Volume 16 Number 4 1988 Nucleic Acids Research

Identification of a repeated sequence in the genome of the sea urchin which is transcribed by RNA polymerase III and contains the features of a retroposon

Paul E.Nisson, Robert J.Hickey, Mark F.Boshar+ and William R.Crain, Jr*

Cell Biology Group, Worcester Foundation for Experimental Biology, 222 Maple Avenue, Shrewsbury, MA 01545, USA

Received September 28, 1987; Revised and Accepted December 10, 1987

ABSTRACT

A repeated sequence element which is located about 200 nucleotides upstream from the protein-coding portion of the muscle actin gene (probably within a large 5' intron) in the genome of the sea urchin, Strongylocentrotus purpuratus has been characterized, and shown to contain the sequence features which indicate that it has been transposed by means of an RNA intermediate. This retroposon-like sequence, SURF1-1, is a member of a family which is dispersed and repeated about 800 times in the genome, referred to as SURF1 (sea urchin retroposon family 1). In vitro transcription of this sequence by RNA polymerase III defines a 300 nucleotide transcription unit which is bounded by a short direct repeated sequence. The 3' end of this unit contains a simple 21 nucleotide A+T-rich sequence characteristic of retroposons, and a consensus B box portion of an internal RNA polymerase III promotor is located 60 to 80 nucleotides downstream from the two sites of transcription initiation. This sequence also contains a 40 nucleotide region that is related to several tRNA sequences (containing the B box), and a 79 nucleotide sequence which is homologous to a repeated sequence previously shown to be present within the 3' untranslated portions of the Spec1 and Spec2 mRNAs of this species (1). Analysis of transcripts of this sequence family in RNA from several embryonic stages indicates that its expression is highest at 11 hours postfertilization (about 128 cells) and drops as development procedes. Furthermore, most or all, transcription of this sequence family in nuclei isolated from 11 hour embryos is by RNA polymerase III, and is from the same strand which is transcribed in vitro.

INTRODUCTION

SINES, short interspersed repetitive sequence elements, range in length from about 70 to 300 nucleotides and are present in most and probably all eukaryotic genomes (2). In mammalian genomes there are generally a small number of different but highly repetitive SINE families and the analysis of their nucleotide sequence indicates that most or all have been derived from known RNA polymerase III genes (reviewed in 3,4,5). These sequences appear to have been amplified and dispersed in the genome via RNA intermediates (6), and are therefore referred to as retroposons (4). The most widely studied example of this type of sequence is the Alu sequence family in the human genome, which has been shown to derive from 7SL RNA (7,8). Virtually all of

the other examples appear to have derived from tRNAs including the monomer family from Galago (9), the rodent B2 family (9,10,11), the rat identifier (ID) family (9,10,11), the rabbit C family (11), and the artiodactyl C repeat family (4,10). In addition to these mammalian SINES it has been reported that there are highly repeated sequences which are related to tRNAs and are transcribed by RNA polymerase III in the genomes of the tortoise, newt and salmon (12,13). The highly repeated sequence in the salmon genome also contains sequence features which are considered to be diagnostic of retrotransposition (ie. short terminal direct repeats and a simple A+T-rich sequence at the 3' end). Furthermore, cloned members of the repeated sequence family Bm1 from the silk moth, Bombyx mori, also contain regions related to a tRNA gene and some members are flanked by direct terminal repeats and have an oligo(A) sequence at their 3' ends (14). Whether all of the short repeated DNA sequences within these non-mammalian genomes is composed of RNA polymerase III gene-derived retroposons, as is thought to be the case for mammalian genomes (4,5), cannot be answered yet. It is apparent, however, that RNA polymerase III retroposons are generated within the genomes of diverse species, and that this is an important mechanism for amplifying certain sequences during the evolution of the genome.

