Nucleic Acids Research, 1994, Vol. 22, No. 22 4697–4704

Transcription of satellite 2 DNA from the newt is driven by a snRNA type of promoter

Stephen R.Coats[1], Yi Zhang[1] and Lloyd M.Epstein[1,2,*]
[1]Institute of Molecular Biophysics and [2]Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA

Received July 5, 1994; Revised and Accepted September 29, 1994

ABSTRACT

The transcriptional promoter of satellite 2 from the eastern newt, *Notophthalmus viridescens*, was analyzed by assaying the activity of deleted or mutated satellite 2 clones in *Xenopus laevis* oocytes. Two elements in the promoter were found to be important for transcription. These elements have sequences that are similar to the sequences of the octamer and the proximal sequence element of vertebrate snRNA genes transcribed by RNA polymerase II. Furthermore, the organization of these elements and their respective roles in transcription are the same as their organization and roles in the snRNA genes. To further investigate the relationship between the satellite 2 and snRNA gene promoters, the ability of the satellite 2 promoter to drive transcription of a true snRNA gene was tested. The satellite 2 promoter initiated transcription of the *Xenopus* U1b2 snRNA gene as efficiently as the native U1b2 promoter, and the 3′ ends of the resulting U1b2 transcripts were accurately formed. This latter result confirms that the satellite 2 promoter is a functional analog of the snRNA promoter, since 3′-end formation of snRNA genes transcribed by RNA polymerase II requires that transcription be initiated from a compatible promoter. The structural and functional similarities between the satellite 2 and the snRNA gene promoters suggest that these elements are evolutionarily related. These findings were used to extend a previously proposed model concerning the nature and derivation of satellite 2.

INTRODUCTION

Satellite 2 is a 300–350 bp repetitive DNA that is highly conserved in families of salamanders that have diverged between 65 and 200 million years ago (1). It was first found in the eastern newt, *Notophthalmus viridescens*, where tandem repeats of satellite 2 are dispersed throughout the genome (2). Much of the interest in this sequence has focused on the ability of satellite 2 transcripts to catalyze their own site-specific self-cleavage using an 'extended hammerhead' domain (3, 4). Further interest stems from the recent report that cloned satellite 2 DNA is transcriptionally active when assayed in oocytes from *Xenopus laevis* (5). Transcription and/or self-cleavage are apparently regulated *in vivo* so that repeat-length monomeric transcripts with different permutations of the same primary sequence are produced in a tissue-specific manner (2, 6). The monomeric transcripts found in tissues such as the liver and testes have ends that correspond to the site of self-cleavage, while the monomers in the ovary have 5′ and 3′ ends located 42 nt upstream from intact self-cleavage sites.

It is not clear if satellite 2 has a cellular function, but the heterogeneity of the genomic repeats indicates that most of them are defective in whatever activity is ultimately associated with this sequence. Similarities in the genomic organization and sequence divergence patterns between the satellite 2 and primate ALU elements led us to propose that dispersed arrays of satellite 2 were generated by retroposition of an active subset of satellite 2 (1–3), just as the dispersed ALU repeats are believed to be derived from the retroposition of a single or small subset of active ALU elements (7, 8). The active set of satellite 2 might be composed of true cellular genes whose products happen to be well-suited for retroposition, or they might be pseudogene-like copies of a cellular gene as is the case for the master ALU elements which are rearranged 7SL RNA genes (9).

To clarify the nature and derivation of satellite 2, we are characterizing its genetic properties and relating them to the properties of known cellular genes. The present paper examines the relationship between the transcriptional promoter and the promoters of vertebrate snRNA genes. Cremisi *et al.* found that transcription of satellite 2 from the European newt, *Triturus vulgaris meridionalis*, showed sensitivity to α-amanitin characteristic of transcription by RNA polymerase II (5). The authors also noted similarities within the satellite 2 sequence to regions in the promoters of vertebrate small nuclear RNA (snRNA) genes transcribed by RNA polymerase II. This well characterized class of promoters contains two functional domains referred to as the proximal sequence element (PSE) and the distal sequence element (DSE) (10). The PSE is centered approximately 60 bp upstream from the RNA cap site and is required for transcription and determining the site of initiation. The DSE, located 200–250 bp upstream from the cap site, is not absolutely required, but acts as an enhancer to stimulate transcription. The

*To whom correspondence should be addressed at: Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA

DSEs of all major snRNA gene promoters that have been analyzed contain a conserved octamer motif as an essential component. One or more additional factor binding sites are often located in close proximity to the octamer motif, but the nature of these sites, and the factors that bind to these sites, vary between individual snRNA genes (11–14). The sequences found in satellite 2 by Cremisi *et al.* were similar to the octamer and PSE motifs. A detailed analysis of the satellite 2 promoter was not performed, however, and the roles of these putative promoter elements were not determined.

Here we show that the octamer and the PSE are involved in the transcription of satellite 2, their roles are similar to the roles of the corresponding elements in the snRNA promoters, and the entire satellite 2 promoter is functionally interchangeable with the *Xenopus* U1b2 snRNA gene promoter. These findings suggest that an evolutionary relationship exists between satellite 2 and snRNA genes.

MATERIALS AND METHODS

Satellite 2 clones

The satellite 2 clones pGD, pGD1,2M, and pG−336/+4 were constructed previously (15). pG−336/+4 corresponds to pG47:473 by the previous nomenclature. The clones diagrammed in Figure 3A were constructed by subcloning deletions of the insert in pG−336/+4 into a *Sma*I/*Eco*RI digested pGEM3Zf(−) vector (Promega Corporation). The *Eco*RI ends of the deleted inserts were derived from the single *Eco*RI site in the polylinker downstream of the transcription initiation site in PG−336/+4. *Sma*I compatible blunt ends were generated in three different manners. For pG−110/+4, a blunt end at position −110 (relative to the initiation site) was formed by digestion with *Rsa*I. For pG−285/+4 and pG−195/+4, blunt ends were generated by digestion with *Sty*I and *Nde*I, respectively, followed by removal of the 5′ overhangs with S1 nuclease. For pG−82/+4, pG−63/+4 and pG−46/+4, the blunt end restriction sites *Stu*I, *Ssp*I and *Sma*I were engineered at the indicated locations by site-specific mutagenesis (16).

pG−336/+4 Oct Mut and pG−336/+4 PSE Mut were created by site-specific mutagenesis, and differ from pG−336/+4 solely at the positions indicated in Figure 3A. The sequences of these and all other modified clones were verified by dideoxynucleotide sequencing (17).

