© 1991 Oxford University Press

The U1 snRNA gene repeat from the sea urchin (*Strongylocentrotus purpuratus*): the 70 kilobase tandem repeat ends directly 3′ to a U1 gene

Jin-Chen Yu+, Brian Wendelburg, Sameer Sakallah and William F.Marzluff*
Department of Chemistry and Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA

Received November 20, 1990; Revised and Accepted January 29, 1991 EMBL accession nos X56629, X56630

ABSTRACT

Lambda phage clones containing multiple copies of the 1.1 kb tandemly repeated unit of the sea urchin (*S. purpuratus*) U1 RNA genes were isolated from a gene library. The 1.1kb repeat unit encodes a single copy of the predominant U1 RNA expressed in oocytes and embryos prior to the blastula stage. The tandem repeat unit is about 80 kb in size and is probably present one time per haploid genome as judged by pulsed-field electrophoresis of sperm DNA digested with restriction enzymes which do not cut in the repeat unit. Two of the phage contained DNA flanking the repeat unit as well as several repeat units. The tandem repeat unit ends just 3′ to the U1 coding region. There is only limited homology in the 5′ flanking region with U1 snRNA genes from the sea urchin *L. variegatus*.

INTRODUCTION

The small nuclear RNAs are required components for the processing of pre-mRNAs in eucaryotes (1,2). The snRNAs are among the most abundant RNAs in higher eucaryotes (3) and they are encoded by multiple genes in most species (4). In mammals (5) and birds (6) there are a limited number of functional U1 genes. In organisms such as sea urchins and frogs, which have large eggs and whose early development is characterized by very rapid cell divisions, there is a much larger demand for snRNAs, both during oogenesis and early embryogenesis. In both *Xenopus* (7,8) and sea urchins (9,10) the genes for U1 RNA are present in tandemly repeated units which encode the major U1 RNAs expressed in oogenesis and early embryogenesis. There are other U1 RNAs expressed late in embryonic development and in somatic cells in both frogs (11) and sea urchins (12), indicating that there must be other genes for U1 RNA, which are present in lower copy number. A minor gene family for U1 RNA has been isolated from *Xenopus* (8).

The tandemly repeated sea urchin U1 RNA genes are an example of developmentally regulated genes, which are expressed early in embryonic development and are not expressed after mesenchyme blastula (13) or in somatic cells (12). The sea urchin U1 RNA genes are not expressed upon injection into frog eggs (14,15) so the sequences required for their expression must differ from those of the vertebrate U1 RNA genes. We have isolated the tandemly repeated U1 snRNA gene set from the sea urchin *S. purpuratus* in phage containing multiple copies of the repeat unit. Two of these phage contain the same flanking DNA which is at the 3′ end of the tandemly repeated clusters. This flanking DNA is present in a 80kb EcoR1 fragment containing the entire U1 repeat unit.

MATERIALS AND METHODS

Screening of phage library

A genomic library prepared in EMBL3 from sperm DNA of *S. purpuratus* (16) by partial digestion with Sau3A was a generous gift of Dr. Bill Klein. 150,000 plaques were screened with the plasmid pU1C (10), which contains nucleotides 30 to 140 of the *L. variegatus* U1 RNA gene. The five positive phage were plaque-purified and DNA was prepared from liquid lysates.

DNA sequencing

Appropriate restriction enzyme fragments were cloned into pUC18. DNA was sequenced by the dideoxynucleotide method (17) using Sequenase (U.S. Biochemicals.). The sequence of both strands were obtained from over 90% of the sequence and where only one strand was sequenced the regions were sequenced at least twice.

Pulsed field gel electrophoresis of sea urchin sperm DNA

The general procedure of Finney (18) was followed with some modifications. Sperm was collected, diluted with 1 volume 0.55 M KCl, and embedded in 1% agarose plugs (UltraPure grade,

* To whom correspondence should be addressed

+ Present address: Tumor Virology Section, Building 37, National Institutes of Health, Bethesda, MD 20205, USA

BRL). Each 18 mm plug contained 4 μl diluted sperm and 176 μl molten agarose. The sperm was lysed by incubating overnight at 50°C in lysis buffer [10 mM Tris-HCl, pH 8, 100 mM Na_2EDTA, 1% SDS] containing 100–200 μg/ml proteinase K (BRL). Lysis was continued with fresh buffer and enzyme at 37°C for a second overnight period. Plugs were then washed 3 times, 1 hour each, with 10 volumes of TE buffer [10 mM Tris-HCl, pH 8, 10 mM Na_2EDTA] containing 1 mM PMSF (Sigma), followed by 3 times, 30 minutes each, with 10 volumes of Tris-HCl, pH 8. All washes were performed at room temperature. Several plugs were lysed and washed at the same time, and were stored in TE at 4°C until further use.

