Nucleic Acids Research, Vol. 20, No. 14 3743–3751

Two promoter elements are necessary and sufficient for expression of the sea urchin U1 snRNA gene

Brian J.Wendelburg and William F.Marzluff*
Department of Chemistry, Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA

Received March 2, 1992; Revised and Accepted June 12, 1992

ABSTRACT

The essential elements of the sea urchin *L. variegatus* U1 snRNA promoter were mapped by microinjection of a U1 maxigene into sea urchin zygotes. Two elements are required for expression: a distal sequence element (DSE) located between −318 and −300 and a proximal sequence element (PSE) centered at −55. Removal or alteration of other sequences conserved in different sea urchin snRNA U1 genes, including deletion of all sequence between −90 and −273, did not affect the expression. Sequences around the start site were not required for expression. Deletion of nucleotides between the PSE and the start site resulted in initiation inside the U1 coding region, suggesting that the PSE determines the start site of transcription. There is no obvious similarity between the sequences required for the sea urchin U1 snRNA expression and the sequences required for the expression of other sea urchin snRNAs.

INTRODUCTION

There are three major classes of genes transcribed by RNA polymerase II; genes encoding polyadenylated mRNAs, genes encoding the non-polyadenylated histone mRNAs and genes encoding many of the small nuclear RNAs involved in RNA processing reactions. The promoters of the vertebrate snRNA genes differ dramatically from the promoters of genes encoding mRNAs (1). The vertebrate nucleoplasmic snRNA genes transcribed by RNA polymerase II lack a TATA box and contain two major elements required for expression; a proximal sequence element (PSE), located at about −50, which is loosely conserved among all the vertebrate snRNAs transcribed by RNA polymerase II and a distal sequence element (DSE) whose position is variable but is generally located about 200 nts from the start of transcription. The DSE element contains binding sites for well-defined transcription factors (e.g., oct-1, NF-1 and Sp1) which are involved in transcription of many genes encoding mRNAs (2–4). In addition, the DSE has many properties similar to an enhancer, functioning in a distance and orientation independent manner (5–7). The start site of transcription is determined by the PSE (8).

Much less is known about the structure of invertebrate snRNA promoters. There is a highly conserved element 5′ of the *C. elegans* snRNA genes which has been proposed to function as a PSE (9) although no functional studies have been done. The sea urchin snRNA genes are not expressed after injection into *Xenopus* oocytes (10) suggesting that the sequence requirements and factors involved in expression of the sea urchin and vertebrate snRNA genes are different.

The sea urchin *L. variegatus* U2 snRNA promoter has recently been characterized and shown to contain four elements involved in transcription, including both an essential TATA box at −25 and a PSE element at −55 (11). There are distal sequence elements but no single distal element has more than four-fold effect on expression. Here we examine the promoter of the *L. variegatus* U1 snRNA gene in an homologous expression system, based on microinjection into sea urchin zygotes. Both proximal and distal sequence elements are essential for expression, but these do not resemble any sequences within the sea urchin U2 snRNA gene or in other vertebrate or invertebrate snRNA genes. The PSE has an analogous function to the vertebrate PSE, determining the start site of transcription.

MATERIALS AND METHODS

Construction of U1 genes with altered promoters

The U1.2 maxigene was constructed by cloning a 19 nt synthetic oligonucleotide into the AvaI site at position 112 in the U1.2 coding region and has been described (12). This gene contains 419 nts 5′ of the start of the U1 RNA and is referred to as clone −419. The U1 maxigene was used for construction of the promoter mutants. Deletion clones were constructed using the Exo III/ Mung Bean Nuclease digestion method (13) or using Bal 31 nuclease.

Construction of insertion and deletion mutants near the U1 start site

Deletion mutants in the proximal promoter region were constructed by digesting the U1 maxigene with EagI which cuts at −19. The DNA was digested with EagI and then with Bal 31 for various time intervals. The DNAs were then religated after using Klenow fragment to blunt the ends. Deletion clones

* To whom correspondence should be addressed at: Program in Molecular Biology and Biotechnology, 132 MBBRL, CB#7100, University of North Carolina, Chapel Hill, NC 29755, USA

Δ−7/−78, Δ−5/−25, and Δ−5/−16 were identified by sequencing. Similar deletion mutations were constructed using a BamHI site introduced at position −167 by site directed mutagenesis. The insertion mutant, −19i, was constructed by digestion with EagI and filling in the 5′ overhang ends with the Klenow fragment of DNA polymerase I and religating with T_4 DNA ligase. All other mutants were constructed by site-directed mutagenesis using the method of Kunkel (14).

Microinjection of DNA into sea urchin zygotes

Microinjection of DNA into *S. purpuratus* and *L. pictus* zygotes was performed by the method of McMahon and coworkers (15), as modified by Colin and coworkers (16) and previously used for analysis of sea urchin snRNA genes (12). The test genes were usually injected together with a control gene, the $LvU2_H$ gene, which contains the LvU2E snRNA promoter (11) and the 5′ half of the *L. variegatus* U2 gene and the 3′ half of the U1 gene. The internal control gene and all test genes were linearized at the 3′ end and the concentration of each adjusted to 50 μg/ml. Test gene and internal control DNAs were then mixed in equal volumes so that the overall DNA concentration remained constant. Ten to 20 picoliters of DNA solution were injected per zygote and on average about 100 to 200 embryos were injected. Embryos were incubated at 15°C until the blastula (20 hr) stage. The embryos were collected at the appropriate stage and total cellular RNA isolated.

Analysis of RNA transcripts

The embryos were processed as described previously to prepare total cell RNA (12). The transcripts from each test gene and internal control gene were assayed using a mixed riboprobe RNase protection assay. The template for synthesis of the U1.2 riboprobe was the U1 maxigene linearized at the EagI site at −19. The template for synthesis of the U2 riboprobe was the $LvU2_M$ gene, a U2 maxigene constructed by inserting a synthetic oligonucleotide into the LvU2L gene and described previously (11). The riboprobes were mixed and the RNase protection assay carried out as described previously (11). The protected fragments were analyzed by electrophoresis on 8% polyacrylamide-7M urea gels and detected by autoradiography.

