Structure of the sea urchin U1 RNA repeat

David T. Brown*, Gilbert F. Morris, Nunta Chodchoy, Cindy Sprecher and William F. Marzluff

Department of Chemistry, Florida State University, Tallahassee, FL 32306, USA

Received 10 August 1984; Revised and Accepted 27 November 1984

ABSTRACT

The genes coding for U1 RNA in the sea urchin *L. variegatus* are present in a 1400 base pair tandem repeat. One member of the repeat has been cloned and its sequence determined. The repeat unit contains a single copy of the gene for *L. variegatus* U1 RNA. This gene encodes an RNA which is 75% homologous to mammalian U1 RNA. The *L. variegatus* U1 RNA could assume a secondary structure similar to that proposed for other U1 RNAs. In addition the *L. variegatus* U1 RNA is precipitated by anti-SM and anti-RNP antisera. Analysis of the *L. variegatus* genomic DNA using the cloned U1 gene as a probe reveals a major and a minor type of repeat unit. The two repeated units are the same length but differ in a number of restriction enzyme sites clustered 200-500 bases downstream from the gene. The monomer we have cloned and sequenced is a representative of the minor repeat. A sequence (GATAA) which is -41 to -37 bases 5' to the gene has homology to the putative RNA polymerase II promoter. Fifteen bases 3' of the gene is a sequence (CAAAGAAAGAAAA) which is very similar to the sequence found 3' of the sea urchin histone genes. The two Hha I, Hpa II and Ava I sites in the repeat are all unmethylated in sperm DNA.

INTRODUCTION

A major class of small nuclear RNAs is characterized by the presence of cap structures at the 5' terminus which contain 2,2,7 trimethyl guanosine (1). These RNAs are present in ribonucleoprotein particles which are precipitated by certain classes of antisera from patients with lupus erythematosus (2). These RNAs are presumably present in all eucaryotic cells. They have been described in *Drosophila* (3), *Dictyostelium* (4), dinoflagellates (5), peas (6), chickens (7), and sea urchins (8) as well as in mammalian cells.

The best studied small nuclear RNA is the homologue of mammalian U1 RNA. The sequence of the U1 RNA has been conserved

among a variety of species. More strikingly, all the U1 RNAs sequenced thus far can be drawn in a similar secondary structure which is maintained by compensatory substitutions (3, 9). Considerable evidence suggests that this RNA is involved in splicing of mRNA precursors (10-13).

Genes for U1 RNA have been isolated from a number of species. In Xenopus laevis there is a tandemly repeating dimer of U1 genes of about 500 copies (14). Other genes which are not part of the tandemly repeating unit have also been identified. There are a number of gene copies in mammals. In rodents there are U1 genes that are linked but not in a simple repeating unit (15, 16). In humans, there is a family of U1 genes which have long homologous flanking regions but which are not tightly (within 10kb) linked (17, 18). The genes for U2 RNA show a similar organization. Two of the U2 genes in mice are closely linked (19), while in humans the U2 genes are tandemly repeated in a large tandem repeat (20).

We previously reported that the sea urchin U1 genes, like many other sea urchin genes, are encoded in a tandemly repeated unit about 1400 bases long (21). We report here the structure of the repeat unit and the sequence of the RNA which would be encoded by this gene. This RNA is 70-75% homologous to the mammalian and drosophila U1 RNA. The sea urchin U1 RNA is found in a particle which can be precipitated by both anti-RNP and anti-Sm sera from patients with lupus erythematosus.

MATERIALS AND METHODS

Sea Urchins

L. variegatus sea urchins were collected from the Gulf of Mexico and maintained at the FSU marine laboratory at Turkey Point. Embryos were grown and labeled with $^{32}\text{PO}_4$ as previously described (8).

Preparation of RNA

Nuclei were prepared as described by Nijhawan and Marzluff (8). RNA was prepared by extraction with phenol-SDS at pH5 at room temperature. The 4-8S RNA was prepared by sucrose gradient centrifugation and resolved by electrophoresis on 10% polyacrylamide gels in 7M urea (22).

To release ribonucleoprotein complexes nuclei were incubated as described by White et al. (23). The extracted material was immunoprecipitated with anti-RNP, a monoclonal antibody (a gift of Dr. S. Hoch) or anti-Sm antisera (a gift from Dr. J. Steitz) by the procedure of Kessler (24). The precipitated RNA was analyzed by polyacrylamide gel electrophoresis (22).

DNA Sequencing

DNA was sequenced by the method of Maxam and Gilbert (25). The DNA was end-labeled either with γ - $^{32}\text{PO}_4$ ATP and polynucleotide kinase or with α - $^{32}\text{PO}_4$ dCTP and the Klenow fragment of DNA polymerase I (26). 85% of the sequences were done on both strands and those where only one strand was sequenced, the sequences were done at least twice. All the restriction enzyme sites used for sequencing were sequenced directly. The presence of the predicted restriction enzyme sites was confirmed by digestion with the appropriate enzymes.