For digestion with restriction enzymes, plugs were cut into 4 mm pieces, and the plugs were placed in ice-cold microcentrifuge tubes containing 140 μl of 1× restriction enzyme buffer and 20 μg BSA (Sigma). After 30 minutes on ice, 100 units of restriction enzyme were added, and the tubes were kept on ice for another 30 minutes before they were transferred to a 37°C water bath. Digestion was first carried out overnight, then continued for a second overnight period with fresh buffer and enzyme as described before. At the end of digestion, plugs were placed in fresh tubes and soaked in 1 ml lysis buffer at 50°C. After 8–12 hours, plugs were placed in 1 ml fresh lysis buffer containing 100–200 μg proteinase K, and incubated at 50°C overnight. Plugs were stored in this solution at 4°C until further use.

Electrophoresis of the digested sperm DNA was performed on a GeneLine instrument (Beckman) which uses the Transverse Alternating Field Electrophoresis system (19,20). Electrophoresis was carried out at 14°C for 20 hours in 1× TAFE buffer [10 mM Tris-acetate, pH 7.6, 0.5 mM Na_2EDTA]. Pulse times started at 5 seconds for the first 2 hours, and continued with 5 second increments for each of the 9 remaining 2 hour stages. At the end of electrophoresis, the gel was stained with ethidium bromide for 15 minutes and photographed. The gel was then depurinated with 0.25 N HCl, denatured with 0.5 N NaOH, 1.5 M NaCl, neutralized with 1 M Tris-HCl, pH 8, 1.5 M NaCl, and finally soaked in the transfer buffer [10× SSC]. DNA fragments were then transferred to ZetaBind membrane filters (CUNO) as per manufacturer's instructions. After transfer was complete, the filters were washed at room temperature in 2× SSC for 10 minutes, followed by a 1 hour wash at 65°C in 0.1× SSC, 0.5% SDS, 10 mM Tris-HCl, pH 7.5.

The filters were pre-hybridized for 4–6 hours at 60°C in 5× SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 0.4 mg/ml heparin (Sigma). The probe (U1.1 or U1.2, labeled by the random primer method) was added directly to this solution and hybridization was carried out at 60°C for 18 hours. The filters were then washed twice at 60°C in 2× SSC, 0.1% SDS and exposed to X-ray film.

RESULTS AND DISCUSSION

The sea urchin U1 RNA genes are present in 1.1 (*S. purpuratus*) to 1.4 kb (*L. variegatus*) tandemly repeated units. We isolated the U1 genes from the sea urchin, *S. purpuratus*, by screening a genomic library constructed by partial digestion with Sau3A (16). We obtained five positive phage, four of which have been extensively characterized. Three of these clones contained solely DNA from the U1 repeat unit, while the other two phage contained flanking DNA in addition to several repeat units.

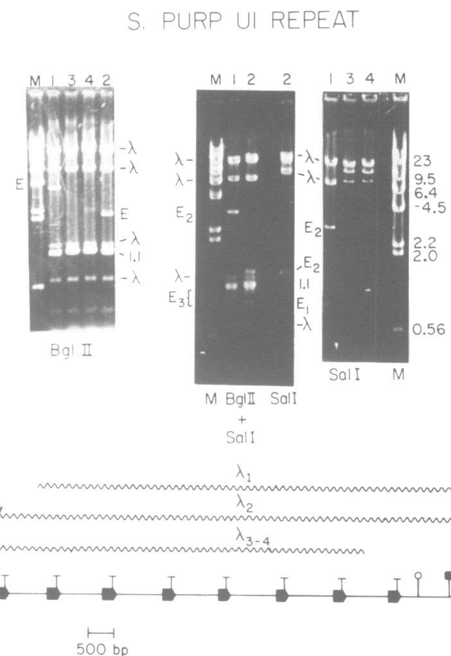

Fig. 1. Characterization of phage containing the U1 repeat unit. Ethidium bromide stained gels of restriction digests of four lambda phage containing the U1 genes of *S. purpuratus* repeat unit are shown. Above each lane is given the number of the phage (1–4). Below the lanes are the restriction enzymes used. Lanes M are marker lambda DNA digested with HindIII. The fragment labeled 1.1 represents the repeat unit. Fragments labeled E are the end fragments. E is the BglII end fragment. E_1 is the BglII-SalI fragment from phage 1 and 2 which was sequenced. The bands labeled E_2 are the SalI flanking fragments from phage 1 and 2. The band labeled E_3 are the flanking fragments from the 5′ portion of the repeat in phage 1 and 2. Bands labeled λ are derived from the phage DNA. A schematic of the repeat unit showing the structure of phage 1–4 is shown below. ┬ BglII; ⫯ BamH1; ┳ SalI.