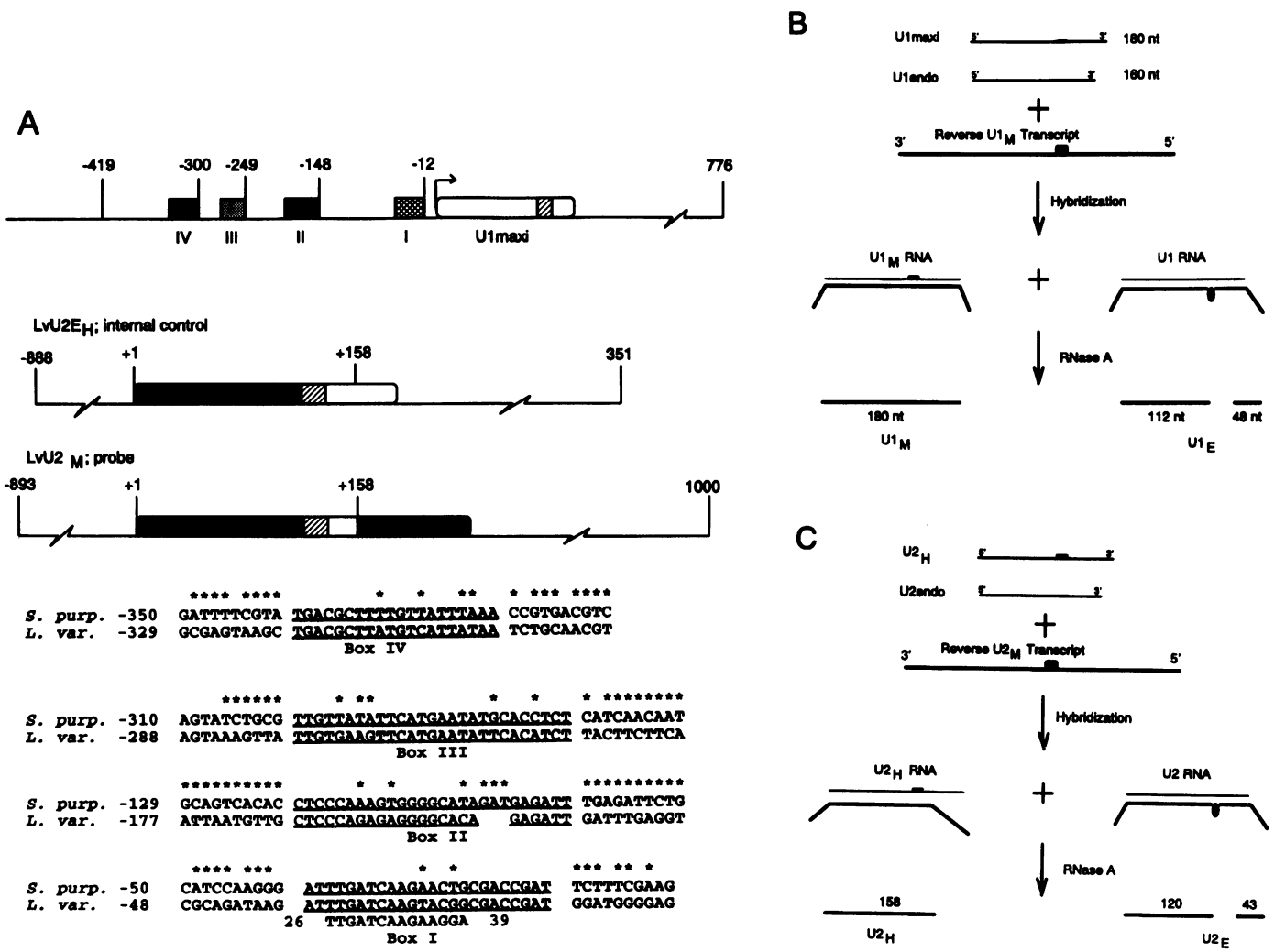

Fig. 1. Assay for expression of the U1 gene. A. The U1 maxigene, $U1_M$, and the two different internal control genes $U2E_H$ and $U2L_M$ are diagramed. The open box represents the U1 coding region, the solid box the U2 coding region and the hatched box the synthetic oligonucleotide inserted into the genes. The boxes indicate the conserved regions between the *L. variegatus* and *S. purpuratus* U1 genes. The sequences around the conserved boxes are shown below the figure. B. A schematic of the riboprobe assay used to detect the expression of the $U1_M$ gene is shown. The $U1_M$ transcript protects a 180 nt fragment of the $U1_M$ riboprobe and a 112 nt fragment from the endogenous sea urchin U1 snRNA. C. The $U2_H$ transcript was assayed using the $U2E_M$ riboprobe as a probe. The $U2_H$ transcript protects a 158 nt fragment of the $U2E_M$ probe and the endogenous sea urchin U2 snRNA protects a 120 nt fragment.

Gel mobility shift assays

Synthetic double-stranded oligonucleotides with protruding 5′ termini were synthesized and labeled by filling in the overhanging ends using α-dCTP and α-dATP (3000Ci/mmol) and the Klenow fragment of DNA polymerase I. Gel retardation assays were performed by mixing 2.5 ng of probe, 4.0 μg poly-dI/dC as a nonspecific competitor, and any competitor oligonucleotides in 10 mM Tris-Cl pH 7.5, 50 mM NaCl, 1.0 mM DTT, 1.0 mM EDTA, 5% glycerol. Sixteen micrograms of nuclear protein extracted from blastula embryos was added and the samples were incubated on ice for 30 min. The reaction was analyzed by electrophoresis on 5% or 8% nondenaturing polyacrylamide gels in 40 mM Tris-borate, 1 mM EDTA, pH 8.0. Electrophoresis was at 20 ma until the tracking dye, bromophenol blue, reached the bottom of the gel. The gels were fixed in 10% acetic acid-10% methanol for 1 hr, dried and the complexes detected by autoradiography.

RESULTS

Sea urchin U1 promoters

There are at least two sets of U1 snRNA genes in sea urchins, a tandemly repeated gene set expressed only in oocytes and in early embryogenesis (17) and a low-copy number gene set expressed constitutively, which encodes the only U1 snRNA expressed in adult sea urchins (18). We have previously isolated representatives of the tandemly repeated gene set from two different sea urchin species, *S. purpuratus* and *L. variegatus* (19,20). These species diverged about 35 million years ago (21). A representative of the tandemly repeated U1 gene family in *L. variegatus* is expressed well when it is injected into *S. purpuratus* zygotes (12), indicating that the factors present in one species recognize the genes from the other species. Comparison of the promoters of the SpU1 gene from *S. purpuratus* with the promoter of the *L. variegatus* early U1 gene revealed four short regions (boxes I−IV) of sequence homology (20)(diagramed in Fig. 1A). Our initial analysis focused on these conserved sequences.

Identification of a Distal Sequence Element (DSE)

To measure the activity of the U1 promoter, we used a U1 maxigene which can be distinguished from the endogenous U1 gene in an RNase protection assay. The U1 maxigene has a synthetic oligonucleotide inserted into the AvaI site at nt 112 of the U1 RNA sequence. To quantify the relative expression of the U1 gene, the genes were injected together with a modified U2 snRNA gene, usually the $LvU2_H$ gene, a hybrid gene containing the 5′ half of the *L. variegatus* U2E snRNA gene (11) and the 3′ half of the $LvU1_M$ maxigene. The U1 maxigene with the position of the conserved elements in the 5′ flanking region, and the LvU2 hybrid ($LvU2_H$) gene and maxigene ($LvU2_M$) are shown in Fig. 1A. Figure 1B shows the RNase protection assay used to assess the expression of the U1 maxigene. Figure 1C shows the assay for the internal control $LvU2_H$ gene. The probe for the $LvU2_H$ internal control gene was the $LvU2_M$ gene. This probe protects a 158 nt fragment from the control gene which is easily distinguished from the 180 nt fragment protected by the $LvU1_M$ gene. These probes were mixed to allow simultaneous assay of the test and control genes. Note that the size markers used on the gels were DNA markers and the RNA transcripts migrated about 10% slower than DNA fragments of similar size.

