RNP motif [2–4]. The conserved residues in RBD are spread over the region with the most conserved sequences in the RNP-1 octamer and the RNP-2 hexamer submotifs [2,4]. While many of the RBD-class proteins perform essential functions in RNA processing, other RBD-class factors carry out specialized roles in many developmental and differentiation processes [2]. Thus, in *Drosophila melanogaster*, the products of two of the regulatory genes in the sex determination pathway, *Sex-lethal* (*Sxl*) and *transformer-2* (*tra2*), contain RBDs [5].

*Corresponding author. Fax: (34) (8) 510-3534.
E-mail: PSL@dwarf1.quimica.uniovi.es

**Present address: Department of Cell Biology and Physiology, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110, USA.

Abbreviations: EST, expressed sequence tag; RBD, RNA-binding domain; RT-PCR, reverse transcription-coupled polymerase chain reaction

The nucleotide sequence presented here has been deposited in the GenBank/EMBL database under accession no. X80232.

SIG41 was isolated from a subtracted mouse macrophage cDNA library constructed to identify genes overexpressed in cells exposed to mineral particles [6]. Two SIG41-hybridizing RNA species of 2.1 and 1.5 kb were detected in macrophages, and both were induced by silica particles more than 2-fold. Interestingly, the SIG41 gene was down-regulated in response to a battery of different stimuli [6]. The comparison of a partial sequence of the SIG41 clone with the GenBank/EMBL databases indicated that SIG41 corresponded to a novel expressed mRNA with significant homology to the members of the RBD-containing proteins.

In this paper we report the complete sequencing of mouse SIG41 cDNA, and identify the protein encoded in SIG41 as a mouse homologue of the *Drosophila* splicing regulator Tra2. We also determined the widespread expression of SIG41 gene in cells and tissues in four types of transcripts.

2. Materials and methods

2.1. Materials

Restriction endonucleases were from Boehringer Mannheim. [α - 32 P]dCTP (3000 Ci/mmol) and [α - 35 S]dATP (1000 Ci/mmol) were purchased from Amersham International. The GeneAmp PCR reagent kit was from Perkin-Elmer. All other chemicals were reagent or molecular biology grade and were obtained from Boehringer-Mannheim, Merck, or Sigma.

2.2. Cell culture

Cell lines were cultured in Dulbecco's modified Eagle's medium (Sigma) supplemented with 10% fetal bovine serum, 100 units/ml of penicillin, and 100 μ g/ml of streptomycin. For transcript stability experiments, RAW 264.7 cells were plated at 5×10^5 cells per well in 6-well plates (Corning) and actinomycin D (Sigma) was added at 5 μ g/ml.

2.3. Isolation of SIG41 and screening of a RAW 264.7 cDNA library

Isolation of SIG41 cDNA by differential screening of a subtracted RAW 264.7 cDNA library constructed in the pBluescript SK- vector (Stratagene) has been described previously [6]. Independent SIG41 cDNA clones were isolated after three rounds of screening of 5×10^5 clones from a RAW 264.7 cDNA library constructed in λ ZAPII [6] following standard procedures [7]. The cDNA insert of SIG41 was labeled by random priming using the 'RediPrime' (Amersham) labeling system in the presence of [α - 32 P]dCTP to a specific activity of 1×10^9 cpm/ μ g DNA. pBluescript SK- phagemids containing SIG41 cDNAs were excised in vivo from the λ ZAPII vector by coinfection of *E. coli* XL1-Blue cells with VCS M13 helper phage (Stratagene), according to the supplier's instructions.

2.4. DNA sequencing and analysis

Nucleotide sequences were determined by the dideoxy chain termination method [8] on single- and double-stranded template DNA using Sequenase 2.0 (USB Amersham) with custom-designed oligonucleotide primers. Sequences were manually read from both strands and in some cases sequence ambiguities were solved by substituting dITP for dGTP in the sequencing reactions. Sequence analysis and comparison were performed with the University of Wisconsin's Genetics Computer Group version 8.0 package of programs [9].