In Figure 1, we show the ethidium bromide stain of restriction enzyme digests of four lambda phage. If a phage contains only copies of the repeat unit, then digestion with an enzyme which cuts within the repeat unit one time will release multiple copies of the repeat unit.

The library was constructed from a partial digest of genomic DNA with Sau3A. Sau3A cuts the repeat 4 times, once at the BglII site and at 3 additional sites all of which are within 150 nts of the BglII site. Cleavage of the phage DNA with an enzyme, BglII, which cuts once in the repeat unit will release multiple copies of the repeat unit (fragments labeled 1.1 in Fig. 1). There will also be fragments produced from the lambda phage vector as well as fragments which extend from the BglII site in the repeat to the first BglII site in the vector. BglII digests the phage DNA within 100 nts of the polylinker site near the short arm, so these fragments should be about 1.1 kb or less than 250 nts. If there is additional DNA present which is adjacent to the repeat units than these will yield fragments which are longer than 1.1 kb. This was observed in phage 1 and 2 (Fig. 1, left panel, fragments labeled E).

Digestion with BglII released a major fragment of 1.1 kb which is clearly present in multiple copies in each phage. BglII cleaves the parent phage about 100 bp from the SalI site in the short arm. From phage 1 and 2, BglII released longer fragments (4.3 and 2.1kb respectively, designated E in Fig. 1). These fragments must represent pieces of DNA which do not contain the repeat unit and may be end fragments of the tandemly repeated cluster.

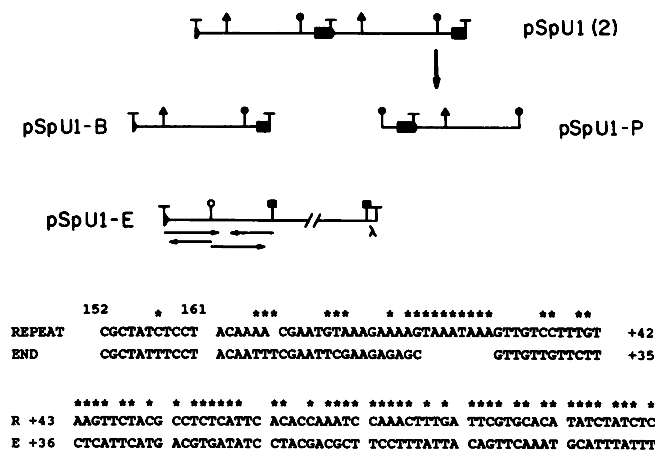

Fig. 2 The *S. purpuratus* U1 repeat unit. (Top). The sequencing strategy used for determining the sequence of the U1 repeat is shown. The plasmids pSpU1(2) and pSpU1-B were constructed by cloning 2.2 and 1.1 kb fragments from a partial digest of phage 3 with BglII into the BamH1 site of pUC18. The plasmid pSpU1-P was constructed by subcloning the PstI fragment from pSpU1(2) into pUC18. The BglII-BglII end fragments were subcloned from phage 1 and 2 and sequenced from the 3′ end of the gene. The subclone pSpU1-E was constructed by inserting the 0.86 kb BglII-SalI fragment from phage 2 into pUC18. A BglII-BamH1 subclone and a BamH1-Sal1 subclone from this plasmid were also sequenced. ↑ AccI; ⚲ BamH1; T BglII; ↑ PstI; ↑ SalI.
(Bottom). Sequence of DNA flanking the U1 repeat. The sequence of the BglII to SalI fragment flanking the U1 repeat is shown. The numbers refer to the distance from the end of the U1 RNA gene. The end fragment (E) is compared with the repeat (R) for about 100 nucleotides starting with the last 10 nucleotides of the U1 coding sequence.

Digestion of the phage DNA with SalI, which cuts in the polylinker, will release the insert sea urchin DNA if only the repeat units have been cloned, since there is no SalI site in the repeat unit. If there is additional sea urchin DNA beside the repeat unit and this DNA contains a SalI site then an additional DNA fragment will be obtained. This DNA fragment represents DNA which flanks the repeat unit in genomic DNA. This was observed in phage 1 and 2 (Fig. 1, center and right panel, phage 1 and 2, fragments labeled E_2). Phage 3 and 4, which contained only repeat DNA, released a single fragment of about 15 kb upon digestion with SalI. Phage 1 and 2 had a SalI site in the insert, resulting in two fragments of sea urchin DNA. Phage 1 contained fragments of 9 and 3 kb, while phage 2 contained fragments of about 15 and 1.3 kb. The shorter SalI fragments are derived from flanking DNA and are designated E_2 in Fig. 1.