Subclones were constructed from the sea urchin U1 repeat (cloned in the Hind III site of pBR 322) by cutting the U1 insert with Sau 3A. The four fragments generated were inserted into the polylinker of pUC9 which had been cut with Hind III and Bam HI (Hind III - Sau 3A fragments) or Bam HI (Sau 3A-Sau 3A fragments).

Analysis of U1 Gene Organization

DNA was prepared from L. variegatus sperm by the method of Brown et al (27). Five μgms of DNA was digested with appropriate restriction enzymes and separated by electrophoresis on a 1.2% agarose gel. The DNA was transferred to nitrocellulose (28). The filter was hybridized to the subcloned fragments of the sea urchin U1 gene labeled by nick-translation with α - $^{32}\text{PO}_4$ -dCTP (26). Hybridization was for 18 hours in 6xSSC at 68° as previously described (21). The filter was washed extensively and exposed to x-ray film using an intensifying screen.

S1 Nuclease Mapping

S1 nuclease mapping was done by the method of Weaver and Weissman (29) as previously described (30) using DNA labeled at either the 5' or 3' end. S1 nuclease mapping was also done by the same procedure using RNA labeled in vivo with $^{32}\text{PO}_4$ (8). In addition in vivo labeled sea urchin U1 RNA was hybridized with the pLv U1.1 DNA immobilized on nitrocellulose. Some of the RNA-

DNA hybrids were treated with ribonuclease T_1 prior to elution. The radioactive RNA was then analyzed by polyacrylamide gel electrophoresis.

Northern Hybridization

20 μ gms of *L. variegatus* whole cell RNA was separated on a 2% agarose-formaldehyde gel (32). The RNA was transferred to Gene Screen Plus (New England Nuclear) as described by the manufacturer. The filter was prehybridized for 6 hrs at 68°C in 0.9M NaCl, 0.09M $\text{Na}_3\text{citrate}$, 50mM Tris·HCl, pH 7.5, 1% SDS, 0.1% ficoll, 0.1% polyvinylpyrrolidone, and 100 μ gms/ml denatured *E. coli* DNA. The prehybridization solution was discarded. The filter was hybridized with pLvU1.1C (labeled by nick-translation with $^{32}\text{PO}_4$ dCTP) for 24 hrs at 68°. The filter was washed extensively and exposed to x-ray film. The amount of U1 RNA was quantified by densitometry.

Materials

Enzymes were purchased from Bethesda Research Labs.

γ - $^{32}\text{PO}_4$ ATP was purchased from ICN. α - $^{32}\text{PO}_4$ dCTP was purchased from New England Nuclear.

RESULTS

The *L. variegatus* U1 repeat was isolated from sperm DNA digested with Hind III. DNA fragments of 1.1-1.5 kb were purified by gel electrophoresis and inserted into the Hind III site of pBR322. A single colony was selected by hybridization with mouse U1 DNA and the *L. variegatus* U1 cDNA(21). We have adopted the nomenclature U1 RNA, rather than the N1 RNA used previously (8,21), to be consistent with the nomenclature of small nuclear RNAs in other species.

Characterization of the U1 DNA Clone

The pLv U1.1 plasmid contained an insert of 1398 base pairs. The pLv U1.1 plasmid selected only U1 RNA when $^{32}\text{PO}_4$ labeled RNA from blastula embryos was hybridized to the immobilized DNA (Fig. 1). Nuclease treatment of the hybrids with T_1 ribonuclease did not cleave the majority of the U1 RNA suggesting that this DNA coded for U1 RNA.

The sequence of the 1.4kb Hind III fragment was determined. Figure 2 shows the restriction enzyme map of the insert and the

FIG. 1. Selection of $^{32}\text{PO}_4$ -RNA by LvU1.1

Sea urchin embryos were labeled from 8-48 hours with $^{32}\text{PO}_4$ from blastula to gastrula. Nuclear RNA was prepared from the whole embryo and hybridized to pLvU1.1 DNA or pLu305(41, which codes for 5S RNA) immobilized on nitrocellulose. Duplicate filters were treated with T_1 ribonuclease prior to elution of the RNA. The bound RNA was analyzed by polyacrylamide gel electrophoresis. Lane 1 and 2: RNA bound by pLv U1.1 analyzed with (lane 2) or without (lane 1) T_1 treatment. Lane 3 and 4: RNA bound by pLu 305 analyzed with (lane 4) or without (lane 3) T_1 treatment.

FIG. 2. Restriction Map of pLvU1.1

The restriction map of pLv U1.1 is shown together with the sequencing strategy used. Solid lines are fragments sequenced by 5' labeling and the dotted line by 3' labeling. The position of the U1 gene is indicated by the arrow. The restriction enzymes used were: $\circ$ Hind III; $\circ$ Hinc II; $\circ$ Bcl I(Sau 3A); $\circ$ Hpa II; $\circ$ Hinf I; $\circ$ Ava I; $\circ$ Bgl II(Sau 3A); $\circ$ Dde I; $\circ$ BstNI.

sequencing strategy used. The coding region of the U1 gene is indicated by the arrow.