The overall repetitive sequence composition of the sea urchin genome has been described extensively, and there appears to be several thousand different repetitive sequence families (15,16). This is in contrast to the studied mammalian genomes, which appear to contain a much smaller number of different repetitive sequences. The nucleotide sequence of cloned members of at least 12 different sea urchin short repetitive sequence families has been reported (1,17,18,19,20), but none have been reported to be transcribed by RNA polymerase III, or have features of retroposons. On the other hand, the sea urchin genome does appear to contain transposons which do not utilize RNA intermediates. One long (3 kb) repeated sequence element which has sequence characteristics similar to the foldback transposon in Drosophila has been discovered in the sea urchin genome (21,22). Extensive analysis of several copies of this transposon, TU, has revealed the presence of another element, Tsp, which has been inserted into one of the TU elements (17,23). The Tsp elements lack the most typical features of transposons, however, Cohen et al., (17) concluded that they are a new type of transposable element on the basis of their presence at unrelated locations in the genomes of different individuals. We report here the first evidence of a short interspersed repeated sequence which can be transcribed by RNA polymerase III, and which has the

sequence properties required for RNA-mediated transposition (retroposons) in the sea urchin genome (referred to as SURF1-1). The relatively low number of copies of the SURF1 elements in the S. purpuratus genome (about 800) and the apparent absence of retroposon-like characteristics in several other cloned S. purpuratus repeats indicate that this type of repeated sequence does not dominate the genome of the sea urchin as is the case in mammals and some other vertebrates. Nevertheless, the presence of these elements in the sea urchin genome demonstrates that sequence amplification and dispersion through RNA intermediates is widespread among eukaryotic organisms, and must play a role in the evolution of the sea urchin genome.

MATERIALS AND METHODS

Description of Cloned Sequences

The clone pSpG28 consists of a 7 kb genomic HindIII fragment which contains the entire protein coding sequence of the S. purpuratus muscle actin gene and 5' and 3' adjacent sequence inserted into the HindIII site of pBR322 (24,25). A 1.117 BamHI-BglII fragment from pSpG28, which contains the first 45 nucleotides of actin-coding sequence and 1.077 kb of adjacent 5' sequence, was subcloned into the plasmid vector pUC8. A restriction map of this clone, called pUC8-14, is shown in Figure 1A. This same BamHI-BglII fragment was also cloned into the plasmid vector pGEM1 and is referred to as pGEM1-14. pGem4-7.6 (a gift from David S. Durica) carries a 7.6 kb BamHI-BamHI fragment which contains the first 45 nucleotides of actin coding sequence of the muscle actin gene and about 7.5 kb of adjacent 5' sequence. The plasmid clone pXltmet1 consists of a 0.78 kb DNA fragment carrying a Xenopus laevis methionine tRNA gene cloned into pBR322.

In vitro Transcription using a HeLa Cell S100 Nuclear Extract

S100 extracts were prepared from HeLa cells as described by Weil et al. (26). Transcription reactions were carried out in a final volume of 0.020 ml containing 1 µg of plasmid DNA, 0.5 mM ATP, 0.5 mM GTP, 0.5 mM CTP, 0.025 mM unlabeled UTP, 10 µCi of [α-^{32}P]UTP (Sp.act.$\geq$600 Ci/mmole) (transcriptions without label had 0.5 mM UTP), 20 mM Hepes (pH 7.9), 75 mM KCl, 2.2 mM $MgCl_2$, 0.25 mM dithiothreitol, and 0.010 ml of S100 extract. Samples were incubated at 30° C for 1 hr, and nucleic acids were isolated by phenol/chloroform extraction.

Synthesis of Riboprobes and RNase Protection Assays

Riboprobe was synthesized in a final volume of 0.020 ml containing 1 µg of linearized plasmid DNA, 0.5 mM ATP, 0.5 mM CTP, 0.5 mM GTP, 0.012 mM unlabeled UTP, 50 µCi of [α-^{32}P]UTP (Sp.act.$\geq$600 Ci/mmole), 40 mM Tris-HCl (pH 7.9), 6 mM $MgCl_2$, 10 mM dithiothreitol, 2 mM spermidine, and 10 to 20 units of SP6 or T7 RNA polymerase. Samples were incubated at 37° C or 40° C when the T7 or SP6 polymerase was used, respectively. The template was removed by digestion with DNase RQ1 (1 µ/µl) for 15 min at 37° C. Nucleic acids were isolated by phenol/chloroform extraction and unincorporated nucleotides were removed by centrifugation through a 1 ml bed of Sephadex G-50. After ethanol precipitation the dried pellet was dissolved in hybridization buffer to a final concentration of 5-10 x 10^5 cpm/µl.