Finally digestion with both SalI and BglII will release the repeat units plus any end DNA fragments. These fragments must be less than 1.1 kb if only the repeat units are cloned and indeed must be either about 1 kb or less than 180 nts given the location of Sau3a sites in the repeat unit, relative to the BglII site. Any other fragments, particularly larger fragments, must come from the flanking DNA. Digestion of phage 1 and 2 with SalI and BglII resulted in the SalI-SalI fragment (labeled E_2) described above and an additional BglII-SalI fragment of 0.9kb (labeled E_1 in Fig. 1 center panel). The 0.9kb fragments from phage 1 and 2 are identical (see below) and result from cleaving fragment E (Fig. 1, left panel) with SalI. In addition to fragment E_1, the expected fragments from the other end of the phage (labeled E_3 in Fig. 1, center panel) were obtained. Fragments E_3 contain a partial repeat unit.

Based on these restriction enzyme digests, a model for the repeat unit is presented at the bottom of Figure 1. Phage 1 and 2 contain 8 and about 11–13 copies of the repeat followed by flanking DNA from the 3′ end of the repeat. Digestion of phage 3 and 4 with BglII and SalI gave no fragments of sea urchin DNA larger than 1.1kb consistent with these phage containing only multiple copies of the tandem repeat (not shown). A partial digest of phage 3 and 4 with BglII gave a ladder of fragments which were multiples of 1.1 kb (not shown).

Sequences of the S. purpuratus U1 repeat

Both a 1.1 kb BglII monomer, pSpU1-B, and 2.2 kb BglII dimer, pSpU1[2] of the repeat derived from a partial digest of phage 3 were subcloned into the BamH1 site of pUC18. Since the sea urchin U1 coding region contains a BglII site at nucleotide 140, only the dimer 2.2kb clone contains an intact gene. A 1.1 kb repeat unit containing an intact U1 gene (pSpU1-P) was constructed by digesting the dimer pSpU1[2] subclone with PstI which cuts once in the repeat unit outside of the gene. The sequence of 800 nts of the Pst1 subclone (pSpU1-P) was determined by dideoxy sequencing using the strategy shown in Fig. 2. The sequence of 900 nucleotides of the BglII subclone (pSpU1-B) was determined. From these two sequences, the complete sequence of the 1.1 kb repeat unit was deduced. In addition the 0.9kb BglII-SalI fragments were cloned from both phage 3 and 4, as was the 4.0kb BglII fragment from phage 1 and the 2.3 kb BglII fragment from phage 2. The restriction map of the two 0.9 kb BglII-SalI fragments were identical, as was the sequence for 380 nucleotides starting at the BglII site, indicating that these two fragments were identical.

The 1.1 kb tandem repeat unit contained a single copy of the U1 gene. The coding region differs in two nucleotides (nucleotides 122 and 123) from the sequence of the tandemly repeated *L. variegatus* U1 genes (21), as we previously reported from studies based on RNA sequencing of U1 RNA from *S. purpuratus* eggs (12).

The tandem repeat ends immediately 3′ to a U1 gene

The 4.0 and 2.1 BglII fragments from phage 1 and 2 were subcloned into pUC18. The 0.86 kb BglII to SalI fragment was subcloned from each phage and sequenced. The sequences of the two fragments were identical. At the BglII site there was the 3′ end of a U1 gene starting at nucleotide 140 (Fig. 2). There was one change in the coding region sequence a T for a C at nucleotide 157. There was some similarity between the flanking DNA and the normal repeat DNA for 35 nucleotides from the 3′ end, although much of the purine-rich sequence required for 3′ end formation has been deleted or altered. The sequence of the first 95 nucleotides of the end fragment is compared with the repeat unit in Figure 2. There were no significant similarities to the sequence of the U1 gene or the rest of the 1.1 kb repeat present in this 860 nt 3′ flanking sequence. The complete sequence of the repeat unit and of the end fragment has been deposited in the data base (Accession numbers X56629 and X56630, respectively).

These results strongly suggest that we have cloned the DNA 3′ to the *S. purpuratus* U1 repeat unit. The repeat unit ends directly 3′ to a U1 RNA gene. Since the presumed signal required for 3′ end formation is altered significantly in this gene, it is possible that the last gene in the repeat is not functional. We did not isolate a fragment which contains the other end of the tandem repeat and hence do not know that the boundaries of the repeat