Figure 3A compares the sequence of the cloned sea urchin DNA with the U1 RNAs of rat, chicken and drosophila. This 161 base sequence was 70-75% homologous with each RNA. The first base of the sea urchin RNA has been assigned at the AUA sequence common to the 5' end of all U1 RNAs. The 3' end of the sea urchin U1 RNA is assigned from the predicted secondary structure. In Figure 3B the same sequence is shown in a possible secondary structure adapted from Mount and Steitz (3), similar that pro-

A

					30					60
DROSOPHILA			T A	-TAACG	C	CG TT	T G A T			
URCHIN	AT	ACTTACCTGGCGCAGGGGTCGC	--ATTGATCA	AAGAAGGATGCACCCCCAGGGCGAGGCTT						
RAT		--	AGATACCA	C	TG	TTTT				
CHICKEN		--	AGACACCA	G C	TG	TTTT				C
					90					120
DROSOPHILA	G	CT	-GAGTTG	C	ATT	T T	AAT	T		
URCHIN	-GCCATTGCACTCCGGCT	-TGCTGAACCT	-TGC	GATTCCCCCAACG	TGGGGA	AACTCGGGCGT				
RAT	AT		A G	C CC	T	T C	A			ACTGC
CHICKEN	AT	CC	G G	C CC	T	T C	A			ACTGC
					150					
DROSOPHILA	GT	T	C G	A- G	CG	- G				
URCHIN	ACAATTTATGGTAGCGGAGATCT	-GCGTTGCGCTATCTCCT-								
RAT	T	G	T G	GA -	C C	G				
CHICKEN	T	G	T G	GA -	C C	G

C

```

                                30                                60
AAGCTTTCGG GTCTACTACC AACTAGAGCC CATGGAACGA GATATTAAG AGAGCAAAGT
                                90                                120
ATAATTTTGC CCAAGTAAGA CGACGACCGA TAAAGAAAG GTTGACCCGA TAACAGGATG
                                150                                180
CGAGTAAGCT GACGCTTATG TCATTATAAT CTGCAACGTC AGTAAAGTTA TTGTGAAGTT
                                210                                240
CATGAATATT CACATCCTTA CTTCTTCACA TCTTACGTGC AATTTTGTAC TTGTTATTAC
                                270                                300
GTTTAAATTA TAGAATTACA TGGATACGTC ATTAATGTTG TCCCAGAGAG GGGCACAGAG
                                330                                360
ATTTGAGGTT TCTTCTCTC AAAACTTGCC AAAATTGTAC AAATTCATAA AATCATGTGA
                                390                                420
CTGTTTTCTC TTGTAGTGT GAGATGTTT ACTGCTTTCT ATCGCAGATA AGATTTGATC
                                450                                480
                                *
AAGTACGGCG ACCGATGGAT GGGGAGCATA CTTACCTGGC GCAGGGGTG CATTGATCAA
                                510                                540
GAAGGATGCA CCCCCAGGGC GAGGCTTGCC ATTGCACTCC GGCTTGCTGA ACCTTGGGAT
                                570                                600
TCCCCAAAC GTGGGGAAC CGGGCGTACA ATTTATGGTA GCGGAGATCT GCGTTGCGGC
                                630                                660
                                *
TATCTCCTAC AATTCAACTA ATGCAAGAA AGAAAATAAA AGCTCCCCGA AAAACTCTGA
                                690                                720
TGGTGACATG AATGACTAA TCCCTTCCAA TTTTGGGGA ATTTCAACCT TATCAGGTTT
                                750                                780
TTCAAAACCAC TAGTTTCTTT ATTACTTAAG GGGACATGAT TGGCATTCCCT TTCATAAAAA
                                810                                840
GCTCGGGTTT AAATCCGTGT GACAAAAATA GCATACTACT TTAACATCTG CTTCATCTCA
                                870                                900
ACTGCGAAAT GTATGTCGAA ATAACTTTG CTAAGTTTGT AAAGATTCTA ATGAGATTTA
                                930                                960
AAAAAAATGA TTTTCCGGCA GATTGGCTCG CTCATTTTAC TCGGTCGCAA GTTTCCGATA
                                990                                1020
AAAACCCCTT CCAAATAGTT TTCTCATTGA AACACGTCA TCCAAAATAC GTTTCAAGTT
                                1050                                1080
AATCCTTCCA ATCAATACAT ATTGTAACGT GGTATATTTT AAAACCTATT GGCCCATCGT
                                1110                                1140
TGTTTCTTTT CCATATGAAC ATTAGTTAGT GGTTAAAAGT TGCATCATGA TTCGTAACTA
                                1170                                1200
AAAAAAAATA AATGACCCGT TTAAGCGTGC AGCTTGGACC AACCAACATT TTCGTACCA
                                1230                                1260
ATCTATAAAA AGGTAAATA TCCAGGTGCA TGGTCATTTC CCAGGGGTGG AGTCCAGGTT
                                1290                                1320
ATTGACTACT CATTTAATGT GTATGAACTA GTTTCGATTT TGTGTTGCAT TTTCTCTTTA
                                1350                                1380
TCCCTTGATA TGGAAATAAA CAATCATTTT AACTTTATCG AGATATTAAG CCTCACATTT
                                1400
CATTCTAAAA TGAGACGAAA GCTT