5S, U1b2, and chimeric clones

p115/77 is a *Xenopus borealis* somatic 5S RNA maxigene (18). A 364 bp *Hin*dIII fragment from p115/77, containing the entire coding region, was subcloned into the *Hin*dIII site of pGEM3Zf(−) to generate pG115/77. pXlU1b contains a tandem array of the *X. laevis* embryonic U1b1 and U1b2 genes (19). A 693 nt *Pst*I fragment containing the U1b2 gene from pXlU1b was subcloned into *Pst*I digested pGEM3Zf(−) to generate pGU1b2. pGU1b2 maxi was constructed by inserting a synthetic double-stranded linker into the *Bcl*I site at position +27 of the coding region in pGU1b2. The sequence between the duplicated *Bcl*I sites in the resulting clone is 5′-TCGCGAGAATT-CGCCGGCCCTAGG-3′.

The chimeric clone pGsat−26/U1−25 was constructed by the directed ligation of three separately prepared fragments of DNA. The first was derived from pG−336/+4, and contained position −336 to position −26 of the satellite 2 promoter. Position −336 was generated by digestion with *Hin*dIII; position −26 was generated by digestion with *Eco*O109I and filling in the 5′ overhang with T4 DNA polymerase. The second fragment was a 391 bp *Pvu*II/*Pst*I fragment from pGU1b2 maxi consisting of position −25 to +366 of the U1b2 maxigene. These two fragments were ligated simultaneously with *Hin*dIII/*Pst*I digested pGEM3Zf(−) to generate pGsat−26/U1−25. For the construction of pGsat+43/U1+39, a 361 bp fragment was excised from pGU1b2 maxi by digestion at the *Eco*RI sites in the maxigene insert and in the downstream polylinker region. This fragment was inserted into the *Eco*RI site at position +43 in pG−336/+43 2M. This latter clone was previously constructed and called pG47:331 2M (15). 2M designates the deletion of an A residue in the conserved GAAAC motif of the hammerhead domain which completely abolishes self-cleavage activity *in vitro*.

Oocyte microinjection and RNA preparation

Mature *X. laevis* were obtained from *Xenopus* I. Techniques for the preparation and injection of *Xenopus* oocytes were modified from Colman (20), and described previously (1). Oocytes were injected with approximately 50 nl of water containing cloned DNA at the indicated concentrations. RNA was prepared from batches of 10–15 microinjected oocytes by homogenization in SDS and precipitation with lithium chloride as previously described (1).

Primer extensions

Primer extension reactions using 5′-end-labeled oligonucleotide primers were described previously (2). The primers used to analyze satellite 2 transcription were the 24mer described previously (6), or the universal M13 forward sequencing primer. The primer used to quantify transcription from the 5S RNA maxigene was complementary to the maxigene insert and had the sequence 5′- CCCATCCAAGTCCGGATC-3′. Two different primers were used to analyze transcription of the U1b2 maxigene. One had the sequence 5′-GGGGAAATCGCAGGGGTCAG-3′ and was complementary to a region downstream of the maxigene insert and thus annealed to endogenous U1 RNA as well as to the maxigene transcripts. The other primer, 5′-GATCACCTA-GGGCCGGCGAATTCTCGCGAT-3′, was complementary to the maxigene insert and did not interact with endogenous U1 RNA.

For determining the relative expression of the satellite 2 promoter deletions, primer extension reactions were electrophoresed on denaturing polyacrylamide gels, and extension products were quantified using a Betascope 603 Blot Analyzer (Betagen Corporation). Satellite 2 extension products were standardized to the extension products of transcripts from the coinjected 5S maxigene to control for experimental fluctuations. These standardized values were then used to determine the relative expression of the deletion clones.

Deoxyoligonucleotide-directed RNase H cleavage

In order to map the 3′ ends of transcripts from the U1b2 maxigene and the chimeric constructs, and to generate smaller fragments of RNA that were easier to resolve on acrylamide gels, we cleaved the transcripts at fixed, internal positions using deoxyoligonucleotide-directed RNase H cleavage. The deoxyoligonucleotide used for this purpose was the primer used for primer extensions that was complementary to the insert in the U1b2 maxigene. To anneal the primer to the template RNAs, the specified amount of RNA from injected oocytes was combined

with 0.3 μg of unlabeled primer in 15 μl mixtures consisting of 250 mM KCl, 10 mM Tris (pH 8.0), and 1 mM EDTA. The mixtures were then incubated at 65°C for 1.5 h, cooled to 32°C over a 1.5 h period, and incubated an additional 8–12 h at 32°C. Cleavage with RNase H was performed by adding 30 μl of cleavage buffer (7.5 mM $MgCl_2$, 1.5 mM DTT, 15 U RNase H) to the annealing reactions and incubating the mixtures at 37°C for 1 h. Reactions were terminated by adding 2 μl 0.5 M EDTA, and the RNA was precipitated with ethanol.

The RNase H cleavage products were analyzed by Northern blots as described previously (2), except that the RNA was separated on denaturing polyacrylamide gels and electrophoretically transferred to GeneScreen Plus Transfer Membranes as recommended by the supplier (NEN). Radioactively labeled probes were prepared by *in vitro* transcription with either SP6 or T7 RNA polymerase (3).

RNase protection analyses

Radioactively labeled antisense RNA probes were prepared from linearized pGD DNA by transcription with T7 RNA polymerase (3). Probe and target RNAs were combined, dried, resuspended in 30 μl of hybridization buffer [40 mM PIPES (pH 6.4), 1 mM EDTA, 0.4 M NaCl, 80% formamide], and overlaid with 25 μl of mineral oil. To anneal the probe to the target, samples were heated at 90°C for 10 min and then incubated at 50°C for 12 h. 300 μl of RNase digestion mixture [300 mM NaCl, 10 mM Tris (pH 7.4), 5 mM EDTA, 90 U/ml RNase T1, 40 μg/ml RNase A] was then added to each sample and the reactions were incubated at 30°C for 1 h. The reactions were terminated by adding 20 μl of 10% SDS and 10 μl of 10 mg/ml proteinase K and incubating the mixture at 30°C for 30 min. Samples were phenol extracted, precipitated with ethanol, and analyzed on 7.5% polyacrylamide, 7 M urea gels.

RESULTS

Satellite 2 from *N.viridescens* is transcriptionally active in *Xenopus* oocytes

To investigate the transcriptional activity of satellite 2, we used a cloned satellite 2 dimer from the newt, *N.viridescens*, for *Xenopus* oocyte injections. This clone, named pGD (Fig. 1A), has the same organization as the *T.vulgaris meridionalis* clone that was shown to be transcriptionally active by Cremisi *et al.* (5). The products of satellite 2 transcription in *Xenopus* oocytes were first analyzed by primer extensions (Fig. 1B). The 24mer primer used in this analysis gave two extension products with RNA from injected oocytes (lane 4) but not with RNA from uninjected oocytes (lane 3). The less abundant 110 nt product is derived from transcripts which have undergone self-cleavage in *Xenopus* oocytes after transcription. This was evident from the size of the extension product which is the same as that obtained with transcripts synthesized and self-cleaved *in vitro* (lane 1). Furthermore, this extension product was not found when *in vivo* transcription was directed by a mutant satellite 2 dimer, pGD1,2M, in which both self-cleaving domains were inactivated (lane 5).