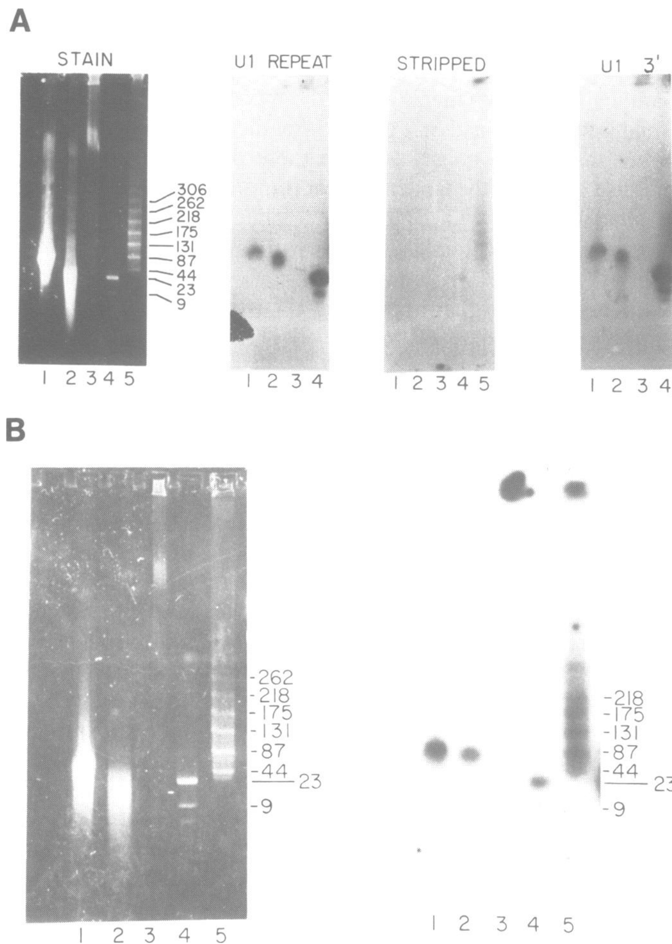

Fig. 3. Analysis of the U1 repeat by pulsed-field electrophoresis. Sperm was immobilized in agarose plugs, digested with restriction enzymes and then analyzed by pulsed-field electrophoresis. A. *S. purpuratus*. On the left is the ethidium bromide stained gel. The filter was hybridized to a labeled plasmid DNA containing the U1 coding region and then exposed to X-ray film for 1 hour (left). Labeled DNA was included so that the markers could be visualized. The filter was then stripped and exposed to X-ray film for 24 hrs (center). The filter was then hybridized to the 3′ flanking 800 nt EcoR1 to SalI fragment and exposed to x-ray film for 24 hours (far right). Lane 1. Sperm DNA digested with BamH1; Lane 2: Sperm DNA digested with EcoR1; Lane 3: undigested sperm DNA; Lane 4: phage DNA digested with HindIII; Lane 5: phage concatamers. The sizes in kb are indicated alongside the gel. B. *L. variegatus* sperm DNA was analyzed by pulsed-field electrophoresis and hybridized with labeled *L. variegatus* U1 plasmid DNA. Radiolabeled phage DNA was included in the hybridization to visualize the markers. On the left is the ethidium bromide stained gel. On the right is the autoradiograph of the filter. Lane 1: *L. variegatus* sperm DNA digested with BamH1; Lane 2. *L. variegatus* DNA digested with EcoR1; Lane 3: undigested sperm DNA; Lane 4: λ phage DNA digested with HindIII; Lane 5: λ phage concatamers. The sizes in kb are indicated beside the gel.

are the same at each end of the repeat. If they are then we would predict that the repeat unit would initiate with sequences just 3′ to the U1 gene and hence the first gene in the tandem repeat would have an intact promoter sequence (containing all the flanking sequences). There is no particular feature of this sequence which suggests a reason for it marking the boundary of the repeat unit. As we show below this fragment is attached to the single large cluster containing the U1 snRNA genes.

In our study of repeated sea urchin small nuclear snRNA genes, we have isolated 37 phage from the sea urchin *L. variegatus* which contain repeat units. Of the 37 phage obtained only one might have contained flanking DNA (22), despite the fact that