We have previously shown that 345 nts of 5′ flanking sequence were sufficient for maximal expression of the $LvU1_M$ gene and that there was no expression from a U1 gene containing 203 nts of 5′ flanking sequence (12). Further deletion analysis of the U1 promoter is shown in Fig. 2. These genes were injected into *S. purpuratus* zygotes together with the $LvU2_H$ gene as an internal control and the expression of the U1 transcripts compared with the expression of the control gene. A U1 maxigene containing 318 nts of 5′ flanking sequence was expressed at wild-type levels, while genes containing 300 or 240 nts of 5′ flanking sequence

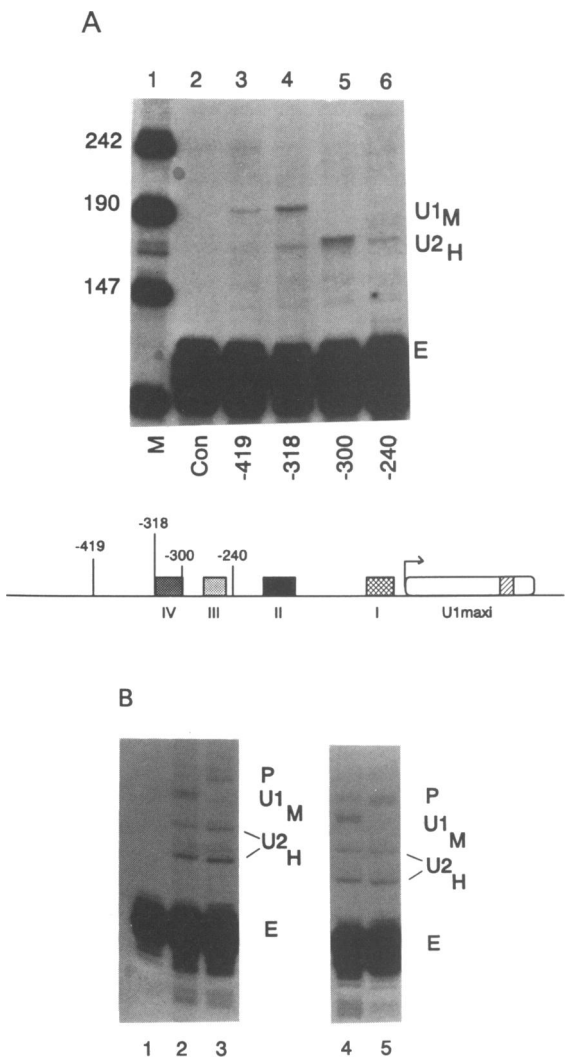

Fig. 2. Box IV is required for expression of the U1 gene. A. $U1_M$ genes containing 419, 318, 300 or 240 nts (lanes 3−6) of 5′ flanking sequence mixed with the $U2_H$ gene and injected into *S. purpuratus* zygotes. Lane 2 (Con) is uninjected control embryos. The RNAs were analyzed using the mixed riboprobe assay described in Fig. 1B and 1C. The protected fragments are: $U1_M$—transcript from the $U1_M$ gene; $U2_H$—transcript from the $U2_H$ gene; E—protection due to endogenous U1 and U2 snRNA. A schematic of the 5′ deletions is shown below. Lane 1 (M) is marker pUC18 digested with HpaII. B. $U1_M$ genes containing 318 (lanes 2 and 4) or 300 nts (lanes 3 and 5) of 5′ flanking sequence mixed with the $U2_H$ gene and injected into two separate batches of *S. purpuratus* zygotes. The RNAs were analyzed using the mixed riboprobe assay described in Fig. 1B and 1C. The protected fragments are: $U1_M$—transcript from the $U1_M$ gene; $U2_H$—transcript from the $U2_H$ gene; E—protection due to endogenous U1 and U2 snRNA. Lane 1 is RNA from uninjected embryos.

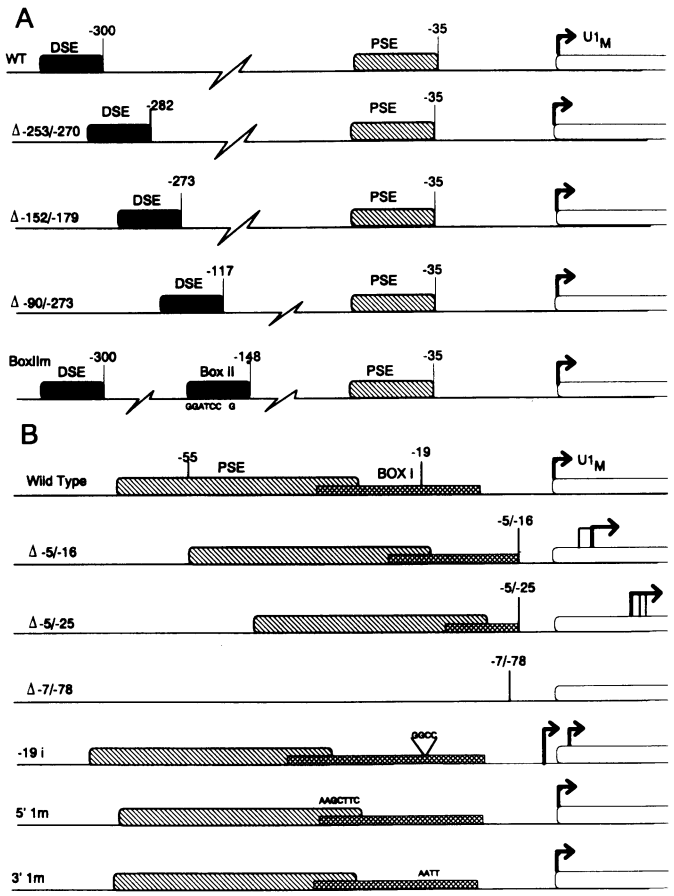

Fig. 3. Structure of promoter mutants. A. Mutants of box II and box III were constructed by site-directed mutagenesis or by Bal31 deletion as described in Materials and Methods. Box II_M had the sequence of box II altered. Δ−253/270 has box III deleted, Δ−152/−179 has box II deleted and Δ−90/−273 has both box II and box III deleted. B. Deletions in the proximal region of the promoter and mutations in the 5′ ($5'I_M$) and 3′ ($3'I_M$) portions of box I were constructed. The arrow indicates the start sites of the transcripts observed (see Fig. 4).

were not expressed (Fig. 2A, lanes 4−6). Figure 2B shows a similar result with two independent injections of U1 maxigenes constructs together with the $LvU2_H$ gene as an internal control. into separate batches of *S. purpuratus* embryos. Often when the $LvU2_H$ gene is injected into *S purpuratus* embryos, there were two fragments protected by the transcripts from this gene, which were present in a constant ratio within a particular assay (see below, Fig. 4B,C, Fig. 5). The U1 maxigene containing 318 nts was expressed (Fig. 2B, lanes 2 and 4) while the U1 maxigene containing 300 nts of 5′ flanking sequence was not expressed (Fig. 2B, lanes 3 and 5). These results localize the DSE to the region between −318 and −300, corresponding to the conserved box IV sequence. A factor which binds to this region has previously been identified by DNaseI footprinting (12).