The more abundant 150 nt extension product was produced equivalently using RNA from oocytes injected with either pGD or pGD1,2M (Fig. 1B, lanes 4 and 5), and is the same size as the major extension product obtained with newt ovary RNA (lane 2). The 5′ ends represented by these extension products are indicative of transcription driven by upstream snRNA-like promoter elements. By adding dideoxynucleotides to the primer extension reactions, we determined that the predominant 5′ nucleotide in both the newt ovary RNA and the products of transcription in *Xenopus* oocytes was an adenine residue located 53 nt downstream from the 3′ border of the putative PSE (data not shown). There are two of these sites in the pGD dimer, and the second is 196 nt downstream from the 3′ end of an octamer motif (see Fig. 1A). This spacing agrees well with the organization of the promoter elements and the initiation site in most vertebrate snRNA gene promoters (10). Thus, transcription of satellite 2 in *Xenopus* oocytes is consistent with the use of its snRNA-like promoter, and transcription from this promoter is likely to be responsible for 5′-end formation of satellite 2 transcripts in the newt ovary.

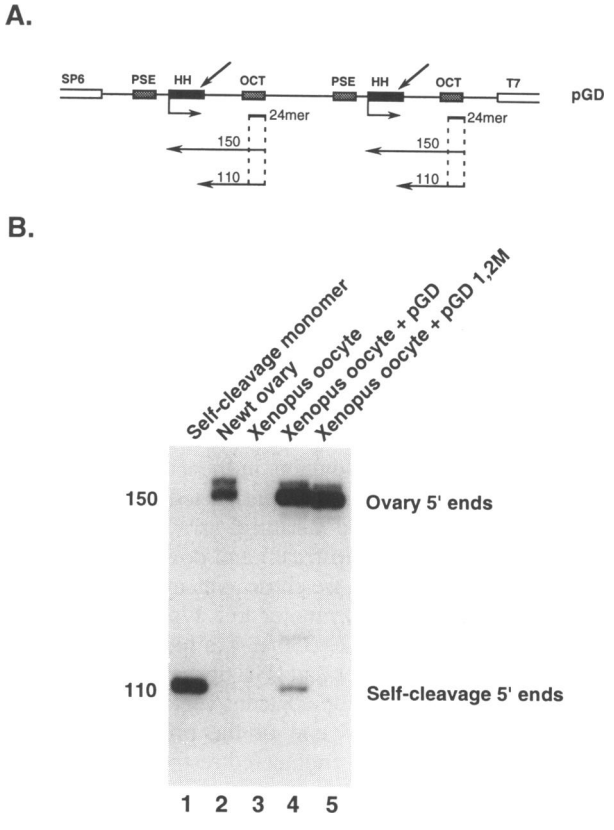

Figure 1. Primer extension analysis of transcripts produced from injected satellite 2 clones. **A)** Organization of the satellite 2 dimer clone pGD that was used for injections into *Xenopus* oocytes. The relative locations are indicated for the proximal sequence elements (PSE), the octamer motifs (OCT), and the self-cleaving hammerhead domains (HH). Diagonal arrows indicate the sites of self-cleavage in transcripts from pGD; horizontal arrows indicate the positions that correspond to both the 5′ ends of the ovary monomers and the sites of transcription initiation as determined in part B and Figure 2. Also shown are the regions of homology to the 24mer primer used in the primer extension analysis and the previously determined sizes of the extension products using the 24mer and either the ovary-specific or self-cleavage generated monomers (6). pGD1,2M is identical to pGD except for single nucleotide deletions which inactivate each of its hammerhead domains (15). **B)** Approximately 50 nl of cloned DNA at a concentration of 400 pg/nl was injected into stage 5–6 *Xenopus* oocytes, and RNA was pooled from 15–20 oocytes for analysis by primer extensions using the 24mer. Lane 1, extension products using 1 ng of purified monomers generated by self-cleavage of pGD sense transcripts; lane 2, extension products using satellite 2 transcripts present in 18 μg of newt ovary RNA; lane 3, extension products using 4 μg of RNA from uninjected *Xenopus* oocytes; lanes 4 and 5, extension products using 4 μg of RNA from oocytes injected with the indicated satellite 2 clone.

To determine if 3′ ends characteristic of newt ovarian monomers were also generated in *Xenopus*, RNA from injected oocytes was subjected to a RNase protection analysis (Fig. 2). If transcripts with ovary-type 5′ and 3′ ends were produced, they would have been monomers that started and ended at the same site in adjacent repeats of the pGD dimer. Monomeric transcripts were detected in oocytes injected with pGD (lane 5). However, these appear to have been generated by the self-cleavage of readthrough transcripts from the first initiation site. Evidence for this is that intact self-cleavage sites were required for their formation (note the near absence of the monomer in lane 6 which shows the result of injecting pGD1,2M), and an end analysis of the purified monomers indicated that they had ends that mapped to the site of self-cleavage (data not shown). They therefore did not have the configuration of the newt ovarian monomers, and the termination or processing event that generates the 3′ ends of the ovarian monomers did not occur, or was extremely inefficient in *Xenopus*.

Several other fragments of the antisense pGD probe were protected by the RNA from injected oocytes. Two of these had sizes that corresponded to the 554 and 221 nt fragments expected from protection by readthrough transcripts originating at the first and second initiation sites, in agreement with the assignment of these as initiation sites. The prevalence of the 221 nt fragment indicates that transcription was more efficient from the downstream site, possibly due to its upstream octamer. The identities of the other minor fragments evident in lanes 5 and 6 of Figure 2 were not determined, but they may represent cryptic or inefficient termination sites or RNase protection artifacts.