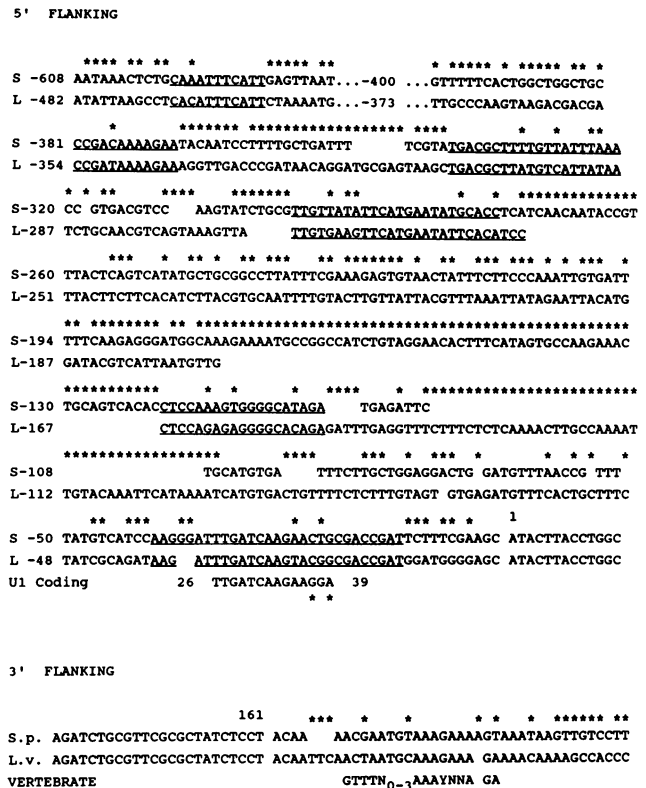

Fig. 4. Comparison of sequences 5′ and 3′ to the U1 gene from *L. variegatus* and *S. purpuratus*. The regions of homology are underlined. The sequence from −19 to −34 is similar to a sequence from nts 26 to 39 in U1 RNA. Asterisks indicate where the two sequences differ. S and L refer to the *S. purpuratus* and *L. variegatus* sequences respectively. The 3′ end sequences are compared with the vertebrate consensus sequence (29). The purine-rich region at the 3′ end of the signal is similar to the sequences 3′ to the sea urchin genes.

the repeat unit is only 70−90 kb (see below). Screening of additional phage from the *S. purpuratus* library containing U1 repeats did not result in isolation of phage containing DNA flanking the 5′ end of the repeat. Some tandemly repeated clusters (e.g. the sea urchin U2 snRNA gene cluster) are not represented in these same gene libraries (our unpublished results) and hence it is likely that certain DNA fragments containing tandemly repeated units are incompatible with cloning in phage.

Size of the U1 gene cluster

To determine the size of the U1 gene cluster and to definitively show that the sequence flanking the U1 gene cluster in the phage clones also flanks the cluster in the genome, we analyzed sea urchin sperm DNA using pulse-field electrophoresis (20). The sperm DNA was digested with two restriction enzymes, EcoR1 and BamH1 which do not cut within the repeat unit. The DNA was resolved by pulse-field electrophoresis and the blot probed with the U1 repeat and with the 0.8kb SalI to fragment 3′ of the repeat unit. In Figure 3A the size of the tandemly repeated unit is estimated from the size of the fragment released by EcoR1 (lane 2), since the BamH1 digest was not complete. The EcoR1 fragment was estimated at 71 kb which represents the U1 gene cluster (Fig. 3A). This fragment contains about 60 copies of the U1 repeat, allowing for the flanking DNA of unknown length which extends to the first EcoR1 or BamH1 site in this region. A fragment of identical size was detected by the 0.8 EcoR1-SalI fragment which was at the end of the cluster in the cloned phage

DNA (Fig. 3A, right). This confirms that this fragment does indeed represent the 3′ flanking sequence to the tandemly repeated U1 snRNA cluster.

Similar results were obtained when the *L. variegatus* sperm was analyzed using the same procedure. Figure 3B shows that both BamH1 and EcoR1 released a 90 kb fragment containing the tandemly repeated U1 snRNA genes. Since the repeat unit containing the *L. variegatus* U1 gene is 1.4 kb, larger than the 1.1kb *S. purpuratus* repeat, it is likely that the copy number is about the same, 60 genes per repeat unit. This number agrees well with previous estimates of the U1 copy number by saturation hybridization (23). These results suggests that there is a single cluster of tandemly repeated U1 genes per haploid genome in each sea urchin species and that the cluster contains about 60 copies of the U1 gene.

Sequence similarities between the U1 repeat units of the two species

In the 400 nucleotides 5′ to the start of the U1 gene, there are four regions which have been highly conserved between the *L. variegatus* and *S. purpuratus* U1 repeat (Fig. 4A). The first of these is just 5′ to the gene (−12 to −35 in both species) and is over 90% conserved (2 changes out of 24 nucleotides) between the two species. Eleven nucleotides (−33 to −23) of this conserved region are identical (in *S. purpuratus*) to the sequence from nucleotides 26 to 36 in the U1 coding region. It is possible that this internal sequence is involved in the expression of the U1 genes. This is a particularly attractive possibility since under some conditions the sea urchin U1 genes (Yu et al., unpublished results) and a modified frog U2 gene (24) can be transcribed by RNA polymerase III. This region includes the area where a TATAA sequence would be expected in most genes transcribed by RNA polymerase II. There is no TATAA sequence conserved between the two genes, suggesting that a TATAA sequence is not important in the function of this promoter.