Fig. 1. Nucleotide sequence of SIG41 cDNA and deduced amino acid sequence. The ORF is shown in upper case; the 5' and 3' UTRs are shown in lower case. The deduced amino acid sequence of SIG41 in the one-letter code is indicated above the nucleotide sequence. The numbers at the right and left hand sides indicate nucleotide and amino acid residues, respectively. The RS domains are underlined, and the RBD is doubly underlined with the RNP-1 and -2 subdomains indicated by asterisks flanking the labels. The Gly stretch is in italics. Arrows and roman numerals indicate the 3' ends of the corresponding SIG41 cDNA types. Consensus polyadenylation signals are underlined.

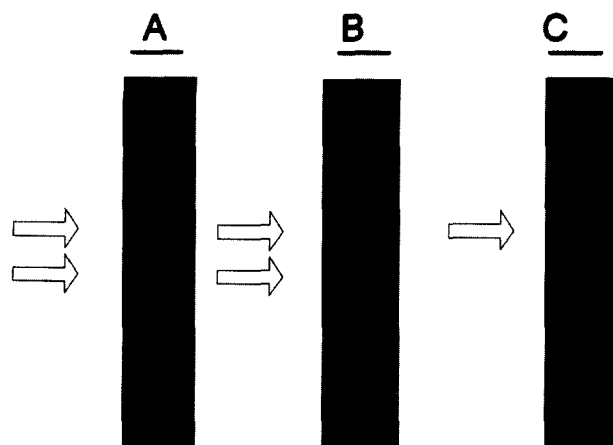


Fig. 2. Northern analysis of 15 µg of total cellular RNA from RAW 264.7 cells hybridized with the full-length SIG41 cDNA (A), an RBD-specific probe (B), and a type I- and II-specific 3' UTR probe (C). Arrows indicate the 2.1 and 1.5 kb transcripts.

riched in G+C residues (63%) including 16 CpG dinucleotides. The enrichment in CpG dinucleotides is usually associated with the 5' regions of genes coding for housekeeping proteins [12]. The SIG41 open reading frame (ORF) comprises 864 bp, beginning with the Met at position 146 (Fig. 1) that is located within a favorable environment for translation initiation [13], and is preceded by four in-frame stop codons. The conceptual translation of the ORF results in a 288 amino acid protein with a calculated mass of 33642 Da, and rich in basic amino acids (net charge +49 in the absence of phosphorylation, but see below). A hydropathy analysis of the SIG41 with the Kyte-Doolittle [14] algorithm predicts a highly hydrophilic peptide (data not shown). Examination of the SIG41-derived sequence revealed a 76 residue region that conforms to the RBD consensus sequence [2], extending from positions 118 to 193. An RNP-1 octamer (RGFAFVYF) and an RNP-2 hexamer (LGVFGL) are present within the RBD region. Two S+R-rich sequences or RS domains [2] are located at both ends of the predicted protein. The N-terminal RS1 domain spans 96 amino acids and is 68% R+S, including 21 RS pairs. The C-terminal RS2 domain is considerably shorter (28 amino acids) and similarly enriched in R+S residues (64%, including 6 RS pairs). Both RS domains contain a number of putative phosphorylation sites that may dramatically alter the net charge of the SIG41 protein. Finally there is a stretch of eight consecutive Gly located between the RBD and RS2 domains. This segment is similar to the glycine 'hinge' present in the SR family of splice factors [2] where it probably contributes to the flexibility of the peptide chain during binding [2].

A FASTA search of the non-redundant SwissProt/NBRF databases revealed the high similarity of SIG41 to the *Drosophila* Tra2 [15] splicing regulator (45% identity and 58% similarity). When RBDs in SIG41 and Tra2 are compared, identity reaches 57% (74% similarity), including five identical residues in the RNP2 hexamer, and six identical residues in the RNP2 octamer (Fig. 3A). The same amino acid is present in both proteins in 10 out of the 14 critical positions in the RBD consensus sequence of Birney et al. [2]. As expected, both RS1 and RS2 domains in SIG41 show high identity (30% and 61% respectively) with that of *Drosophila* Tra2 (Fig. 3B,C). Moreover, the overall structural features (N- and C-terminal RS domains, and the unique RBD) are lo-

cated in identical order in both proteins excluding the glycine hinge that is absent in Tra2 (Fig. 3D). The comparative analysis suggests that SIG41 represents a mouse homologue of *Drosophila* Tra2. Moreover, SIG41-related sequences are expressed in non-murine mammals since a search of the GenBank/EMBL nucleotide database found a number of ESTs derived from human (19 ESTs, average identity 83%) and rat (one EST, 94% identical) cDNA libraries (data not shown). Presumably more than one SIG41-related transcript might be present in human cells since at least three of the consensus polyadenylation signals in mouse SIG41 (at positions 1956, 1477, and 1388) are conserved in human EST sequences.