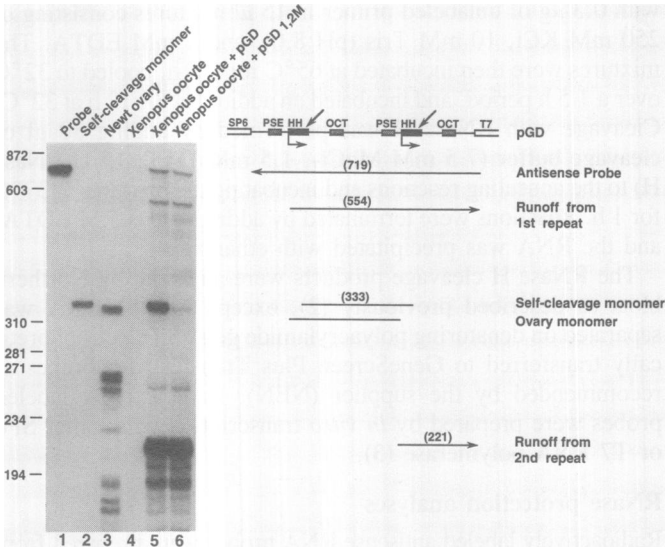

Figure 2. RNase A/T1 protection analysis of transcripts produced from injected satellite 2 clones. Lane 1, the antisense probe prepared by transcribing pGD *in vitro* with T7 RNA polymerase; lane 2, protection of probe by 1 ng of purified monomers produced by *in vitro* self-cleavage of pGD sense transcripts; lane 3, protection by 12 μg of total newt ovary RNA; lane 4, protection by 8 μg of RNA from uninjected *Xenopus* oocytes; lanes 5 and 6, protection by 8 μg of RNA from oocytes injected with the indicated satellite 2 clones. Sizes in nt of coelectrophoresed ϕX *Hae*III fragments are indicated on the left. The expected sizes of the various protection fragments are indicated in parentheses in the diagram on the right.

The putative PSE and octamer sequences are functional transcriptional elements

To examine the functional domains in the satellite 2 promoter, a series of promoter deletions was constructed for microinjection into *Xenopus* oocytes. While a natural array of satellite 2 has promoter elements in both upstream and downstream positions relative to each initiation site, we chose to assay a single promoter in the same context as the promoter in a typical snRNA gene. For this purpose we used pG−336/+4, a monomeric subclone of pGD which has an octamer and PSE upstream of an initiation site that is poised to transcribe adjacent vector sequences (Fig. 3A). Deletions from the 5′ end of the promoter were then constructed using existing or engineered restriction enzyme sites in pG−336/+4 (Fig. 3A).

The transcriptional properties of these constructs were assayed by primer extensions of total RNA from injected oocytes (Fig. 3B). A *Xenopus borealis* somatic 5S RNA maxigene was coinjected with each sample to serve as an internal standard, and is the source of the 136 nt extension product seen in all lanes of Figure 3B. Transcription was consistently greater from either pG−336/+4 or pG−285/+4 (lanes 1 and 2) than from any of the other constructs. The first significant drop in transcription occurred from construct pG−195/+4 (lane 3) in which the consensus octamer motif was deleted and which expressed at about 40% of the level of pG−336/+4. Expression levels did not change with the next two deletions (lanes 4 and 5) but dropped another 50% when the deletion extended to the 5′ border of the PSE (pG−63/+4, lane 6). Finally, transcription was not detected when the PSE was entirely deleted (pG−46/+4, lane 7).

Included in Figure 3A are the relative expression levels of the deletion clones calculated from the experiment shown in Figure 3B (experiment II), as well as from an independent replicate of the same experiment (experiment I). While the general trends observed in these experiments are similar to the trends found in similar studies of snRNA gene promoters, the 2−3 fold reduction in transcription from clones deleted for the region encompassing the octamer motif is less dramatic than the 10−20 fold reductions reported for deletions of the DSE (and its component octamer motif) from the promoters of some amphibian snRNA genes (21−23). This result does not necessarily reflect functional differences in the satellite 2 and snRNA promoters. When a *Xenopus* U1b2 maxigene was tested using the same methodology as that used to analyze satellite 2 transcription, deleting the DSE resulted in a similar, 1.5 fold reduction in transcription (Fig. 3B, lanes 10 and 11). Interestingly, the measured effect of the DSE deletion changed as the amount of template injected was decreased. In one experiment, 20 ng/gv of the U1b2 DSE deletion was expressed at 37% of the wildtype levels while 2 ng/gv and 0.2 ng/gv were expressed at 13.5% and 5.9% of the wildtype levels, respectively (data not shown). Similar results were obtained for the satellite 2 clone deleted for its octamer motif when it was tested at lower concentrations (data not shown). Concentration effects of this nature were reported for DSE mutations in the *Xenopus* U3 snRNA gene promoter (14), and we believe that they are the primary cause of the discrepancies between our measurements and those of previous studies. More importantly, when assayed under similar conditions and by the same methodology, transcription by the satellite 2 and U1b2 promoters is enhanced to similar degrees by their upstream elements.

To confirm that the reduced levels of transcription from the deletion clones were due to the loss of the octamer and PSE