Total embryo RNA (5 µg) was hybridized with 5-10 x 10^5 cpm of probe in 0.030 ml of a hybridization buffer containing 80% formamide, 40 mM PIPES (pH 6.7), 0.4 M NaCl, 1 mM EDTA at 45° C overnight. Following hybridization,

0.30 ml of 40 µg/ml RNase A, 2 µg/ml RNase T1 (Sigma), 10 mM Tris-HCl (pH 7.5), 5 mM EDTA and 300 mM NaCl, or 2.5 µg/ml RNase A, 0.125 µg/ml RNase T1, 10 mM Tris-HCl (pH 7.5), 5 mM EDTA and 900 mM NaCl was added and the mixture was incubated at 20° C for 1 hr. The reaction was stopped by the addition of 50 µg proteinase K, and SDS to a concentration of 10% followed by incubation at 37° C for 15 min. Nucleic acids were isolated by phenol/chloroform extraction and after ethanol precipitation the resistant RNA was dissolved in a 90% formamide-dye mixture containing 0.3% bromphenol blue and 0.3% xylene cyanol, electrophoresed on a 5 or 10% polyacrylamide, 8 M urea gel, dried and autoradiographed.

RNA Synthesis in Isolated Nuclei and Northern Blotting

Nuclei were isolated and RNA synthesis allowed to run-on in the presence of [α-^{32}P]-UTP as described by Hickey et al., (27). The newly synthesized, ^{32}P-labelled RNA was purified by phenol extraction at 65°C (28), and the unincorporated nucleotides were removed by chromatography on a Bio-Rad P-60 column. The ^{32}P-labelled RNA was used as a hybridization probe in Northern blotting analysis of the opposite strands of pGEM1-14. Hybridization was in 50% formamide, 50 mM sodium phosphate pH 6.5, 0.8 M NaCl, 1 mM EDTA, 5X Denhardt's solution and 200 µg/ml of denatured salmon sperm DNA at 65°C for 60 hr.

DNA Sequencing

DNA sequencing was performed by the procedure of Maxam and Gilbert (29), using the restriction map of the clone pUC8-14 shown in Figure 1 as the basis for the sequencing strategy.

RESULTS

Identification and Mapping of an RNA Polymerase III Transcription Unit Upstream of the Muscle Actin Gene Coding Sequence

We previously discovered a repeated sequence within the first 0.670 kb upstream of the muscle actin gene coding sequence, which is a member of a sequence family that is transcribed by RNA polymerase III during embryonic development (27). A map of the cloned DNA fragment which contains this repeated sequence is shown in Figure 1A (pUC8-14). The analysis of Hickey et al., (27) indicated that the repeated sequence lies on the right hand DdeI-DdeI fragment and the DdeI-BamHI fragment shown in Figure 1A, and that the portion on the DdeI-DdeI fragment is more repeated in the genome than that on the DdeI-BamHI fragment. To ask whether both strands of this repeated sequence family are transcribed in early embryos, we purified ^{32}P-labelled RNA which was synthesized in nuclei isolated from 11 hour (128 cell) embryos, and then hybridized these transcripts to each strand of the repeated sequence (Figure 2). Regardless of whether transcription was allowed to 'run on' in the absence of α-amanitin or in the presence of 1 µg/ml of α-amanitin, most, if not all, of the newly synthesized RNA hybridized to only one strand of the repeated sequence. Most or all members of this repeated sequence family which are transcribed at this stage, must therefore be transcribed asymmetrically