3.2. Expression of SIG41 mRNAs in cell lines and tissues

The range of expression of the SIG41 mRNAs was assessed by Northern blot analysis of total RNA extracted from a number of mouse (RAW 264.7 and J774A.1 macrophages, NIH 3T3 fibroblasts, and splenic B lymphocytes) and human (K562 erythroleukemic cells) cell lines, probed with the SIG41 insert (Fig. 4A). Two major hybridizing species of 2.1 and 1.5 kb were detected in all cells examined, regardless of their lineage. Densitometric scanning of the autoradiograms indicated that the relative proportion of the hybridizing species varied in the different cell lines from a 2.1 kb/1.5 kb relation > 9 in RAW 264.7 to approximately 1 in K562 and splenic B lym-

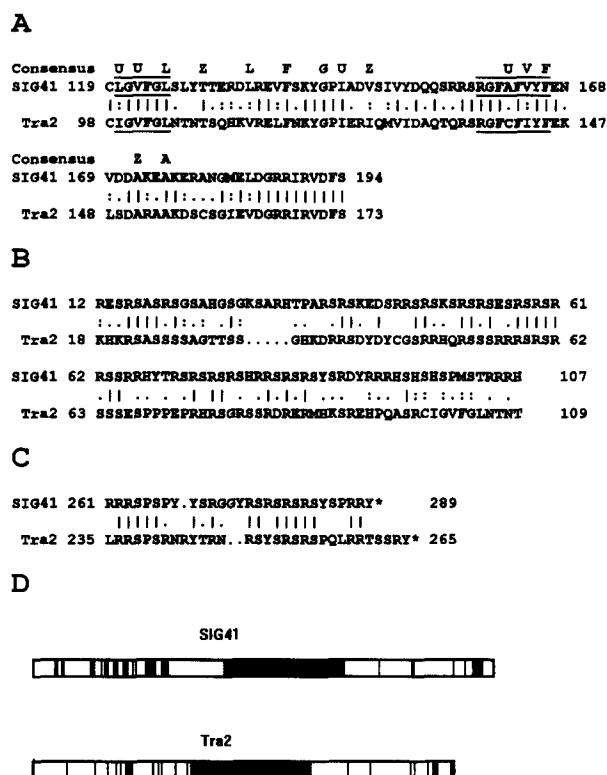


Fig. 3. Alignment of mouse SIG41 and *Drosophila* Tra2 and comparison of their structural motifs. Alignment of (A) RBD sequences, (B) RS2 sequences and (C) RS1 sequences. Conserved amino acids are shown by vertical bars, and conservative substitution by dots. The RNP submotifs are underlined. The RBD consensus sequence follows Birney et al. [2], where U = uncharged residues, and Z = U+S,T. (D) Comparison of structural motifs. Black shading represents the RBD. Vertical lines represent individual RS or SR dipeptides. The thickness is proportional to the number of consecutive repeats.