```

FIG 3. Sequence of the Sea Urchin U1.1 Gene

A. The sequence of pLvU1.1 homologous to the rat (1), chicken (9) and Drosophila (3) U1 RNAs is shown.

B. The sequence of the sea urchin U1 RNA is drawn in the secondary structure proposed by Mount and Steitz (3). The extensive regions of sequence which are invariant among sea urchin, rat, chicken and Drosophila are boxed.

C. The complete sequence of the 1398 base insert of pLvU1.1. The underlined sequences are sequences 5' and 3' to the gene which may be involved in the expression of the U1 gene (see Discussion). The first and last nucleotides of the U1 RNA are indicated by asterisks.

posed by Branlant et al. (9). Regions of sequence (>5 nucleotides) which are invariant between sea urchin, rat, chicken, *Drosophila* and human U1 RNAs are indicated in boxes. The sequence of the entire repeat unit is presented in Fig. 3C. The sequences 5' and 3' to the gene thought to be involved in formation of the RNA are underlined and will be discussed below. (see Discussion)

Sea Urchin U1 RNA is Homologous to pLv U1.1

The position of the U1 gene shown in Fig. 3A was deduced from the DNA sequence. To confirm that sea urchin U1 RNA has a sequence similar to this gene, S1 nuclease mapping was used. pLv U1.1 was 5' end-labeled at the Bgl II site at nucleotide 140 (Fig. 2) in the predicted gene region with γ - $^{32}\text{PO}_4$ -ATP and polynucleotide kinase. The labeled DNA was hybridized with total RNA from various developmental stages. The S1 resistant DNA was analyzed by gel electrophoresis and is shown in Fig. 4A. A single major DNA fragment 140 nucleotides long was protected. An additional minor cleavage at position 30 in the RNA (yielding a 110 nucleotide protected DNA fragment) was also observed in RNA from embryos. This cleavage may be due to the detection of a U1 RNA with a sequence which is different from the pLv U1.1 clone and is present in the embryo. It could also be due to breakdown products of the RNA. Similar results were obtained when pLv U1.1 was cut with Bcl I, which cuts at nucleotide 30 of the RNA (Fig. 2), and 3' end-labeled to probe the 3' end of U1 RNA. A major fragment of 135 nucleotides was protected (Fig. 4B). Using either probe we obtained a similar result with RNA from eggs, embryos and somatic cells, suggesting that the majority of the sea urchin

FIG. 4. S1 Nuclease Mapping of the Sea Urchin U1 RNA

A. The pLvU1.1 DNA was cut with Bgl II and labeled at the 5' end with γ - $^{32}\text{PO}_4$ -ATP and polynucleotide kinase. The DNA was hybridized with 100 nanograms of total RNA from sea urchin oocytes (lane 1) or eggs (lane 2), and morula (lane 3), blastula (lane 4) and gastrula (lane 5) embryos as well as somatic RNA from tubefeet (lane 6). The S1 resistant DNA was analyzed by polyacrylamide gel electrophoresis. The marker was pBR 322 digested with Hinf I (lane M). On the right is a darker exposure which clearly shows the minor S1 cleavage.

B. The pLv U1.1 DNA prepared from a dam^- strain of *E. Coli* was digested with Bcl I and labeled with α - $^{32}\text{PO}_4$ dCTP and the Klenow fragment of DNA polymerase I. The same RNAs were analyzed as in A.

FIG. 5. Analysis of Embryo U1 RNA Synthesized in vivo

A. 4-8S nuclear RNA from embryos labeled with $^{32}\text{PO}_4$ RNA for 10 hours was hybridized in solution with pLvU1.1 DNA cut with Hind III (lane 2), Hha I (lane 3), or Sau 3A (lane 6). The hybrids were treated with S1 nuclease and the S1 resistant RNA was analyzed by gel electrophoresis. Lanes 1 and 5 are marker cytoplasmic and nuclear RNA respectively. Lane 4 is the RNA subjected to the hybridization and S1 without added DNA. The expected fragment sizes are 141 nucleotides (Hha I) and 110 nucleotides (Sau 3A). The band near U4 RNA (140 bases) in lane 6 resulted from hybridization to DNA which had only been partially cleaved with Sau 3A yielding a 140 base fragment.

U1 RNA is not dramatically different in structure in different tissues.