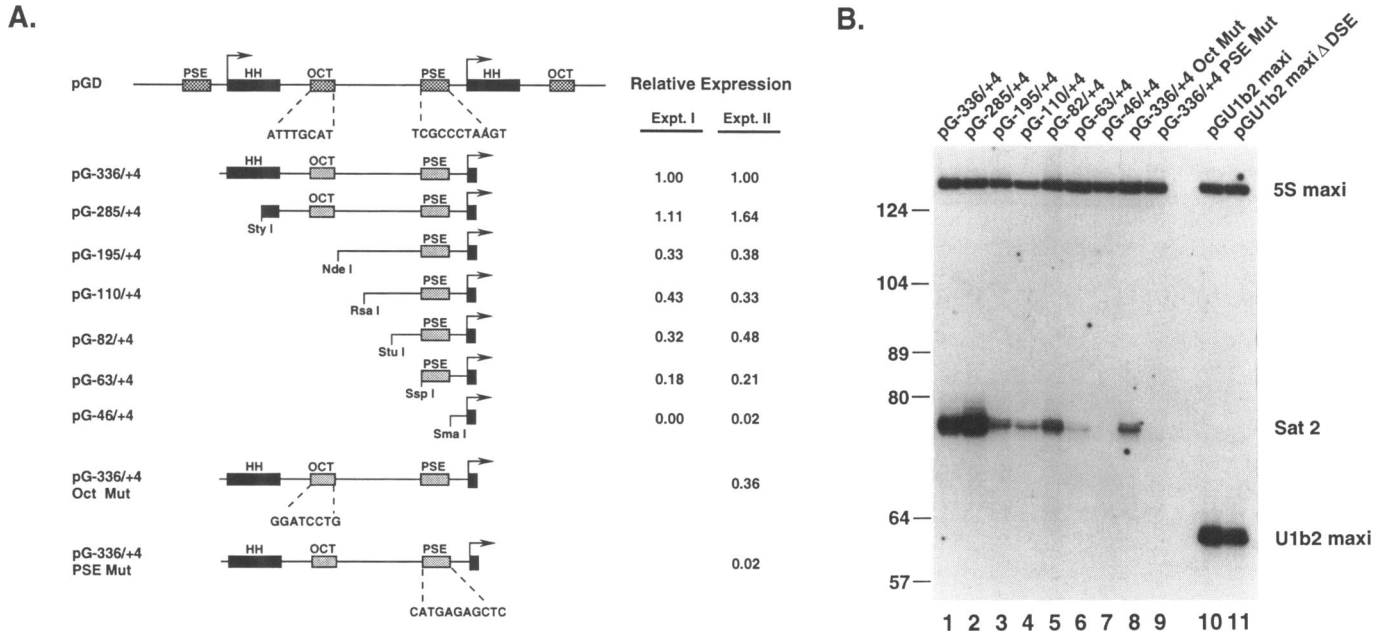

Figure 3. Deletion and mutation analysis of the satellite 2 promoter. **A)** Constructs used for the injections shown in part B. The names of the deletion clones refer to the number of satellite 2 nucleotides upstream (−) and downstream (+) of the transcription initiation sites (horizontal arrows). The sequences of the octamer and PSE elements in the Oct Mut and PSE Mut clones are shown, and can be compared to the native octamer and PSE sequences shown below the pGD diagram. Also shown are the levels of expression of these clones relative to pG−336/+4 as determined from the experiment shown in part B (Expt. II) as well as from an independent replicate of this experiment (Expt. I). **B)** *Xenopus* oocytes were injected with approximately 50 nl of the indicated satellite 2 or U1b2 maxigene constructs at concentrations of 400 pg/nl. A *Xenopus borealis* 5S maxigene was included in each sample at a concentration of 2 pg/nl to serve as an internal standard. RNA was prepared from oocytes and mixed primer extensions were performed on 8 μg of total RNA. For these mixed extension reactions, the primer specific for the 5S RNA maxigene insert was used in conjunction with either the M13 forward sequencing primer to measure transcription from the satellite 2 promoter into the adjacent vector (lanes 1−9), or the primer specific for the insert in the U1b2 maxigene (lanes 10 and 11). Sizes in nt of coelectrophoresed pBR322 *Hae*III fragments are indicated to the left, and the positions of the satellite 2 (Sat2), U1b2 maxigene (U1b2 maxi), and the 5S maxigene (5S maxi) extension products are indicated to the right.

motifs, we tested the expression of clones which were mutated specifically in these regions. pG−336/+4 Oct Mut had 8 nucleotide substitutions in its octamer motif (Fig. 3A), and was expressed to a similar level as the deletion clone pG−195/+4 (Fig. 3B, lanes 3 and 8). pG−336/+4 PSE Mut had 11 substitutions in its PSE (Fig. 3A), and was transcriptionally inert (Fig. 3B, lane 9). From the deletion and mutational studies we conclude that these two elements are important components of the satellite 2 promoter. As with the snRNA gene promoters, the octamer enhances transcription but is dispensable, while the PSE is absolutely required for transcription.

The satellite 2 promoter can substitute for a snRNA gene promoter in 3′ end formation

The data presented in the previous sections indicate that the satellite 2 promoter shares structural and functional properties with the typical vertebrate snRNA gene promoter. However, a true snRNA gene promoter is further characterized by its ability to elicit recognition of the snRNA 3′-end signal (3′ box) located 10−30 nt downstream from the coding region (24−26). In the oocyte injection studies, no prevalent 3′ ends other than those generated by self-cleavage were observed. However, this does not demonstrate that the satellite 2 promoter is not compatible with a snRNA 3′ box, since it is not clear if or even where a 3′ box should be in satellite 2. Given the uncertainties of what constitutes the 'coding' and 'flanking' regions of satellite 2 (see Discussion), we sought to test the ability of the satellite 2 promoter to interact with a functional 3′ box in the context of a known snRNA coding region. For this purpose, the chimeric genes pGsat−26/U1−25 and pGsat+43/U1+39 were constructed which placed the *Xenopus* U1b2 maxigene under the transcriptional control of the satellite 2 promoter (Fig. 4A). The difference in these clones was that pGsat−26/U1−25 retained 25 nt of the U1b2 maxigene promoter, while the entire promoter in pGsat+43/U1+39 was derived from the satellite 2 clone, as were the first 43 nt of the coding region.

The RNAs produced after injecting the U1b2 maxigene or the chimeric constructs were first analyzed by primer extensions to verify the fidelity of 5′-end formation (Fig. 4B). RNA from uninjected oocytes gave the 101 nt extension product that was expected for extension from endogenous U1 transcripts (lane 1). When pGU1b2 maxi (lane 2), pGsat−26/U1−25 (lane 3), or pGsat+43/U1+39 (lane 4) were injected, additional extension products were generated that had the sizes expected if transcription initiated correctly at the U1b2 initiation site (pGU1b2 maxi or pGsat−26/U1−25), or at the satellite 2 initiation site (pGsat+43/U1+39). Primer extension sequencing confirmed that the initiation sites were identical for both the pGU1b2 maxigene and the pGsat−26/U1−25 chimera (data not shown). Since comparable quantities of extension products were detected using all three constructs, the satellite 2 promoter initiated transcription as accurately and efficiently as the native U1b2 maxigene promoter.