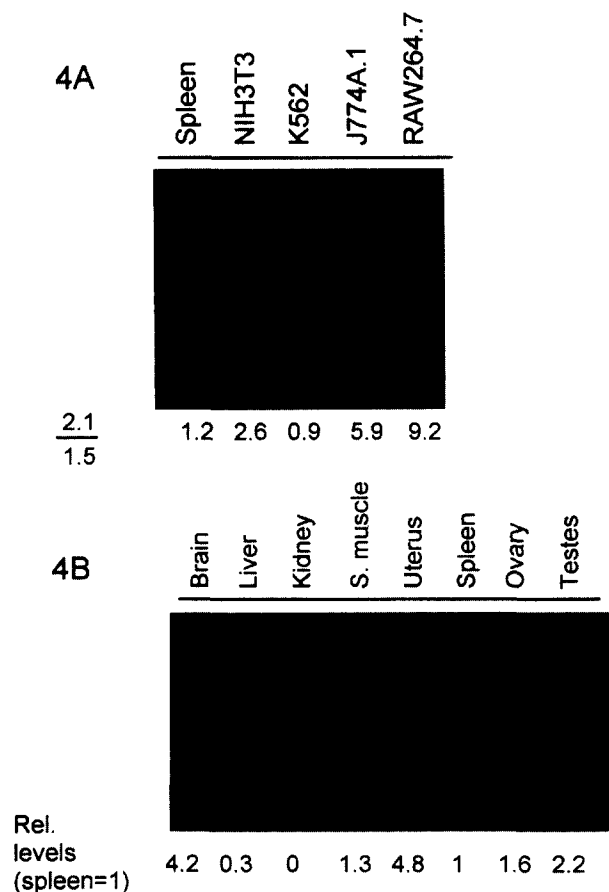


Fig. 4. Expression of SIG41 mRNAs in cell lines and mouse tissues. (A) Northern analysis of 15 µg of cellular RNA from the indicated cell lines, hybridized with the SIG41 cDNA insert. Numbers below the panel indicate the ratio between the two major hybridizing bands (2.1 and 1.5 kb) obtained by densitometric scanning of the autoradiogram. (B) RT-PCR of the RBD region of SIG41 from total RNA prepared from the indicated tissues. The amplification products after 15 cycles were hybridized with the SIG41 insert and autoradiographed. Transcript abundance is expressed as relative to levels in spleen and is the sum of the four cellular SIG41 mRNAs.

phocytes. The 2.1 kb/1.5 kb ratio was mainly influenced by the abundance of the 1.5 kb species.

A semi-quantitative reverse transcription-coupled PCR assay was performed to survey the relative abundance of SIG41 mRNAs in normal mouse tissues. The RT-PCR was carried out using two RBD-specific oligonucleotides as primers. To ensure that relative differences detected in the amplified products mirrored the levels of the SIG41 mRNAs in tissues, we limited the number of amplification cycles to 15 [16]. The 260 bp amplification products were detected by Southern hybridization with a ³²P-labeled SIG41 probe since no bands could be discerned in ethidium bromide-stained agarose gels. The autoradiogram in Fig. 4B shows that SIG41 mRNAs are present in most tissues assayed. When present, however, relative levels differed sharply. Thus, in uterus smooth muscle, brain, and testes, transcript levels were 5-, 4-, and >2-fold higher than in splenic B lymphocytes (in which mRNA levels were assessed by Northern analysis and used as an arbitrary basal level, Fig. 4A) whereas in liver transcript levels were only 30% of that of splenic B lymphocytes. PCR products from kidney RNA were not detected.

3.3. Differential stability of the 2.1 and 1.5 kb SIG41 species

Relative proportions of SIG41 mRNAs could be due to their differential expression or, alternatively, be the result of unequal stabilities, as described for other RNA-binding proteins [17]. To measure mRNA stability, RAW 264.7 cells were incubated up to 10 h in the presence of 5 µg/ml of actinomycin D. The time-point abundance of the transcripts and the rate of degradation was evaluated by densitometric scanning of the appropriate Northern blots probed with ³²P-labeled SIG41 cDNA. As shown in Fig. 5, SIG41 mRNAs were similarly stable during the first 4 h of incubation. Although actinomycin D globally affected the stability of all RNAs (the 28S and the 18S ribosomal RNAs started to be degraded after 4 h, not shown), the 2.1 kb and the 1.5 kb SIG41 mRNA species decayed with different rates after 4 h. Thus, the 2.1 kb/1.5 kb ratio diminished from an initial value of about 6 at 0–4 h to a value of 1.0 at 10 h. Examination of the SIG41 3' UTR revealed that only one ATTTA mRNA destabilizing sequence [18] is present, which could contribute to the differential stability of the SIG41 mRNAs. However, it is unlikely that this motif alone could be responsible for the lower stability of the 2.1 kb transcripts since it is located at position 2045, and therefore present only in type I SIG41 mRNA.

4. Discussion

We have cloned a novel mouse gene from a macrophage cell line. The encoded product of the SIG41 gene is very similar to the product of the *tra2* gene of *Drosophila*. The examination of the amino acid sequence suggests that the SIG41 protein is a member of the RBD-class gene family, which consists of proteins with a loosely conserved RNA-binding domain [2]. SIG41 also contains two RS domains characteristic of many RBD-class proteins. Similar domains in a number of RNA-binding proteins are essential for splicing activity [19], protein-protein interaction [20], or subnuclear localization [21,22].