Sea urchin embryos synthesize a similar U1 RNA to the cloned gene. To demonstrate this, embryos were labeled with $^{32}\text{PO}_4$ and the labeled small nuclear RNA was hybridized in solution to pLv U1.1 and to fragments of pLv U1.1 generated by cleavage inside the U1 coding region with various restriction enzymes. The RNA-DNA hybrids were treated with S1 nuclease. S1 resistant fragments were analyzed by gel electrophoresis. In vivo

FIG. 6. Precipitation of U1 RNA with anti-RNP and anti-SM antisera.

Nuclei were prepared from *S. purpuratus* embryos labeled with $^{32}\text{PO}_4$ for 30 hours from the blastula to the gastrula stage. The nuclei were extracted with 0.15M NaCl and the extract was treated with either a monoclonal anti-RNP antibody (lane 1), anti-RNP antiserum (lane 2), or anti-Sm antiserum (lane 3). The RNA was extracted from the antibody precipitate and analyzed by gel electrophoresis.

labeled sea urchin U1 RNA of the predicted length was protected from S_1 nuclease by hybridization to fragments of pLv U1.1 generated by restriction digests which cut within the coding region of pLv U1.1 (Figure 5).

Presence of the U1 RNA in Ribonucleoprotein Particles.

The sea urchin U1 RNA is present in a ribonucleoprotein particle with properties similar to that of the particles present in mammalian nuclei. Nuclei from embryos labeled with $^{32}\text{PO}_4$ were extracted with 0.15M NaCl and the extract was treated with anti-Sm and anti-RNP antisera. The anti-Sm antiserum precipitated a series of RNAs which presumably correspond to the U series of RNAs found in mammalian cells. The anti-RNP anti-sera precipi-

Table I
Level of U1 RNA During Sea Urchin Development.

	Oocyte	Egg	Morula	Blastula	Gastrula	Somatic
S1	.98(2)	1.0(4)	2.8(3)	3.4(4)	2.9(4)	2.0(3)
Northern		1.0	3.0	3.3	2.5	

The levels of U1 RNA were analyzed by S1 nuclease mapping or Northern blot hybridization. Equal amounts of total RNA (100 nanograms) were assayed from each sample. Oocytes were prepared as described by Giudice et al.(42) and were less than 100 microns in size. Results are expressed relative to the egg = 1. The numbers in parentheses are the numbers of independent determinations. Morula embryos were 100-300 cells, blastula embryos were just after hatching and gastrula were 24 hours after fertilization. The somatic RNA was prepared from tube feet. The number of independent determinations is indicated in parentheses. The deviation in each sample was $\pm 25\%$.

tate only U1 RNA (Fig. 6) from nuclei prepared from blastula embryos labeled with $^{32}\text{PO}_4$. The results in Fig. 6 were obtained with S. purpuratus embryos. Identical results were obtained with L. variegatus embryos.

Change in Amount of U1 RNA During Development

The relative amount of U1 RNA in sea urchin eggs and embryos was determined by S1 nuclease mapping and Northern blot hybridization of egg and embryo RNA. As shown in Table I there were large amounts of U1 RNA in the egg. The amount of U1 RNA increased 3-3.5 fold during development with an increase already detectable in the late morula stage. Since the amount of total RNA is constant as the embryo develops from egg to pluteus, the numbers reflect the fold increase in U1 RNA during development. Similar results were found using the S_1 nuclease assay (Fig. 4A,B) or by detection of U1 RNA after gel electrophoresis and transfer of the RNA to nitrocellulose (Table I). These results are in reasonable agreement with those we reported previously (8). Synthesis of small RNAs is initiated at the 32 cell stage and proceeds at a high rate until the blastula stage. It is between these two stages that most of the small nuclear RNA is synthesized by the sea urchin embryo (8).

Heterogeneity of U1 Repeat:

To confirm that the pLvU1.1 DNA represented a repeat unit present in the L. variegatus genome, the genomic DNA was treated

with enzymes predicted to cut the repeat. We have previously reported that digestion of genomic DNA with Bgl II or Hind III gives the expected 1400 bp fragments (21). Sperm DNA was digested with several restriction enzymes, transferred to nitrocellulose, and hybridized with probes from the coding region and 5' and 3' flanking regions of the U1 gene. The results, shown in Fig. 7, demonstrate that the U1 repeat must be heterogeneous and that at least one other type of repeat unit must exist. The probes used and the fragments expected are shown in Fig. 7C. Digestion of genomic DNA with Hinc II and Hha I gave the expected results; single fragments of 1.4 kb and 1.25 kb respectively. Digestion with Dde I yielded a major fragment of 1.2 kb rather than the expected 1.4 kb, indicating that most of the repeat units have at least 2 Dde I sites. (Fig. 7A) Digestion with Hinf I yielded a major 600 base fragment with the 3' probe, (Fig. 7B, lane 5) which was not predicted from the restriction map of pLvU1.1, in addition to fragments of the predicted sizes. One of the Hinf I sites 3' to the gene is missing in most of the repeat units. Digestion with Hpa II and Ava I, two enzymes which are sensitive to methylation at CG sequences, also gave unexpected results. Both these enzymes yielded a major 1.4 kb fragment, indicating that most of the repeat units are only cut once with these enzymes. The expected fragments from the two Ava I and Hpa II sites predicted by the map of pLvU1.1 are present but are less intense. In a case of Hpa II, the failure to cut the genomic DNA is not due to methylation since the isoschizomer Msp I, which is not sensitive to CG methylation, gave the same result. The results are consistent with two types of U1 repeat in the genome, which differ in a region 3' to the gene. The proposed organization of the two types of repeat unit is shown in Fig. 7C.