Boxes I, II and III are not essential for transcription

To determine whether the other sequences conserved between the two sea urchin species make a contribution to transcription, mutations were created which altered or deleted the box I, II and III sequences. In addition, the entire sequence between −90 to −273, which includes both box II and III (gene Δ−90/−273), was removed. These genes are diagramed in Figures 3A and 3B. These constructs were injected into *L. pictus* (Fig. 4A) or *S. purpuratus* (Fig. 4B,C) embryos together with the $LvU2_H$ gene as an internal control. All the genes with deletions of box II and/or III were expressed well (Fig. 4A, lanes 4−8). In addition to the protected fragments from the injected genes, there is a protected fragment (labeled e) present in the control embryo sample, which is a result of hybridization of the *L. variegatus* probe to a transcript from the U1 gene present in the closely related *L. pictus* embryos (Fig. 4A, lane 2). This transcript extends 3′ of the U1 gene and is not detected in *S. purpuratus* embryos, presumably because the U1 genes have diverged between the two species. There is also a fragment (labeled p) derived from the $LvU2_M$ probe which is present in both the control and tRNA lanes (Fig. 4A, lanes 2 and 9). Deleting box III reduced expression at most 50%, indicating that this sequence by itself is not critical to expression (Fig. 4A, lane 6). Altering most of the nucleotides in box II (Fig. 4A, lane 5) or deleting box II and 12 additional 5′ nucleotides (Fig. 4A, lane 7) also did not affect expression significantly. The gene, Δ−90/−273, which completely deletes box II and III, and moves box IV close to the gene was expressed at 30−40% of the level of the intact gene. These results show that there are no critical sequences other than the box IV sequence distal to −90 in the promoter. The results were quantified by densitometry and are summarized in Table I.

The nucleotide sequence of box I begins at −12 nts from the start site and extends to −38. Changing 4 nts at the 3′ end of this sequence (Fig. 4A and B, lane 4) did not affect expression significantly. Additional mutants in this region were constructed to assess whether this region played any role in expression. In addition to altering the sequence, an insertion of 4 nts at −19 (−19i) or deletions of 20 (Δ−5/−25) or 11 (Δ−5/−16) nts were introduced. Analysis of parallel injections of these genes into the same batch of *S. purpuratus* embryos is shown in Figs. 4B and C.

All of these genes were expressed. Alteration of the 5′ portion of box I reduced expression about 60%−80% (Fig. 4B, lane 3). The experiment in Fig. 4B was done without an internal control gene but analysis of other mutations in the same region (see Fig. 5) showed a clear reduction in expression. Insertion of 4 nts at nt −19 altered the start site of transcription. There were no transcripts which initiated at the proper start site, but rather transcripts were synthesized which initiated just 5′ and just 3′ of the proper start site in equal amounts (Fig. 4B, lane 5; Fig. 4C, lane 6). These transcripts probably initiate at the A's at −2 and +2. The overall expression of the gene was at least 50% of the wild-type gene. Deletion of 20 nts, Δ−5/−25, resulted in formation of two shorter transcripts, which were about 20 nt smaller than the U1 maxigene transcripts (Fig. 4B, lane 6) and again overall expression was not significantly affected. Similarly, deletion of 11 nts, Δ−5/−16, also resulted in two shorter transcripts (Fig. 4C, lane 5). Both of these deletions gave multiple transcripts suggesting that transcription initiated at alternate sites within the U1 coding region. A deletion of 71 nts, Δ−7/−78, completely abolished expression (Fig. 4C, lane 4), and we show definitively below that there is a critical sequence located between −40 and −60 nts required for expression. In Figure 4C the internal control gene $LvU2_H$ was coinjected with the test genes. When the transcripts of the $LvU2_H$ gene are analyzed in *S. purpuratus* embryos, two transcripts are often detected, with the 158 nt transcript the same transcript detected after injection into *L. pictus* embryos. The ratio of the two transcripts is constant in a given batch of embryos but varies between experiments (cf. Figs. 2, 4C and 5).