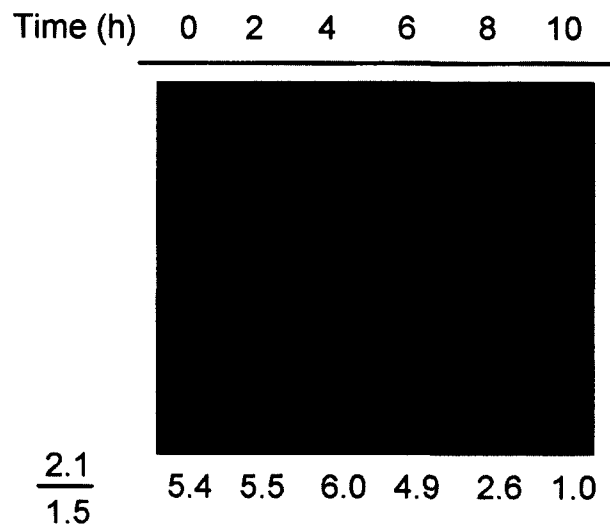


Fig. 5. Stability of SIG41 mRNAs. RAW 264.7 cells were incubated for 2–10 h in the presence of actinomycin D (5 µg/ml). Cellular RNA (15 µg/lane) was subjected to Northern analysis with the SIG41 probe. Densitometric analyses of appropriately exposed autoradiograms were carried out to calculate the ratio between the 2.1 and 1.5 kb mRNA species. The ratio indicated by numbers below the panel corresponds to the average of two independent experiments.

In *Drosophila*, several of the genes controlling sex differentiation perform their function by regulating alternative splicing of certain transcripts in a sex-specific manner [23]. *tra2* is a gene whose encoded products are required for the splicing of the *doublesex* (*dsx*) precursor mRNA into its female form by activation of the *dsx* female splice-site choice [24–28] in female somatic cells, presumably by direct binding of Tra2 to the *dsx* pre-mRNA [29]. In male germ cells, Tra2 auto-regulates processing of its primary transcript [30] and it is required for normal spermatogenesis [31] and processing of some primary transcripts into male-specific forms, as has been demonstrated in *exuperantia* [32]. A general role in splicing has been advanced for Tra2, since the gene product is functional in somatic male cells of *Drosophila* [33] even though its lack of function has no apparent effect on sex differentiation [34]. Although SIG41 seems to be a structural homologue of Tra2 in mammals, a sex-dependent function is not likely given the profound differences in the mechanisms of sexual differentiation between vertebrates and insects [23]. However, the general function of Tra2 in somatic cells might be conserved. Significantly, Tra2 interacts in vitro with the general splicing factors SF1/ASF, SC35, and U2AF [20,35] and, in some transfection studies, Tra2 colocalizes with SC35 in specific regions of the mammalian cell nucleus [22]. Moreover, the widespread expression of SIG41 mRNA in cell lines and tissues points to a general role of the SIG41 protein.

In mouse cells SIG41 is expressed as four transcripts of approximately 2.1, 2.0, 1.5, and 1.4 kb, which conceivably arise from the 3' end processing downstream of each of the four polyadenylation signals in a unique primary transcript. Therefore the same 288 residue protein is encoded in every SIG41 mRNA, although it is conceivable that the differential stability exhibited by the two major SIG41 mRNA species could function as a regulatory mechanism for the expression of SIG41 in specific tissues and thus it might account for the observed differences in the relative abundances of the 2.1 and 1.5 kb species in the cell lines studied. In *tra2*, four alternatively spliced mRNAs are transcribed in a sex- and tissue-specific manner, thus giving rise to four different Tra2 protein variants [15]. Although no alternatively spliced cDNAs were isolated from macrophages, a more complex pattern of mRNA expression in specific tissues and/or developmental stages cannot be completely ruled out. Studies are now in progress to elucidate the role of SIG41 in gene expression during the phagocytosis process in macrophages.