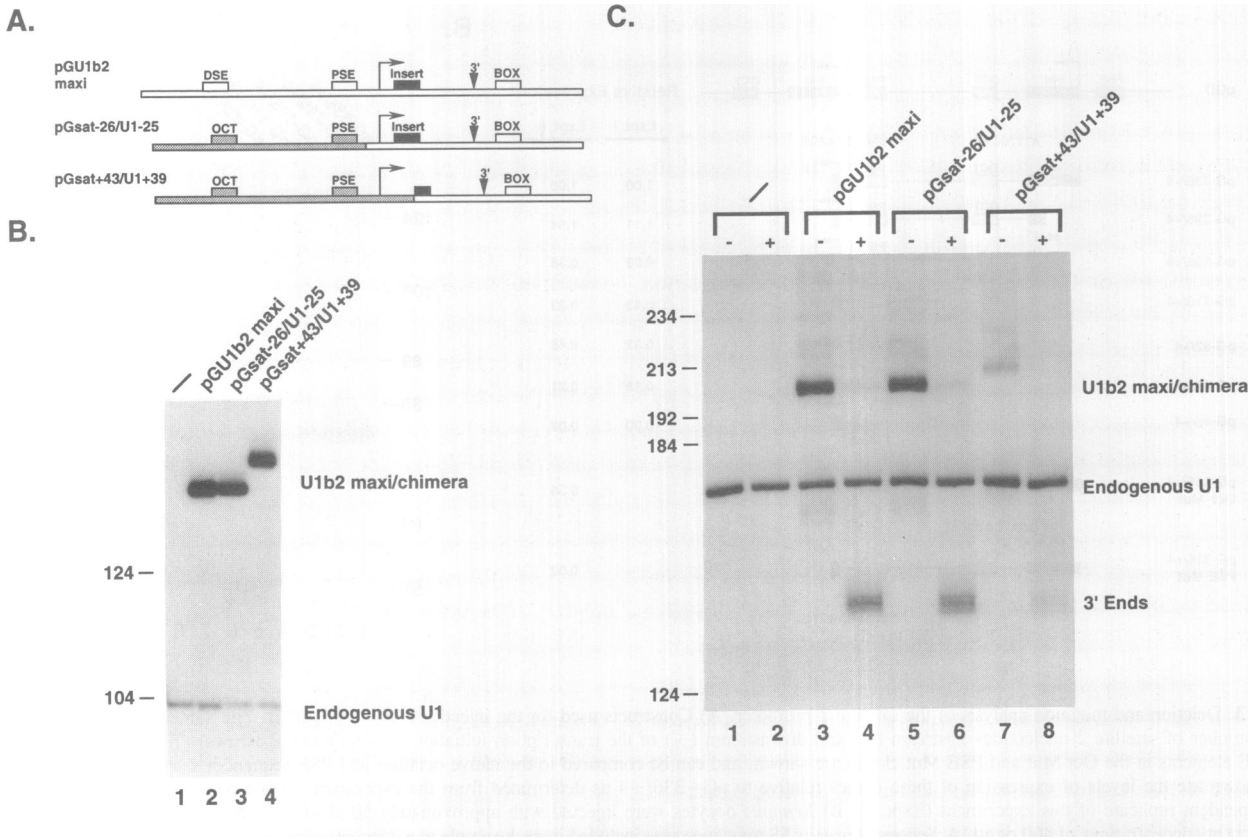

Figure 4. The satellite 2 promoter is compatible with a snRNA gene 3′ box. **A)** U1b2 maxigene construct and satellite 2/U1 chimeric constructs. The distal sequence element (DSE), the proximal sequence element (PSE), and the 3′ box (BOX) of the U1b2 construct are shown as open boxes. The 30 nt insert in the U1b2 maxigene is illustrated as a black box. Horizontal arrows designate the transcription start sites; vertical arrows denote the location of the 3′ end of the mature U1b2 transcript. In the satellite 2/U1b2 chimeras, sequences derived from satellite 2 are represented by gray shading. **B)** Approximately 50 nl of the indicated clones at a concentration of 40 pg/nl was injected into oocytes and 7 μg of total RNA from the injected oocytes was analyzed by primer extensions. The primer used for this analysis hybridized to the U1 coding region downstream of the maxigene insert, and gave extension products with endogenous U1 RNA as well as with transcripts from the U1b2 maxigene (U1b2 maxi) or the chimeric constructs (chimera). Lane 1 shows a primer extension using RNA from uninjected *Xenopus* oocytes. Sizes in nt of pBR322 *Hae*III fragments are shown to the left. **C)** RNase H cleavage analysis of the 3′ ends of transcripts generated from injecting satellite 2/U1 chimeric clones. Each lane contains 8 μg of the indicated RNA samples that were either not treated (−) or treated (+) with RNase H. The primer used to target RNase H digestion was specific for the maxigene insert in the U1b2 maxigene constructs, and therefore did not react with endogenous U1 transcripts. Samples were electrophoresed on a 7.5% polyacrylamide, 7 M urea gel, transferred to a nylon membrane, and probed with the same oligonucleotide used for the primer extensions in panel B. Sizes of pBR322 *Hae*III fragments are indicated to the left, and the positions of the endogenous U1 transcripts (Endogenous U1), the U1b2 maxigene transcripts (U1b2 maxi), the chimeric constructs (chimera), and the 3′-end fragments generated by RNase H treatment, are indicated on the right.

To determine if the 3′ boxes were being properly utilized in the chimeric constructs, the 3′ ends of the transcripts were mapped using RNase H (Fig. 4C). In this assay, a deoxyoligonucleotide primer specific for the insert in the U1b2 maxigene was hybridized to RNA isolated from injected oocytes, and the RNA strands of the resulting DNA/RNA duplexes were digested with RNase H. The transcripts and their RNase H cleavage products were then visualized by a Northern blot analysis using a probe that was complementary to a region downstream from the maxigene insert, and was therefore specific for the 3′ RNase H digestion products. Lanes 1 and 2 are controls that demonstrate that the 164 nt endogenous U1 transcripts were not affected by treatment with RNase H since they did not contain the maxigene insert. The 194 nt U1b2 maxigene transcripts, on the other hand, were successfully targeted for digestion and three major 3′-end fragments were generated which differed in size by one or a few nucleotides (lane 4). While this triplet of products may have resulted from imprecise digestion by RNase H, similar reaction conditions resulted in singlet products from a homogeneous population of control transcripts synthesized *in vitro* (data not shown). Furthermore, a shorter exposure of the autoradiogram depicted in Figure 4C indicated that the full length U1b2 maxigene transcripts were heterogeneous in size before digestion with RNase H. It is therefore more likely that the heterogeneity of the RNase H digestion products is the result of

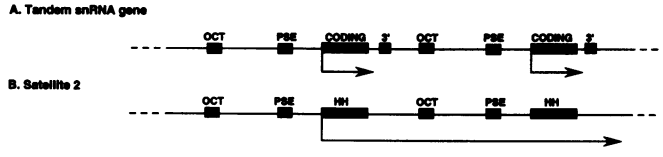

Figure 5. Similar organization of *Ambystoma mexicanum* U1 and satellite 2 tandem repeats. The U1 and satellite 2 genes were drawn to different scales so that similar elements within these genes could be aligned. Abbreviations are as in Figure 1.

bona fide heterogeneity of the 3′ ends of the U1b2 transcripts as has been previously reported (27, 28).

Transcripts produced from pGsat−26/U1−25 were identical in size and abundance to the transcripts from pGU1b2 maxi (compare lane 5 to lane 3), and had 3′ ends that were indistinguishable by the RNase H assay from the 3′ ends of the pGU1b2 maxi transcripts (compare lane 6 to lane 4). This result indicates that nucleotides −26 to −336 of the satellite 2 promoter, which include the PSE and octamer motifs, can replace the analogous region of the U1b2 maxigene promoter and elicit utilization of the snRNA 3′ box and correct 3′-end formation.

pGsat+43/U1+39 produced transcripts that were slightly larger than the transcripts from the other templates since 38 nt of the U1b2 maxigene coding region were replaced with 43 nt of satellite 2 in this chimera (lane 7). Despite the fact that the pGsat+43/U1+39 transcripts accumulated to lower levels than transcripts from the other clones, the 3′ ends of these transcripts were accurately formed (lane 8). Since the primer extension assay indicated that transcription initiation was as efficient from pGsat+43/U1+39 as from the other clones (Fig. 4B), the decreased production of mature transcripts was due to either a decrease in the efficiency of termination or to a destabilization of the mature transcripts caused by the satellite 2 sequence at their 5′ ends. In either case, the satellite 2 promoter can replace the entire U1b2 promoter and elicit accurate 3′-end formation.