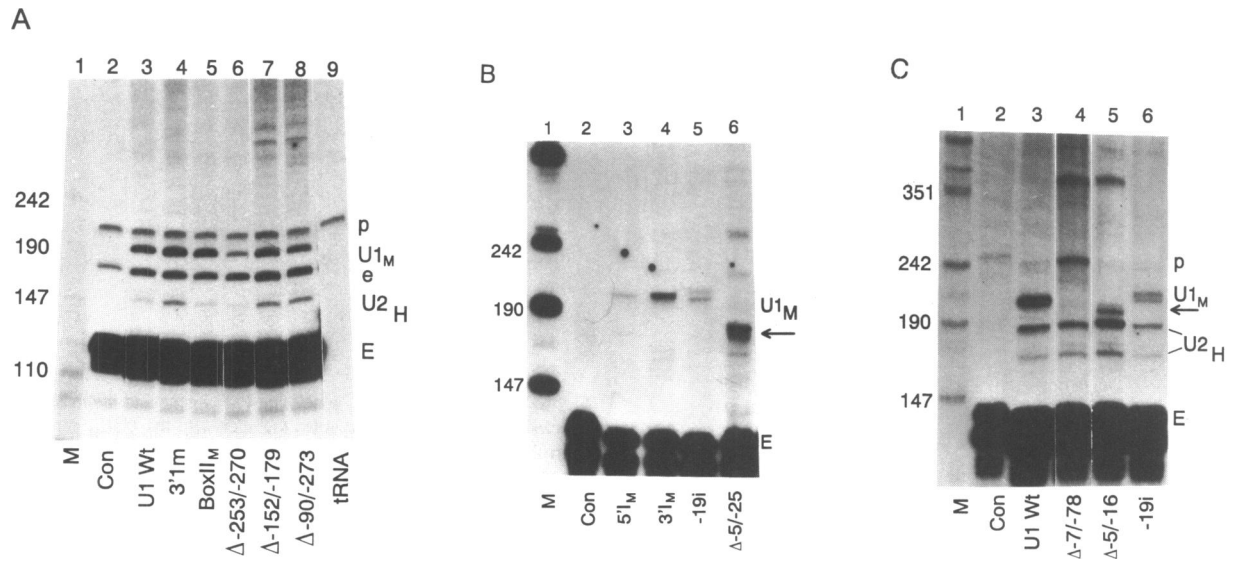

Fig. 4. Expression of promoter mutants. A. The mutant genes diagramed in Fig. 3A were injected in *L. pictus* zygotes together with the $U2E_H$ gene as an internal control. The RNA was analyzed using the mixed riboprobe assay described in Fig. 1B. Lane 2 (Con) is analysis of RNA from control uninjected embryos and lane 9 is analysis of yeast tRNA. Lanes 3–8 are analyses of RNA from embryos injected with the genes in Fig. 3 (the gene is indicated beneath the lane). The RNase resistant fragments are: p—the parent $U2_H$ probe ; e—protection by a transcript from endogenous U1 gene in the *L. pictus* embryos; $U1_M$—protection by the $U1_M$ transcript; $U2_H$—protection by the $U2_H$ transcript; E—protection by endogenous U1 and U2 snRNAs. Lane 1 is pUC18 digested with HpaII. B. The mutant genes described in Fig. 3B were injected into *S. purpuratus* zygotes and the RNA was analyzed using the $U1_M$ riboprobe as shown in Fig. 1B. Lane 2 (Con) is analysis of RNA from control uninjected embryos. Lanes 3–6 are analyses of RNA from embryos injected with the genes in Fig. 3B (the gene is indicated beneath the lane). The RNase resistant fragments are: protection by a transcript from endogenous U1 gene in the *L. pictus* embryos; $U1_M$—protection by the $U1_M$ transcript; E—protection by endogenous U1 snRNAs. Lane 1 is pUC18 digested with HpaII. The arrow indicates the transcripts from the Δ−5/−25 gene. C. Mutant genes shown in Fig. 3B were injected into *S. purpuratus* zygotes together with the $U2E_H$ gene as an internal control. The RNA was analyzed using the mixed riboprobe assay described in Fig. 1B. Lane 2 (Con) is analysis of RNA from control uninjected embryos. Lanes 3–6 are analyses of RNA from embryos injected with the genes in Fig. 3B (the gene is indicated beneath the lane). The RNase resistant fragments are: p—the parent $U2_H$ probe; $U1_M$—protection by the $U1_M$ transcript; $U2_H$—protection by the $U2_H$ transcript; E—protection by endogenous U1 and U2 snRNAs. Lane 1 is pUC18 digested with HpaII. In this experiment two protected fragments were found from the control $U2E_H$ gene. The arrow indicates the transcripts from the Δ−5/−16 gene.

These experiments show that the box I sequences are not essential for expression. Moreover, they indicate that the start site of transcription is determined by sequences 5′ of −25.

Identification of the U1 snRNA proximal sequence element (PSE)

Since sequences in the region from −90 to −270 and 3′ of −25 were not critical for expression, a series of scanning mutations spanning the region from −30 to −80 were constructed and analyzed. Each mutation spanned a region of 10 nts. The scanning mutants were injected into *S. purpuratus* embryos together with the $LvU2_H$ gene (Fig. 5) and compared with the expression of the $U1_M$ gene (Fig. 5, lane 2). Altering the nucleotides between −60 and −80 did not have a major effect on expression (Fig. 5, lanes 6 and 7). In contrast, alteration of the 10 nt between −50 and −60 completely abolished expression (Fig. 5, lane 8), while alteration of the sequences between −40 and −50 reduced expression more than 90% (Fig. 5, lane 9). Alteration of the sequences between −30 and −40 reduced expression about 3–6 fold (Fig. 5, lane 10). The 30LS mutant overlaps the 5′ end of box I, mutation of which also resulted in reduced expression (Fig. 4B, lane 3). These results suggest that there is an element critical for expression located between nts 40 and 60 and that the 3′ portion of this element extends close to nt 30. This element is located in a similar position to the vertebrate PSE. The results of the deletion experiments presented above (Fig. 4) are consistent with this element being a major determinant of the transcription start site. Thus the sea urchin U1 snRNA promoter has a structure similar to the vertebrate promoter with a DSE and a PSE element, each of which is required for expression.

The *L. variegatus* U1 PSE has very limited homology to the same region of the *S. purpuratus* U1 promoter (see below, Fig. 8), although the *L. variegatus* U1 gene is expressed well in *S. purpuratus* embryos. A core T-rich sequence centered around −50 is present in both species. Five nucleotides in this core sequence were changed to create the gene PSE_M. This gene was injected into *S. purpuratus* zygotes together with the $LvU2_H$ gene. The expression of the PSE_M gene was reduced 5 fold relative to the wild type gene (cf. Fig. 6, lanes 3 and 4), demonstrating that the region around −50 is important for expression. These results were quantified by densitometry and are summarized in Table I.

Protein factors interact with both the DSE and the PSE

Stevenson *et al.* have previously shown that there is a factor in blastula nuclei but not in gastrula nuclei which binds the box IV region as assayed by DNAse I footprinting (12). Here we use a mobility shift assay to confirm this result and to look for a PSE binding factor. Synthetic oligonucleotides corresponding to the DSE and PSE were end-labeled and incubated in a nuclear extract prepared from blastula embryos. The sequences of the oligonucleotides used are shown below the figure. A non-specific competitor(dI/dC) was included at a constant concentration. Two specific complexes (labeled I and II in Fig. 7A) were observed when the box IV oligonucleotide was incubated with the nuclear extract from blastula embryos (Fig. 7A, lane 3). Both of these

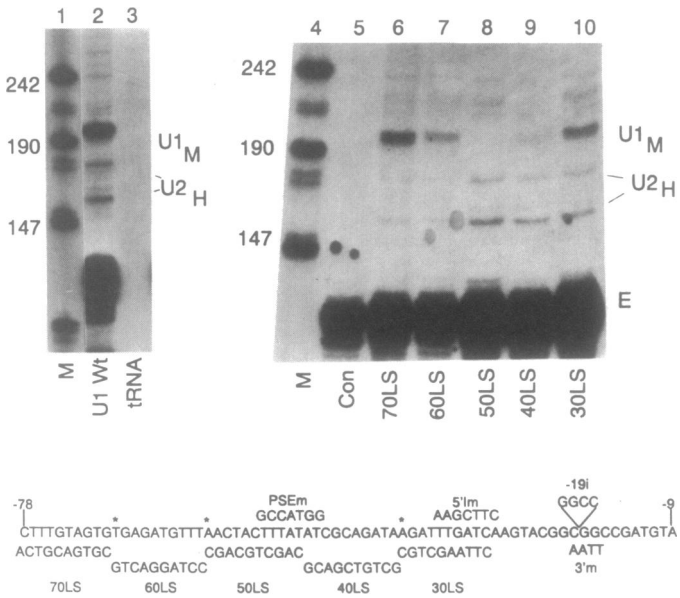

Fig. 5. Linker scanning of the proximal region of the promoter. The LS genes were injected into *S. purpuratus* zygotes together with the $U2E_H$ gene as an internal control. The RNA was analyzed using the mixed riboprobe assay described in Fig. 1B. Lane 2 is analysis of RNA from embryos injected with the intact U1 gene (U1WT), lane 5 (Con) is analysis of RNA from control uninjected embryos and lane 3 is analysis of yeast tRNA. Lanes 6–10 are analyses of RNA from embryos injected with the LS genes shown beneath the figure (the gene is indicated beneath the lane). The RNase resistant fragments are: $U1_M$—protection by the $U1_M$ transcript; $U2_H$—protection by the $U2_H$ transcript; E—protection by endogenous U1 and U2 snRNAs. Lanes 1 and 4 are pUC18 digested with HpaII. There is much less of the other transcript from the $U2E_H$ gene in this analysis compared with Fig. 4C. The sequence of the proximal region of the promoter is shown with the mutations created listed underneath the sequence.