Acknowledgements: We are grateful to Dr. Y.S. López-Boado for her excellent advice and critical reading of the manuscript, and to Dr. C. López-Otin for the synthesis of primers. This work was supported by Comisión Interministerial de Ciencia y Tecnología Grant SAF94-0389. B.H. was supported by a Ministerio de Educación y Ciencia predoctoral fellowship, and E.C. was supported by a Fundación para la Investigación Científica y Técnica predoctoral fellowship.

References

- [1] Dreyfuss, G., Matunis, M.J., Piñol-Roma, S. and Burd, C.G. (1993) *Annu. Rev. Biochem.* 62, 289–321.
- [2] Birney, E., Kumar, S. and Krainer, A.R. (1993) *Nucleic Acids Res.* 21, 5803–5813.
- [3] Mattaj, J. (1993) *Cell* 73, 837–840.
- [4] Nagai, K., Oubridge, C., Ito, N., Avis, J. and Evans, P. (1995) *Trends Biochem. Sci.* 20, 235–240.
- [5] Mattox, W., Ryner, L. and Baker, B.S. (1992) *J. Biol. Chem.* 267, 19023–19026.
- [6] Segade, F., Claudio, E., Wrobel, K., Ramos, S. and Lazo, P.S. (1995) *J. Immunol.* 154, 2384–2392.
- [7] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- [8] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- [9] Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acids Res.* 12, 387–395.
- [10] Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* 162, 156–159.
- [11] Wahle, E. and Keller, W. (1992) *Annu. Rev. Biochem.* 61, 419–440.
- [12] Gardiner-Garden, M. and Frommer, J. (1987) *J. Mol. Biol.* 196, 261–282.
- [13] Kozak, M. (1991) *J. Cell Biol.* 115, 887–903.
- [14] Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 105–132.
- [15] Amrein, H., Maniatis, T. and Nöthiger, R. (1990) *EMBO J.* 9, 3619–3629.
- [16] Botney, M.D., Parks, W.C., Crouch, E.C., Stenmark, K. and Mecham, R.P. (1992) *J. Clin. Invest.* 89, 1629–1635.
- [17] Sureau, A. and Perbal, B. (1994) *Proc. Natl. Acad. Sci. USA* 91, 932–926.
- [18] Shaw, G. and Kamen, R. (1986) *Cell* 46, 659–667.
- [19] Zamore, P.D., Patton, J.G. and Green, M.R. (1992) *Nature* 355, 609–614.
- [20] Amrein, H., Hedley, M.L. and Maniatis, T. (1994) *Cell* 76, 735–746.
- [21] Li, H. and Bingham, P.M. (1991) *Cell* 67, 335–342.
- [22] Hedley, M.L., Amrein, H. and Maniatis, T. (1995) *Proc. Natl. Acad. Sci. USA* 92, 11524–11528.
- [23] Ryner, L.C. and Swain, A. (1995) *Cell* 81, 483–493.
- [24] Burtis, K.C. and Baker, B.S. (1989) *Cell* 56, 997–1010.
- [25] Nagoshi, R.N., McKeown, H., Burtis, K.C., Belote, J.M. and Baker, B.S. (1988) *Cell* 53, 229–236.
- [26] Hoshijima, K., Inoue, I., Higuchi, H., Sakamoto, H. and Shimura, Y. (1991) *Science* 252, 833–836.
- [27] Nagoshi, R.N. and Baker, B.S. (1990) *Genes Dev.* 4, 89–97.
- [28] Ryner, L.C. and Baker, B.S. (1991) *Genes Dev.* 5, 2071–2085.
- [29] Hedley, M.L. and Maniatis, T. (1991) *Cell* 65, 579–586.
- [30] Mattox, W. and Baker, B.S. (1991) *Genes Dev.* 5, 786–796.
- [31] Belote, J.M. and Baker, B.S. (1983) *Dev. Biol.* 95, 512–517.
- [32] Hazelrigg, T. and Tu, C. (1994) *Proc. Natl. Acad. Sci. USA* 91, 10752–10756.
- [33] McKeown, M., Belote, J.M. and Boggs, R.T. (1988) *Cell* 53, 887–895.
- [34] Schüpbach, T. (1982) *Dev. Biol.* 89, 117–127.
- [35] Wu, J.Y. and Maniatis, T. (1993) *Cell* 75, 1061–1070.