DISCUSSION

Satellite 2 has a snRNA type of promoter

Cloned satellite 2 sequences from the European newt, *T. vulgaris meridionalis* (5), and from the eastern newt, *N. viridescens* (this work), have internal elements which promote accurate and efficient transcription upon injection into *Xenopus* oocytes. Cremisi *et al.* determined that transcription of the *T. vulgaris meridionalis* clone was sensitive to α-amanitin concentrations that are diagnostic of transcription by RNA polymerase II (5). Furthermore, the authors identified sequences in satellite 2 that resembled the octamer and PSE elements in the promoters of vertebrate snRNA genes transcribed by RNA polymerase II (10). We verified that these sequences are involved in transcription by satellite 2 and that their roles in transcription are consistent with the established roles of the DSE and PSE in the snRNA gene promoters.

In addition to the octamer motif, many snRNA gene DSEs have other protein binding domains that are required for maximal activity (11−14). It is highly probable that analogous domains exist in the satellite 2 DSE, since the satellite 2 promoter was as efficient as the native U1b2 promoter in promoting transcription of a *Xenopus* U1b2 snRNA gene (Fig. 4). Furthermore, the satellite 2 promoter was an efficient competitor of the U1b2 promoter in mixed oocyte injection studies (S.R.Coats and L.M.Epstein, unpublished data). While there are no obvious similarities between the satellite 2 and snRNA gene promoters outside of their octamer motifs, there are regions upstream and downstream of the octamer in satellite 2 that are highly conserved in five different families of salamanders (1; B.A.Green, Y.Zhang, and L.M.Epstein, unpublished data), and are good candidates for being components of the satellite 2 DSE. Similarly, a conserved region lies just upstream of the PSE, and may be responsible for the small but reproducible reduction in transcription when the region between −63 and −82 was deleted (Fig. 3).

Further evidence that the satellite 2 and major snRNA gene promoters are functionally related was obtained from an analysis of chimeric genes made by fusing portions of satellite 2 to the *Xenopus* U1b2 maxigene. We found that the entire U1b2 promoter could be replaced with the satellite 2 promoter without a dramatic reduction in 3′-end formation. This verifies that the satellite 2 promoter is the functional counterpart of a major snRNA gene promoter since a functional snRNA gene promoter is required for recognition of the 3′ box and correct 3′-end formation (24−26).

Relation between transcription, self-cleavage and tissue-specific transcript permutations

While self-cleavage by the 'extended hammerhead' domain appears to be involved in the formation of satellite 2 monomers in tissues such as the liver and testes (6), an alternative mechanism is involved in the maturation of the predominant monomers in the newt ovary since the 5′ ends of these transcripts are 42 nt upstream from an intact cleavage site (2, 6). Transcription from the snRNA-like promoter can at least partially explain their derivation since transcription in *Xenopus* oocytes initiates at a site that coincides with the 5′ ends of the ovarian monomers. If transcription by RNA polymerase II is responsible for the formation of their 5′ ends, the ovarian monomers should have methylated cap structures (10, 29). We previously reported that the ovarian monomers were efficient substrates for T4 polynucleotide kinase (6), which suggests that they are not capped. However, we have since learned that the labeling observed in those experiments was due to the presence of low levels of self-cleavage generated monomers in newt oocytes, and bona fide ovarian 5′ ends are blocked from reacting with T4 polynucleotide kinase (S.R.Coats and L.M.Epstein, unpublished data). Furthermore, the block on some, but not all of the ovarian transcripts can be removed with tobacco acid pyrophosphatase (TAP) which effectively removes cap structures characteristic of RNA polymerase II transcripts (30, 31). Therefore, it is likely that some of the ovarian monomers have 5′ ends that are generated by transcription initiation. Interestingly, however, some of the ovarian monomers have 5′-blocking groups that appear to be resistant to TAP treatment.

The formation of the 3′ ends of the ovarian monomers remains an enigma. Despite having a promoter that is compatible with a snRNA gene 3′ box, there are no obvious 3′ boxes downstream from the position of the ovarian 3′ end, and the only 3′ ends that were formed efficiently in *Xenopus* oocytes were generated by self-cleavage. It is possible that satellite 2 uses a modified 3′-end signal and the truncated upstream promoter in pGD was unable to recruit the transcription factors needed to recognize this signal. However, when upstream sequences including the octamer were added to the pGD construct, ovary-type 3′ ends were still not formed in *Xenopus* oocytes (S.R.Coats, Y.Zhang, and L.M.Epstein, unpublished data). If transcription alone is responsible for ovarian 3′-end formation, it must involve novel species-specific termination factors. Alternatively, 3′ ends could be formed by the processing of readthrough transcripts. The generation of 3′ ends by this mechanism might also contribute to the formation of ovarian 5′ ends, since the 5′ and 3′ ends of the ovarian monomers occur at adjacent sites in a tandem array. An alternative mechanism for 5′-end formation might also explain the fact that some of the ovary monomers appear to have unusual blocking groups that are resistant to TAP treatment as described above.

Similarity between satellite 2 and snRNA genes

As described in the Introduction, the present study is part of a larger effort to elucidate the origins of the highly repetitive, tandemly repeated satellite 2 family. Because of similarities to certain plant viroid, virusoid, and satellite RNAs which use hammerheads or other self-cleaving motifs to process multimeric intermediates during the replication of their RNA genomes (32), we earlier speculated that satellite 2 was also derived from an infectious RNA (1–3). The finding of an integral transcriptional promoter that is functionally interchangeable with the U1b2 snRNA gene promoter argues against this simplistic model and suggests that an evolutionary relationship exists between satellite 2 and the major snRNA genes. In Figure 5, the organization of an array of satellite 2 repeats is compared to an array of *Ambystoma mexicanum* U1 genes. The *A.mexicanum* U1 array is typical of tandemly repeated snRNA genes in amphibians, and consists of 164 bp coding regions separated by 863 bp of spacer DNA where the transcriptional control elements are located (33). When viewed in this manner, it becomes apparent that in both cases similar promoter elements cause transcription to initiate at the beginning of a concise and functional RNA domain. In the U1 array, the functional domain is the U1 RNA itself; in satellite 2 it is the extended self-cleaving hammerhead. The difference is that while transcription terminates at the end of the U1 RNA, it does not terminate at the end of the hammerhead domain and proceeds into the downstream promoter. How these transcripts are terminated (or processed) in the vicinity of the downstream initiation site to generate the ovarian monomers is not known, but as discussed above, the mechanism does not seem to involve the 3′ box recognition properties of the snRNA-like promoter.