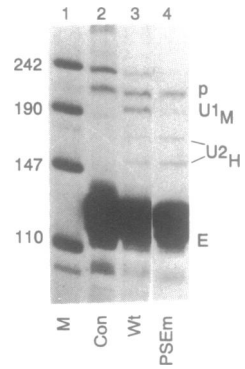

Fig. 6. Analysis of the PSE_M gene. The PSE_M gene and the intact U1 gene were injected into *S. purpuratus* zygotes together with the $U2E_H$ gene as an internal control. The RNA was analyzed using the mixed riboprobe assay described in Fig. 1B. Lane 2 (Con) is analysis of RNA from control uninjected embryos and lane 3 is analysis of RNA from embryos injected with the intact U1 gene (wt). Lane 5 is analysis of RNA from embryos injected with both the PSE_M gene and the $U2E_H$ gene. The RNase resistant fragments are: $U1_M$—protection by the $U1_M$ transcript; $U2_H$—protection by the $U2_H$ transcript; E—protection by endogenous U1 and U2 snRNAs. Lane 1 is pUC18 digested with HpaII. The sequence of the PSE_M gene is shown in the previous figure.

complexes were competed by excess box IV oligonucleotide (Fig. 7A, lanes 4 and 5) and not by the same amount of an unrelated oligonucleotide sequence (Fig. 7A, lanes 6 and 7). Thus there is a specific factor(s) in nuclear extracts from blastula embryos which binds to the DSE sequence.

Incubation of the synthetic PSE sequence with the blastula nuclear extract also resulted in formation of a specific complex. Two complexes (I and II) were formed, both of which were efficiently competed by the homologous unlabeled PSE oligonucleotide (Fig. 7B, lanes 4 and 5). There was partial competition of formation of complex I by a non-specific oligonucleotide (cf. the ratios of complex I and II in lane 3 vs. 6 and 7), while complex II was not competed significantly by a large excess of non-specific oligonucleotide (Fig. 7B, lanes 6 and 7). This result suggested that complex II was a specific complex, while complex I is non-specific.

Further support for this interpretation came from analysis of the ability of the oligonucleotide corresponding to the PSE_M gene to compete for formation of these complexes. The PSE_M oligonucleotide contains the same 5 nucleotide substitutions present in the PSE_M gene, which was expressed poorly relative to the wild-type gene (Fig. 6). The labeled wild-type PSE oligonucleotide was incubated with the blastula nuclear extract and increasing amounts of the authentic PSE oligonucleotide or the PSE_M oligonucleotide were used as competitor. The homologous oligonucleotide competed for the formation of both complexes I and II (Fig. 7C, lanes 2–6). In contrast, the PSE_M oligonucleotide competed for the formation of complex I identically with the wild-type oligonucleotide, but did not compete efficiently for the formation of complex II (Fig. 7C, lanes 7–11). This result further supports the interpretation that complex II is due to binding of a specific factor to the PSE. The failure of the PSE_M oligonucleotide to compete well for the formation of complex II and the reduction in expression when the same substitutions are made into the U1 gene, are consistent with the proposed role of this complex in transcription of the U1 snRNA gene.

Taken together these results demonstrate that there are two important positively acting sequences in the sea urchin U1 snRNA promoter and that each of these sequences interacts with specific protein factor(s), which presumably are critical for the expression of the U1 snRNA gene.

DISCUSSION

Several properties distinguish snRNA genes from other genes transcribed by RNA polymerase II. The transcripts are small, the genes lack introns and the snRNA promoters lack many of the features common to other genes transcribed by RNA polymerase II. While the vertebrate snRNA promoters have been extensively studied, there has been very little work done on defining the invertebrate snRNA promoters. The sea urchin snRNA genes are not expressed in *Xenopus* oocytes, indicating that at least some of the factors are different between the vertebrate and the sea urchin promoters (10). The sea urchin contains at least two sets of genes for the spliceosomal snRNAs. A tandemly repeated set is active in oocytes and early embryos, but is not expressed in pluteus larva or in adult cells (17,18). A low copy number set which is expressed constitutively and is the only gene set expressed in adult cells. The U1 promoter characterized here is from the tandemly repeated gene set and the genes are inactivated between the blastula and the gastrula stage (17).

The sea urchin U1 snRNA promoter

The results reported here demonstrate that the general structure of the sea urchin U1 snRNA promoter is similar to that of the