DISCUSSION

We report here the sequence of a repeat unit containing a sea urchin U1 gene. Like most other tandemly repeated genes, there is a single copy of the gene in the repeat and the great majority of the DNA is spacer sequence. The repeat unit is 68% AT, while the gene region is only 43% AT, in agreement with the general pattern that spacer regions are more AT rich than gene

regions. The GC rich region of the repeat extends from about 100 bases 5' to the gene and ends directly 3' to the gene. The spacer sequence does not appear to be derived from a simple

C

FIG. 7. Analysis of *L. Variegatus* Genomic DNA.

A. Sperm DNA was digested with Dde I (lanes 1 & 4), Hha I (lane 2 & 5) or Hinc II (lanes 3 & 6) resolved on 1.2% agarose gels, transferred to nitrocellulose, and hybridized with the 5' probe (lanes 1-3) or the 3' probe (lanes 4-6).

B. Sperm DNA was digested with the indicated restriction enzymes, resolved on a 1.2% agarose gel, and hybridized to 3 probes from the U1 repeat. The probes were: a 110 base Sau 3A-Sau 3A fragment from the U1 gene (lanes 1-4), a 870 bp Bgl II-Hind III fragment containing 23 bases of coding region and DNA 3' to the gene (lanes 5-8) and a 420 base Hind III-Sau 3A fragment 5' to the gene (lanes 9-12). The enzymes used were Hinf I (lanes 1, 5, 9), Hpa II (lane 2, 6, 10), Msp II (lanes 3, 7, 11) and Ava I (lanes 4, 8, and 12).

C. The structures of the two types of U1 repeat deduced from the mapping experiments in A and B are shown. The three probes used are indicated. A: 5' flanking probe; B: coding region probe; C: 3' flanking probe. The restriction fragments from pLvU1.1 are shown as solid lines and the restriction fragments unique to the other repeat unit are indicated by the wavy lines. Two repeat units are shown in each case since some of the fragments span the Hind III site which we use to define the unit. The restriction enzyme sites are: $\top$ Hind III; $\square$ Hinc II; $\downarrow$ Sau 3A; $\uparrow$ Hha I; $\circ$ Hpa II; $\downarrow$ Hinf I; $\uparrow$ Ava I; $\uparrow$ Dde I.

repeating sequence and there are no significant directly repeated sequences in the repeat.

U1 RNA is synthesized by RNA polymerase II (7,33,34), but the mammalian U1 genes differ from other genes transcribed by RNA polymerase II. They lack a TATAAA box 25-40 bases upstream from the start of the gene (15-18, 20). In addition the 5' flanking regions are highly conserved among mammalian U1 genes. There are three regions of sequence which are of potential interest with

5' END OF SEA URCHIN U1 GENE

```

      81                      55                      30
      TCTCTTTGTAGTGTGAGATGTTTCACTGCTTCTATCGCAGATAAGATTGA
      *      *      *      *
      TCACTTTGTAGAGTGTGATGTTTCTCT
      55                      81
  
```

3' END OF SEA URCHIN U1 GENE

	1	16
U1 GENE	CAATTCAACTAATGCAAAGAAAGAAATAAA	
HISTONE GENES	CA aatatc	CAA-GAAAGAAT
	1	10

Figure 8. Sequences Flanking the U1 Gene

The sequences 5' to the gene(top) and 3' to the gene(bottom) are shown. Upstream of the gene there is a dyad symmetry from base 55 to 81 and a TATAA sequence which is underlined. The sequence 3' to the gene is compared to the sequence 3' of the sea urchin histone genes. The lower case nucleotides are a consensus 3' sequence for the histone genes derived from the sequences presented by Hentschel and Birnstiel (35).

regard to the expression of the sea urchin U1 gene. These sequences are underlined in Fig. 3C, and are shown in Fig. 8. A sequence 37-41 bases upstream from the gene is an AT rich region which is similar to the TATAA sequence found 5' to most genes transcribed by RNA polymerase II. 55-81 bases from the 5' end of the gene there is a 26 base region of dyad symmetry. This is the only significant dyad symmetry region in the whole repeat. Sequences in this region are required for correct expression of human U1 genes after injection into the *Xenopus* oocyte (34). Whether these sequences are required for expression of sea urchin U1 genes is not known. No sequences were present with significant homology to the region I and region II sequences proposed by Weiner from examination of mammalian U1 genes (20). This is not surprising in view of the long separation of mammals and sea urchin in evolution.