While it is possible that satellite 2 was generated by genomic recombination between a snRNA gene and another genetic element (a cDNA copy of an RNA virus, for example), it is equally plausible that its particular arrangement occurred in a preexisting snRNA gene. In this scenario, one of the events that generated satellite 2 was the loss of the termination signal in the snRNA gene. This mutation could have occurred before or during the formation of a tandem array, or could have spread through a pre-existing tandem array by some mechanism of homogenization (34). The significance of this scenario is that it implies that there is a functional snRNA gene that codes for a catalytic hammerhead RNA. This RNA could use self-cleavage as a mechanism of self-regulation, or the RNA could be a *trans*-cleaving RNA involved in the processing of some other cellular RNA in a manner similar to the processing of tRNAs by the M1 RNA component of RNase P (35). *Trans*-cleavage by hammerheads has been shown to be possible (36, 37), although no examples of naturally occurring *trans*-cleaving hammerheads have yet been found.

To further investigate this possibility we are searching the genomes of other vertebrates for sequences related to satellite 2. Southern analyses indicate that highly repetitive homologues of satellite 2 do not exist outside of the amphibian order Caudata. This will facilitate our search for elements that are expected to have moderate to low copy numbers and sequences that have diverged significantly from satellite 2. Using highly conserved regions of satellite 2 as probes in Southerns or as primers for polymerase chain reaction amplifications, we are hoping to identify the genomic elements that have contributed to the formation of satellite 2.

ACKNOWLEDGEMENTS

We thank Matt Andrews for providing the plasmid p115/77, Elsebet Lund and James E.Dahlberg for providing the plasmid pXlU1b and for their useful advice, Joan Hare and J.Herbert Taylor for graciously providing the equipment for microinjecting *Xenopus* oocytes, William F.Marzluff for useful comments, and Lil M.Pabón-Peña for her excellent technical assistance and critical discussions. This work was supported by NIH grant GM 38791 to L.M.E.

REFERENCES

1. Green, B.A., Pabón-Peña, L.M., Graham, T.A., Peach, S.E., Coats, S.R. and Epstein, L.M. (1993) *Mol. Biol. Evol.* **10**, 732–750.
2. Epstein, L.M., Mahon, K.A. and Gall, J.G. (1986) *J. Cell Biol.* **103**, 1137–1144.
3. Epstein, L.M. and Gall, J.G. (1987) *Cell* **48**, 535–543.
4. Pabón-Peña, L.M., Zhang, Y. and Epstein, L.M. (1991) *Mol. Cell. Biol.* **11**, 6109–6115.
5. Cremisi, F., Scarabino, D., Carluccio, M.A., Salvadori, P. and Barsacchi, G. (1992) *Proc. Natl Acad. Sci. USA.* **89**, 1651–1655.
6. Epstein, L.M. and Coats, S.R. (1991) *Gene* **107**, 213–218.
7. Leeflang, E.P., Liu, W.-M., Hashimoto, C., Choudary, P.V. and Schmid, C.W. (1992) *J. Mol. Evol.* **35**, 7–16.
8. Deininger, P.L., Batzer, M.A., Hutchison, C.A.III and Edgell, M.H. (1992) *Trends Genet.* **8**, 307–311.
9. Ullu, E. and Tschudi, C. (1984) *Nature* **312**, 171–172.
10. Dahlberg, J.E. and Lund, E. (1988) in *Structure and Function of Major and Minor SNURPS*, Birnstiel, M.L. Ed. pp. 38–70, Springer-Verlag, New York.
11. Kazmaier, M., Tebb, G. and Mattaj, I.W. (1987) *EMBO J.* **6**, 3071–3078.
12. Parry, H.D., Scherly, D. and Mattaj, I.W. (1989) *TIBS* **14**, 15–19.
13. Roebuck, K.A., Szeto, D.P., Green, K.P., Fan, Q.N. and Stumph, W.E. (1990) *Mol. Cell. Biol.* **10**, 341–352.
14. Savino, R., Hitti, Y. and Gerbi, S.A. (1992) *Nucleic Acids Res.* **20**, 5435–5442.
15. Epstein, L.M. and Pabón-Peña, L.M. (1991) *Nucleic Acids Res.* **19**, 1699–1705.
16. Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) *Meth. Enzymol.* **154**, 367–382.
17. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) *J. Mol. Biol.* **143**, 161–178.
18. Bogenhagen, D.F. and Brown, D.D. (1981) *Cell* **24**, 261–270.
19. Lund, E., Bostock, C.J. and Dahlberg, J.E. (1987) *Genes Dev.* **1**, 47–56.
20. Colman, A. (1984) in *Transcription and translation, a practical approach*, Hames, B.D. and Higgins, S.J. Eds. pp. 49–69, IRL Press, Oxford.
21. Ciliberto, G., Buckland, R., Cortese, R. and Philipson, L. (1985) *EMBO J.* **4**, 1537–1543.
22. Krol, A., Lund, E. and Dahlberg, J.E. (1985) *EMBO J.* **4**, 1529–1535.
23. Mattaj, I.W., Lienhard, S., Jiricny, J. and De Robertis, E.M. (1985) *Nature* **316**, 163–167.
24. Hernandez, N. and Weiner, A.M. (1986) *Cell* **47**, 249–258.
25. Neuman de Vegvar, H.E., Lund, E. and Dahlberg, J.E. (1986) *Cell* **47**, 259–266.
26. Hernandez, N. and Lucito, R. (1988) *EMBO J.* **7**, 3125–3134.
27. Neuman de Vegvar, H.E. and Dahlberg, J.E. (1990) *Mol. Cell. Biol.* **10**, 3365–3375.
28. Yang, H., Moss, M.L., Lund, E. and Dahlberg, J.E. (1992) *Mol. Cell. Biol.* **12**, 1553–1560.
29. Shatkin, A.J. (1976) *Cell* **9**, 645–653.
30. Shinshi, H., Miwa, M. and Sugimura, T. (1976) *FEBS Lett.* **65**, 254–257.
31. Kiss, T., Marshallsay, C. and Filipowicz, W. (1991) *Cell* **65**, 517–526.
32. Symons, R.H. (1989) *TIBS* **14**, 445–450.
33. Murgo, S., Krol, A. and Carbon, P. (1991) *Gene* **99**, 163–170.
34. Weiner, A.M. and Denison, R.A. (1983) *Cold Spring Harb. Symp. Quant. Biol.* **47**, 1141–1149.
35. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) *Cell* **35**, 849–857.
36. Uhlenbeck, O.C. (1987) *Nature* **328**, 596–600.
37. Haseloff, J. and Gerlach, W.L. (1988) *Nature* **334**, 585–591.