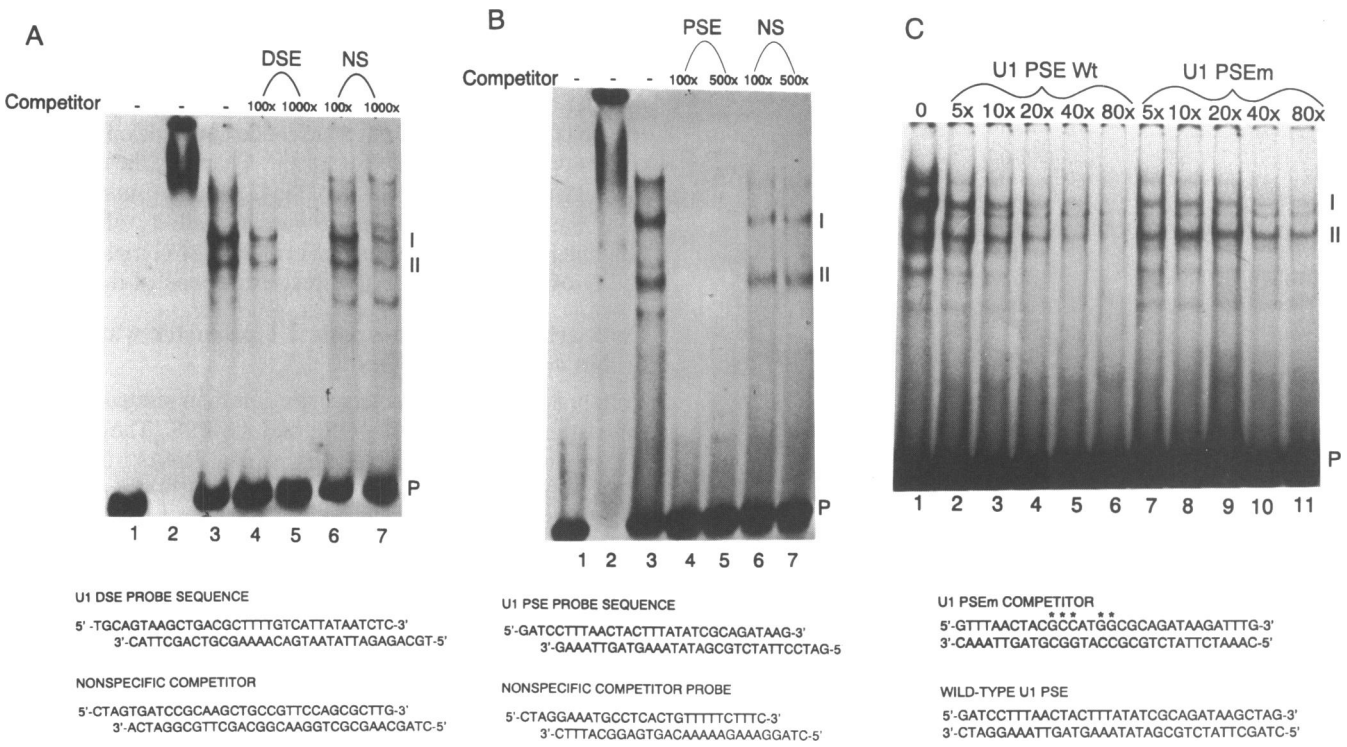

Fig. 7. Mobility shift assay of U1 promoter elements. A. A synthetic double-stranded oligonucleotide containing the box IV sequence was labeled with α-$^{32}PO_4$-dCTP and α-$^{32}PO_4$-dATP (lane 1). The oligonucleotide (2.5 ng) was incubated with 16 μg of blastula nuclear protein and the samples analyzed by electrophoresis on a 5% non-denaturing polyacrylamide gel. Four μg of poly dI/dC were included in all lanes except lane 2. Unlabeled specific (lanes 4 and 5) and non-specific (lanes 6 and 7) competitor oligonucleotides were included at the indicated molar excess. Two specific complexes (I and II) were reproducibly observed. The sequence of the box IV oligonucleotide and the non-specific competitor is shown below the figure. B. A synthetic double-stranded oligonucleotide containing the PSE sequence was labeled with α-$^{32}PO_4$-dCTP and -dATP (lane 1). 2.5 nanogram of oligonucleotide was incubated with 16 μg of nuclear protein from blastula embryos and the samples analyzed by electrophoresis on a 5% non-denaturing polyacrylamide gel. Four μg of poly dI/dC was included in all reactions except lane 2. Unlabeled specific (lanes 4 and 5) and non-specific (lanes 6 and 7) competitor oligonucleotides were included at the indicated molar excess. A specific complex (II) and a non-specific complex (I) were reproducibly observed. The sequence of the PSE oligonucleotide and the non-specific competitor is shown below the figure. C. A synthetic double-stranded oligonucleotide containing the PSE sequence was labeled with α-$^{32}PO_4$-dCTP and -dATP (lane 1). 2.5 nanogram of oligonucleotide was incubated with 16 μg of nuclear protein from blastula embryos and the samples analyzed by electrophoresis on a 8% non-denaturing polyacrylamide gel. Four mg of poly dI/dC was included in all reactions. Unlabeled specific (lanes 2−6) and a competitor oligonucleotide with the PSE_M sequence (lanes 7−11) were included at the indicated molar excess. A specific complex (II) and a non-specific complex (I) were reproducibly observed. The sequence of the PSE oligonucleotide and the PSE_M competitor is shown below the figure.

vertebrate snRNA promoters. There are only two elements required for expression, a DSE which acts in a distance independent manner and a PSE located at about −55 which determines the start site of transcription. Although the overall structure of the sea urchin U1 snRNA promoter and the vertebrate snRNA promoters are similar, the sequences in the sea urchin DSE and PSE are completely different from the vertebrate elements, accounting for the failure of the sea urchin promoters to function in vertebrate systems. Small alterations in the distance between the PSE and the first nucleotide of the U1 RNA altered the start site of transcription, although not in a straightforward manner. Insertion of four nucleotides at −19 resulted in transcription initiating both 5′ and 3′ of the normal start site, with no transcripts which initiated at the normal start site. Larger deletions between the PSE and the start site resulted in heterogenous initiation inside the U1 gene. Thus the precise choice of initiation site characteristic of the normal U1 promoter was lost when deletions or insertions were made. Simply changing the nucleotides around the site of these deletions or insertions, leaving the distance between the PSE and the start site the same, did not affect the start site (Fig. 4). Thus the precise initiation

```
                  ******   *****  **     *    **
LvU1E(-61/-35):   TTTAACTACTTTATATCGCAGATAAGA
SpU1E(-61/-35):   TTTAACCGTTTTATGTCATCCAAGGGA

                   **       *  *   ****  *  *  *  *  *    *
LvU2E(-76/-38):   AAACTCTTTGAAATGCCTCACTGTTTTTCTTTATTAAAA
SpU2E(-76/-38):   TAAAAGCAACATTTTGATCACCGAATCGCCTGGTCCGTA

LvU2L(-76/-38):   AGGGAGTTGTATTTCTCCAACGAGAACTAAAAAGGTTGA

PmU7(-64/-34):    AAACCTAACTRTNNARNTATGCATGTAAAAATT
```

Figure 8. Comparison of the PSEs from different sea urchin snRNA genes. The sequences of the *L. variegatus* U1 (19) and U2E and U2L (11), the *S. purpuratus* U1 (20) and U2E (11) and the *P. milaris* U7 (32) PSEs are aligned, demonstrating that these elements do not share any obvious common features.

of snRNA transcription probably depends upon sequences around the start site as well as the distance between the PSE and the start site.

There are three sequences in the *L. variegatus* U1 snRNA promoter region which have been conserved in the *S. purpuratus* U1 promoter, but which are not essential for expression. The

Table I. Expression of the U1 promoter.

Gene	Expression
−419	100
−318	100
−300	0
−240	0
BoxIIm	120
3'1m	50
5'1m	15
Δ253/270	75
Δ152/179	50
Δ90/273	30
PSE_M	15
70LS	150
60LS	60
50LS	0
40LS	5
30LS	25

The expression of the U1 genes containing different amounts of promoter sequence was determined by injection of the genes into sea urchin zygotes together with a U2 snRNA gene as an internal control. The expression of the gene with the complete promoter (−419) was arbitrarily assigned a value of 100. The results are the average of results obtained from injection into both *L. pictus* and *S. purpuratus* embryos and each represents the average of at least two independent injections. The genes are described in Figs. 3 and 5.

largest of these, box I, extends from −12 to −36, and is 90% conserved. Mutations in the 3' portion of this sequence (−16 to −19) did not affect expression (Fig. 4A, lane 4), while two different mutations in the 5' region (−28 to −36) reduced expression about 3−4 fold (Fig. 5, lane 10; Fig. 4B, lane 3). The 5' region of box I is directly adjacent to the PSE element and mutations directly adjacent to box I abolished expression. Thus it is difficult to assess whether the 5' portion of box I is a separate element from the PSE, although it is very possible given the large size of this region, that multiple factors interact over the PSE and Box I. The other two conserved regions, box II and box III, had less than a 2-fold expression when they were mutated or deleted (lanes 5−7) and both of them could be deleted with only a 3-fold reduction in expression (Fig. 4A, lane 8). These sequences at most only modestly modulate U1 expression of the microinjected genes at the blastula stage, but they could play a role in the shut-off of the genes at the gastrula stage or in the organization of the tandemly repeated gene cluster. Alternatively their conservation could be fortuitous.

Protein factors present in blastula nuclear extracts interact with the two essential elements of the U1 promoter

Using synthetic oligonucleotides containing the DSE and PSE sequences we have shown that there are factors which interact specifically with these sequences present in nuclei from blastula embryos. By DNaseI footprinting, Stevenson and coworkers have previously shown that a factor which interacts with box IV is present in blastula and not gastrula embryos (12), raising the possibility that the activity of that factor is an important part of the regulation of temporal expression of the tandemly repeated U1 snRNA genes.

We also detected a factor which interacts with the PSE element using a mobility shift assay. A complex is formed which is specifically competed by an homologous oligonucleotide, but not by heterologous oligonucleotides. Several other complexes were detected which are non-specific, as judged by the ability of heterologous oligonucleotides to compete for their formation. Supporting the idea that the specific complex (complex II) represents a complex relevant for expression of the U1 gene is the fact that changing 5 nts in the core of the PSE sequence, reduced the ability of the oligonucleotide to compete for formation of the specific complex by 70−80%. Changing these same 5 nts in the gene reduced expression of the U1 gene by more than 80%. Thus there is a correlation between the ability of the mutant oligonucleotide to compete for formation of complex I and the ability of the mutant PSE to direct expression of the U1 gene.