There is a clear-cut region at the 3' end of the U1 gene which may be involved in either termination or processing at the 3' end. The sequence CAAAGAAAGAAAA is found 13 bases 3' of the U1 coding region. A similar sequence is found in the early sea urchin histone genes although it is only 8 bases from the end of the histone mRNA, which also ends in a stem-loop structure (35). Evidence is accumulating that the histone mRNAs are formed by

processing at the 3' end (36, 37) and sea urchin U1 RNA may be formed by a similar mechanism.

Fig. 3A compares the U1 RNA sequences of four widely divergent organisms; rat, sea urchin, chicken and *Drosophila*. There are several large regions (8-15 bases) which are invariant in these species. It is clear that the sea urchin sequence can be fit into the same secondary structure as the other U1 sequences (Fig 3B). The striking feature of this structure is that the loops and single-stranded regions have been highly conserved among all the species while the stem regions vary somewhat in length but primarily in compensatory base changes leading to conservation of overall secondary structure. The first 13 nucleotides of the RNA are conserved in all species. In addition there is a conserved stretch from base 142 to 152 in all 3 species which is found in the 3' stem-loop region. These regions must be important in the U1 RNA function. In particular the single-stranded sequences could play a role in specific RNA-RNA or RNA-protein interactions (3).

Gene Organization

The sea urchin U1 repeat is heterogenous as determined by probing genomic DNA cut with a number of restriction enzymes. The repeat length is uniform but there are at least 4 restriction enzyme sites which vary in a 500 base region 3' to the gene. (Throughout we refer to the 3' regions and 5' regions as arbitrarily defined by the Hind III site. However because of the nature of a tandem repeat the region 3' to one gene is 5' to the next.) Each of these enzymes defines a major and minor repeat. Assuming that there is a major and minor repeat we have isolated a member of the minor repeat. At least 2 other genes in the sea urchin have more than one type of repeat. In the sea urchin *L. Pictus* and *S. Drobrachiensis* there are two types of histone repeat each coding for the α -histone genes (38,39). In *L. variegatus* there are 2 types of 5S RNA gene repeats (40). In these cases, unlike the sea urchin U1 repeat, there are differences in repeat length. Whether or not the two types of U1 repeat unit are interspersed cannot be determined from these data. It is unlikely that they are a result of polymorphism since the same pattern was observed with 5 different *L. variegatus* individuals (not shown). We have

recently isolated a second U1 clone which has the restriction map predicted for the major repeat (Yu and Marzluff, unpublished results).

The Hpa II sites are totally unmethylated in sperm DNA since Hpa II and Msp I digestion gave identical results after digestion of sperm DNA. The Hha I sites are also unmethylated since a single fragment of 1250 bases in length was released from sperm DNA. It is likely that the Ava I sites are also unmethylated although it is possible that one of the Ava I sites is partially methylated while the other is totally unmethylated. Bird and coworkers have shown that the sea urchin genome can be divided into methylated and unmethylated domains (41). The U1 RNA genes, like those of other genes expressed during embryonic development, are probably totally unmethylated.

U1 RNA in Different Stages of Development

The majority of the U1 RNA that we detected appears to be identical in all the sea urchin samples we analyzed; egg, embryos and somatic tissues. Minor cleavages are detected in the S1 assay and may arise from U1 RNA molecules which have a sequence which is different from that of the pLv U1.1 DNA probe. Indeed the stage specific variations in the minor bands would support this possibility (Fig. 4A). However it must be cautioned that the additional cleavages produced are minor and that the S1 assay does not distinguish between a sequence variation or a degradation product of the U1 RNA. In addition sequence variation may exist which cannot be detected by the S1 nuclease assay. We cannot conclude that the gene in pLvU1.1 is actually expressed or at what stage it might be expressed. It is also very possible that there are other dispersed U1 genes present in the genome which are not in a tandemly repeated unit. The level of U1 RNA biosynthesis is regulated during sea urchin development (8) and the transcription of this tandemly repeated gene set may be regulated as well.

ACKNOWLEDGEMENTS

This work was supported by NIH grant GM27789 to W.F.M. We thank Dr. Sallie Hoch for the monoclonal anti-RNP antibody and Dr. Joan Steitz for the anti-RNP and anti-Sm sera. This is publication No. 1017 from the FSU Marine Laboratory.