There are also blocks of sequence similarity further 5′ to the gene. The next 60 nucleotides (−35 to −95) have 60% similarity. Further 5′ there is a 19 nucleotide sequence which is highly conserved (16/19) between the two species. This sequence is located from −148 to −167 in *L. variegatus* and from −101 to −119 in *S. purpuratus*. The 5′ sequences are then completely divergent up to nucleotide −252 where there is an 18 nucleotide sequence which is highly conserved (16/18) with a similar sequence starting at −275 in *S. purpuratus*. The final large conserved sequence block is from nucleotide −321 to −340 in *S. purpuratus* which has 80% (16/20) similarity to a sequence from −288 to −307 in *L. variegatus*.

There are two other short regions of high homology 5′ to the gene. There is a 12 nucleotide sequence at −370 in the *S. purpuratus* gene and at −353 in the *L. variegatus* which show virtual identity (11/12). The final region of similarity is an 11 nucleotide sequence at −473 in the *L. variegatus* gene and at −598 in the *S. purpuratus* gene which are very similar (10/11). It is significant that all of these sequences are included in the highly conserved regions between the two types of *L. variegatus* repeat units (21). The inter and intra-species similarity is consistent with these blocks of conserved sequence having functional importance. In contrast, there is only a limited region 3′ to the gene which is conserved either between or within species (see below).

The vertebrate snRNA genes have a common sequence with the octamer motif, located about 200 nucleotides from the start of transcription in the U1 genes and a conserved proximal element, located about 60 nucleotides from the start of transcription, which are required for expression (25−27). Neither of these sequence elements is present in the sea urchin U1 genes. This is consistent with the failure of sea urchin snRNA genes to be transcribed after injection into *Xenopus* oocytes (14).

Similarities in the 3′ flanking regions

The first 30 nucleotides 3′ to the *S. purpuratus* U1 gene show 83% similarity with the *L. variegatus* sequence. Following the first 30 nucleotides there is no similarity between the two species (Fig. 4B). We had previously noted (10) that the sequence 3′ of the *L. variegatus* U1 gene is a CAAAGAAAGAAA sequence similar to that found 3′ of the sea urchin histone genes (28) and necessary for 3′ end formation of the histone genes (29). The changes in the U1 genes between the two species alter the C-AAAGAAAGAAA sequence so that the 3′ flanking sequence in the *S. purpuratus* gene is not as similar to the sequence at the 3′ end of the histone genes. The sequences required for 3′ formation of the vertebrate snRNAs are also purine-rich and located in a similar position. The sequence requirements for vertebrate snRNA 3′ end formation are not very stringent (30) and a similar situation may exist in sea urchins. The sequences 3′ to the sea urchin U7 genes (31) show moderate similarity to the U1 genes and all of these sequences fit into the vertebrate consensus sequences. Based on the correct functional expression of the sea urchin U7 gene with a *Xenopus* U2 promoter in *Xenopus* oocytes (14), it is likely that the sea urchin 3′ sequences function in vertebrate cells.

Evolution of the tandem repeated units

The two sea urchin species *L. variegatus* and *S. purpuratus* diverged about 100 million years ago. The sequences of the coding region show two nucleotide changes between the two species, the same amount of change as has occurred among mammals (mouse and human) which diverged at about the similar time. Outside of the coding region there are only about 100 nucleotides of sequence which have been highly conserved, 4 short blocks 5′ to the gene and a short sequence 3′ to the gene. This is similar to the sequence conservation found among the mammalian U1 genes (4). One of these sequences immediately 5′ to the gene is similar to a sequence within the 5′ portion of the U1 RNA. It is likely that these sequences are involved in either the expression or correct developmental regulation of these genes. The tandemly repeated gene set is expressed in oocytes and early embryos and then is not expressed in later developmental stages or in somatic cells. In this respect they are similar to the highly repeated histone genes of sea urchins (32) and the 5S rRNA genes (33) of *Xenopus*.

This is first example we are aware of in which the sequences flanking a tandemly repeated sequence have been identified. The U1 tandem repeat ends immediately 3′ to the U1 gene. Indeed, there has been enough variation in the 3′ flanking sequence that it is likely that the terminal U1 gene is not functional. There is no special feature of this sequence which suggests why it might mark the end of the tandem repeat. The relatively high frequency with which the end fragment was obtained (2 out of 5 phage isolated) is consistent with the relatively small size of the tandem repeat. Two independent phage were isolated which contained the same terminal sequence suggesting that this is indeed the end of the repeat. We did not obtain any phage which had the 5′ end of the cluster so it is not yet known whether the two ends of the cluster define an intact repeat unit.