Comparison of the sea urchin U1 promoter with other sea urchin snRNA promoters

Vertebrate snRNA promoters share common sequences required for expression, both in the DSE and the PSE. These are found not only in the genes for the spliceosomal RNAs transcribed by RNA polymerase II but also by the U6 snRNA transcribed by RNA polymerase III (22,23) and in the less abundant snRNAs, U7 (24,25) and U11 (26). Similarly there is a different commonly shared sequence in the expected position for the PSE among the *C. elegans* snRNA genes (9). A role for this sequence in expression of the snRNA genes in these species has not been established. Plant snRNA genes also share common 5' flanking sequences, which are required for expression (27).

The sea urchin snRNA genes may be an exception to this pattern. Four sea urchin snRNA promoters also have been characterized and all have different sequence requirements for expression. While all four characterized sea urchin snRNA promoters, [U1, repeated and low copy number U2 (11) and U7 (28) snRNAs] have in common a PSE located at about −50 which is involved in determining the start site of transcription, these PSE elements do not share any sequence similarity. The tandemly repeated U2 snRNA genes have a PSE and an essential TATA box (11), a feature not found in other animal cell snRNA genes. There is also an upstream element which only has a modest effect on expression (11). The LvU2E gene containing only 75 nts of 5' flanking sequence (the PSE and the TATA box) is expressed at 25% of the intact gene (11). The single copy U2 snRNA gene expressed later in development has a similar promoter structure to the tandemly repeated U2 snRNA genes, but the sequence of the individual elements (other than the TATA box) are completely different (11). Similarly the sea urchin U7 snRNA gene has an essential sequence in the position of the PSE, which differs from both the U2 and U1 PSE sequences (Fig. 8). There is no similarity between any of the elements required for expression in the U1 and U2 snRNA promoters. These results strongly suggest that the sea urchin snRNA genes utilize distinct elements for each snRNA gene.

It is possible that distinct factors are used only for the tandemly repeated snRNA genes and that the low copy number genes will share common factors. However, in *Xenopus* where there are tandemly repeated snRNA genes expressed in early embryogenesis, and in mouse (29,30) and chicken (31), where there are developmentally regulated snRNA genes, the major promoter elements are shared among these genes. The low copy number sea urchin U2 snRNA gene, with a required TATA box, has an unusual promoter structure and is likely to be distinct from the other low copy number snRNA promoters. Thus while the sea urchin snRNA promoters retain the characteristic of an important PSE element at about −50 critical for expression, the factors which interact with this element may well be distinct for different snRNA genes.

ACKNOWLEDGEMENTS

We thank Branko Stefanovic for many helpful discussions. This is publication 1074 from the Florida State University Marine Laboratory. This work was supported by grant GM27789 from the NIH to W.F.M.

REFERENCES

1. Parry, H.D., Scherly, D. and Mattaj, I.W. (1989) Trends Biochem. Sci. 14, 15–19.
2. Roebuck, K.A., Szeto, D.P., Green, K.P., Fan, Q.N. and Stumph, W.E. (1990) Mol. Cell. Biol. 10, 341–352.
3. Ares, M.,Jr., Chung, J.S., Giglio, L. and Weiner, A.M. (1987) Genes Dev. 1, 808–817.
4. Tanaka, M., Grossniklaus, U., Herr, W. and Hernandez, N. (1988) Genes Dev. 2, 1764–1778.
5. Ares, M.,Jr., Mangin, M. and Weiner, A.M. (1985) Mol. Cell. Biol. 5, 1560–1570.
6. Ciliberto, G., Palla, F., Tebb, G., Mattaj, I.W. and Philipson, L. (1987) Nucleic Acids Res. 15, 2403–2416.
7. Mangin, M., Ares, M.,Jr. and Weiner, A.M. (1986) EMBO J. 5, 987–995.
8. Parry, H.D., Tebb, G. and Mattaj, I.W. (1989) Nucleic Acids Res. 17, 3633–3644.
9. Thomas, J., Lea, K., Zucker-Aprison, E. and Blumenthal, T. (1990) Nucleic Acids Res. 18, 2633–2642.
10. Strub, K. and Birnstiel, M.L. (1986) EMBO J. 5, 1675–1682.
11. Stefanovic, B. and Marzluff, W.F. (1992) Mol. Cell. Biol. 12, 650–660.
12. Stevenson, K.A., Yu, J.-C. and Marzluff, W.F. (1992) Nuc. Acids Res. 20, 351–357.
13. Henikoff, S. (1984) Gene 28, 351–359.
14. Kunkel, T.A. (1985) Proc. Natl. Acad. Sci. USA 82, 488–492.
15. McMahon, A.P., Flytzanis, C.N., Hough-Evans, B.R., Katula, K.S., Britten, R.J. and Davidson, E.H. (1985) Dev. Biol. 108, 420–430.
16. Colin, A.M., Catlin, T.L., Kidson, S.H. and Maxson, R. (1988) Proc. Natl. Acad. Sci. USA 85, 507–510.
17. Santiago, C. and Marzluff, W.F. (1989) Proc. Natl. Acad. Sci. USA 86, 2572–2576.
18. Nash, M.A., Sakallah, S., Santiago, C., Yu, J.-C. and Marzluff, W.F. (1989) Dev. Biol. 134, 289–296.
19. Yu, J.C., Nash, M.A., Santiago, C. and Marzluff, W.F. (1986) Nucleic Acids Res. 14, 9977–9988.
20. Yu, J.-C., Wendelburg, B., Sakallah, S. and Marzluff, W.F. (1991) Nucleic Acids Res. 19, 1093–1098.
21. Smith, A.B. (1988) Mol. Biol. Evol. 5, 345–365.
22. Kunkel, G.R. and Pederson, T. (1988) Genes Dev. 2, 196–204.
23. Carbon, P., Murgo, S., Ebel, J.P., Krol, A., Tebb, G. and Mattaj, I.W. (1987) Cell 51, 71–79.
24. Gruber, A., Soldati, D., Burri, M. and Schümperli, D. (1991) Biochim. Biophys. Acta Gene Struct. Expression 1088, 151–154.
25. Phillips, S.C. and Turner, P.C. (1991) Nucleic Acids Res. 19, 1344.
26. Suter-Crazzolara, C. and Keller, W. (1991) Gene Exp. 1, 91–102.
27. Vankan, P., Edoh, D. and Filipowicz, W. (1988) Nucleic Acids Res. 16, 10425–10439.
28. Southgate, C. and Busslinger, M. (1989) EMBO J. 8, 539–549.
29. Lund, E., Kahan, B. and Dahlberg, J.E. (1985) Science 229, 1271–1274.
30. Howard, E.F., Michael, S.K., Dahlberg, J.E. and Lund, E. (1986) Nucleic Acids Res. 14, 9811–9825.
31. Korf, G.M., Botros, I.W. and Stumph, W.E. (1988) Mol. Cell. Biol. 8, 5566–5569.
32. De Lorenzi, M., Rohrer, U. and Birnstiel, M.L. (1986) Proc. Natl. Acad. Sci. USA 83, 3243–3247.