*Present address: Department of Biochemistry, University of Mississippi Medical Center,
2500 N. State Street, Jackson, MS 39216, USA

REFERENCES

1. Busch, H., Reddy, R., Rothblum, L. and Choi, Y. C. (1982) *Annual Review of Biochemistry* **51**, 617-54.
2. Lerner, M. R. and Steitz, J. A. (1979) *Proc. Nat. Acad. Sci.* **76**, 5495-99.
3. Mount, S. and Steitz, J. A. (1981) *Nuc. Acids Res.* **9** 6351-68.
4. Wise, J. A. and Weiner, A. M. (1981) *J. Biol. Chem.* **256** 956-63.
5. Reddy, R., Spector, D., Henning D., Liu, M. and Busch H. (1983) *J. Biol. Chem.* **258**, 13965-69.
6. Krol, A. and Ebel, J.-P. (1984), *Nuc. Acids Res.* **11** 8583-94.
7. Roop, D. R., Kristo, P., Stumph, E. W., Tsai, M.-J., O'Malley, B. W. (1981) *Cell* **23** 671-80.
8. Nijhawan, P. and Marzluff, W. F. (1979) *Biochemistry* **18** 1353-60.
9. Branlant, C., Krol, A., Ebel, J.-P, Lazar, E., Gallinaro, G. Jacob, M., Sri-Widada, J. and Jeanteur, J.-P. (1980) *Nuc. Acids Res.* **8** 4143-54.
10. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. and Steitz, J. A. (1980) *Nature* **283** 220-224.
11. Rogers, J. and Wall, R. (1980) *Proc. Nat. Acad. Sci.* **77** 1877-1879.
12. Padgett, R. A., Mount, S. M., Steitz, J. A. and Sharp, P.A. (1983) *Cell* **35**, 101-107.
13. Hernandez, N. and Keller, W. (1983) *Cell* **35** 89-99.
14. Lund, E., Dahlberg, J. E. and Forbes, D. (1984) *Mol. Cell. Biol.*, in press.
15. Marzluff, W. F., Brown, D. T., Lobo, S. and Wang, S. S. (1983) *Nuc. Acids Res.* **11** 6255-6270.
16. Watanabe-Nagasu, N., Itoh, Y., Toni, T., Okano, K., Koga, N., Okada, N. and Ohshima, Y. (1983) *Nuc. Acids Res.* **11** 1791-1801.
17. Manser, T. and Gesteland, R. (1982) *Cell* **29** 257-264.
18. Lund, E. and Dahlberg, J. E. (1984) *J. Biol. Chem.* **259** 2031-21.
19. Nojima, H. and Kornberg, R. D. (1983) *J. Biol. Chem.* **285** 1851-55.
20. Van Arsdell, S. W. and Weiner, A. M. (1984) *Mol. Cell Biol.* **4** 492-99.
21. Card, C. O., Morris, G. F., Brown, D. T. and Marzluff, W. F. (1982) *Nuc. Acids Res.* **10** 7677-88.
22. Peattie, D. A. (1979) *Proc. Nat. Acad. Sci.* **76** 1760-64.
23. White, P. J., Billings, P. B. and Hoch, S. O. (1982) *J. Immunology* **128** 2751-56.
24. Kessler, S. W. (1975) *J. Immunology* **115** 1617-24.
25. Maxam, A. M. and Gilbert, W. (1977) *Proc. Nat. Acad. Sci.* **74** 560-64.
26. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular Cloning* Cold Spring Harbor Laboratory.
27. Brown, D., Wensink, P. and Jordan, E. (1971) *Proc. Natl. Acad. Sci.* **68** 3175-79.
28. Southern, E. M. (1975) *J. Mol. Biol.* **98** 503-15.
29. Weaver, R. F. and Weismann, C. (1979) *Nuc. Acids. Res.* **7** 1175-93.
30. Sittman, D. B., Graves, R. A. and Marzluff, W. F. (1983)

- Nuc. Acids Res. 11 6679-96.
31. Marzluff, W. F. and Huang, R. C. C. (1984) in Transcription and Translation, A Practical Approach. Ed. by Hames, B. D. and Higgins, S. J. IRL Press, Washington D.C. p. 89-129.
 32. Rave, N., Cikvenzakov, R. and Boedtker, H. (1979) Nuc. Acids Res. 6 3559-67.
 33. Morris, G. F. and Marzluff, W. F. (1983) Biochemistry 22 645-53.
 34. Murphy, J. T., Burgess, R. R., Dahlberg, J. E. and Lund, E. (1982) Cell 29 265-74.
 35. Hentschel, C. C. and Birnstiel, M. L. (1981) Cell 25 301-13.
 36. Birchmeier, C., Folk, W. and Birnstiel, M. L. (1983) Cell 35 433-440.
 37. Galli, G., Hofstetter, H., Stunnenberg, H. G. and Birnstiel, M. L. (1983) Cell 34 823-828.
 38. Cohn, R. H. and Kedès, L. H. (1979) Cell 18 843-53.
 39. Busslinger, M., Rosconi, S. and Birnstiel, M. L. (1982) EMBO J. 1 27-33.
 40. Lu, A.-L., Blin, N. and Stafford, D. W. (1981) Gene 14 51-62.
 41. Bird, A. P., Taggart, M. H. and Smith, B. A. (1979) Cell 17 889-901.
 42. Giudice, G., Sconzo, G., Bono, A., Albanese, I. (1972) Exp. Cell Res. 72 90-94.