ACKNOWLEDGMENTS

This work was supported by grant GM 27789 to W.F.M. We thank Dr. Bill Klein for a gift of the *S. purpuratus* genomic library. This is publication 1059 from the Florida State University Marine Laboratory. The accession numbers for the sequences in this paper are X56629 for the *S. purpuratus* repeat unit (SPU1E) and X56630 for the 3′ end fragment (SPU1END).

REFERENCES

1. Maniatis, T. and Reed, R. (1987) Nature 325, 673−678.
2. Sharp, P.A. (1987) Science 235, 766−777.
3. Reddy, R. and Busch, H. (1988) in Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles, Birnstiel, M.L. Ed. pp. 1−38, Springer-Verlag, Berlin.
4. Dahlberg, J.E. and Lund, E. (1989) in Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles, Birnstiel, M.L. Ed. pp. 38−70, Springer-Verlag, Berlin.
5. Lund, E. and Dahlberg, J.E. (1984) J. Biol. Chem. 259, 2013−2021.
6. Early, J.M.III, Roebuck, K.A. and Stumph, W.E. (1984) Nucleic Acids Res. 12, 7411−7421.
7. Lund, E., Dahlberg, J.E. and Forbes, D.J. (1984) Mol. Cell. Biol. 4, 2580−2586.
8. Zeller, R., Carri, M., Mattaj, I.W. and DeRobertis, E.M. (1984) EMBO J. 3, 1076−1081.
9. Card, C.O., Morris, G.F., Brown, D.T. and Marzluff, W.F. (1982) Nucleic Acids Res. 10, 7677−7688.
10. Brown, D.T., Morris, G.F., Chodchoy, N., Sprecher, C. and Marzluff, W.F. (1985) Nucleic Acids Res. 13, 537−556.
11. Lund, E. and Dahlberg, J.E. (1987) Genes Dev. 1, 39−46.
12. Nash, M.A., Sakallah, S., Santiago, C., Yu, J.-C. and Marzluff, W.F. (1989) Dev. Biol. 134, 289−296.
13. Santiago, C. and Marzluff, W.F. (1989) Proc. Natl. Acad. Sci. USA 86, 2572−2576.
14. Strub, K. and Birnstiel, M.L. (1986) EMBO J. 5, 1675−1682.
15. Birnstiel, M.L. and Schaufele, F.J. (1988) in Structure and Function of Major and Minor Small Ribonucleoprotein Particles, Birnstiel, M.L. Ed. pp. 155−182, Springer-Verlag, Berlin.
16. Hardin, S.H., Carpenter, C.D., Hardin, P.E., Bruskin, A.M. and Klein, W.H. (1985) J. Mol. Biol. 186, 243−255.
17. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463−5467.
18. Finney, M. (1988) in Current Protocols in Molecular Biology, Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. Eds. John Wiley and Sons, New York,NY.
19. Gardiner, K., Laas, W. and Patterson, D. (1986) Somatic Cell Mol. Genet. 12, 185−195.
20. Schwartz, D. and Cantor, C.R. (1984) Cell 37, 67−75.
21. Yu, J.C., Nash, M.A., Santiago, C. and Marzluff, W.F. (1986) Nucleic Acids Res. 14, 9977−9988.
22. Nash, M.A. and Marzluff, W.F. (1988) Gene 64, 53−63.
23. Nijhawan, P. and Marzluff, W.F. (1979) Biochemistry 8, 1353−1360.
24. Mattaj, I.W., Dathan, N.A., Parry, H.D., Carbon, P. and Krol, A. (1988) Cell 55, 435−442.
25. Murphy, J.T., Skuzeski, J.T., Lund, E., Steinberg, T.H., Burgess, R.R. and Dahlberg, J.E. (1987) J. Biol. Chem. 262, 1795−1803.
26. Ares, M., Jr., Mangin, M. and Weiner, A.M. (1985) Mol. Cell. Biol. 5, 1560−1570.
27. Skuzeski, J.M., Lund, E., Murphy, J.T., Steinberg, T.H., Burgess, R.R. and Dahlberg, J.E. (1984) J. Biol. Chem. 259, 8345−8352.
28. Hentschel, C.C. and Birnstiel, M.L. (1980) Cell 25, 301−313.
29. Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Cell 41, 349−359.
30. Ach, R.A. and Weiner, A.M. (1987) Mol. Cell. Biol. 7, 2070−2079.
31. De Lorenzi, M., Rohrer, U. and Birnstiel, M.L. (1986) Proc. Natl. Acad. Sci. USA 83, 3243−3247.
32. Maxson, R., Mohun, T.J., Cohn, R. and Kedes, L. (1983) Annu. Rev. Genet. 17, 239−277.
33. Brown, D.D. (1982) The Harvey Lectures 76, 27−44.