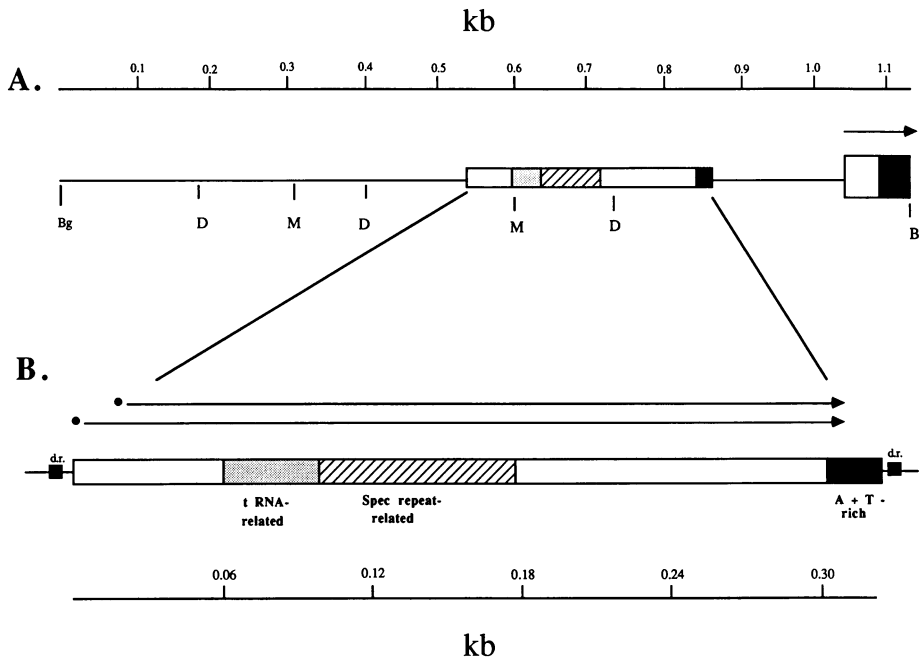

Figure 1. Depiction of the major features of the sea urchin sequence in the recombinant clone pUC8-14. (A) Map of the 1.117 kb sea urchin DNA fragment which is cloned in the plasmid vector pUC8. The larger boxed region on the right shows the location of the muscle actin mRNA sequence. Transcription of the muscle actin gene is from left to right as indicated by the arrow; the open block consists of 5' untranslated mRNA sequence, and the filled block contains actin-coding sequence. The narrow boxed region indicates the location of the SURF1-1 sequence; the depicted domains are indicated in panel B. The following restriction enzyme sites are shown: BglII (Bg), DdeI (D), MspI (M) and BamHI (B). (B) Enlargement of SURF1-1 region showing the regions of tRNA homology, Spec-repeat homology, the simple A+T-rich 21 nucleotide sequence and the short direct repeats (d.r.) found at the ends. The location and direction of transcription of the in vitro transcripts are indicated by the arrows just above the map.

with the same strand being represented in the new transcripts. It should be noted also that the strand of this repeated sequence which is transcribed is the same strand which is copied in the nearby muscle actin gene. The transcription of this sequence family is reduced somewhat by 1 μg/ml of α-amanitin (Figure 2). This was noted previously in a similar experiment in which it was also demonstrated that transcription of actin genes is eliminated by this amount of α-amanitin, and that high concentrations of α-amanitin (360 μg/ml) elimininate transcription of the repeated sequence family (27). The

Figure 2. Asymmetric transcription of SURF1 sequences in nuclei isolated from 11 hour (128 cell) embryos. ^{32}P-labelled RNA synthesized by nuclei from 11 hour embryos in the absence (lanes 1 and 2) or presence of 1 µg/ml (lanes 3 and 4) α-amanitin hybridized to a blot containing RNA transcribed from the opposite strands of the plasmid clone pGEM1-14. Lanes 1 and 3 contain RNA transcribed by SP6 RNA polymerase. In this case two different size RNAs are synthesized, one which is the full length of the sea urchin sequence (1.117 kb), and one which is due to a strong stop site at about 0.680 kb. Lanes 2 and 4 contain RNA transcribed by T7 RNA polymerase which is the full length of the sea urchin sequence. Visualization of these RNAs by UV illumination after electrophoresis demonstrated that approximately equal amounts were present in each lane. The faint signal associated with the T7 polymerase synthesized strand in lane 2 is probably due to a low level of non-specific binding of probe, but it is possible that this strand is transcribed from a few members of this repeated sequence family. In each case the total ^{32}P-labelled RNA synthesized by 9 x 10^8 nuclei was hybridized to the blot: 4.4 x 10^6 cpm for no α-amanitin and 1.5 x 10^6 cpm for 1 µg/ml α-amanitin.

reduced transcription of these sequences by RNA polymerase III in the presence of low concentrations of α-amanitin might be a fundamental property of the promotors associated with these elements. It is also possible that the reduced transcription is due to repeated sequence family members which are embedded within RNA polymerase II transcription units (discussed later).

To ask whether the particular repeated sequence which lies near the muscle actin gene is a functional RNA polymerase III transcription unit, we used DNA containing this sequence as the template in an *in vitro* RNA polymerase III transcription system and looked for the presence of transcripts. DNA from the clone pUC8-14 was transcribed in a HeLa S-100 nuclear extract which has been shown to lack RNA polymerase II activity (26,30). The newly synthesized,

^{32}P-labelled transcripts were separated by electrophoresis on a polyacrylamide gel and then detected by autoradiography (Figure 3A). The major products of this reaction are two transcripts of approximately 300 nucleotides and one about 400 nucleotides. Furthermore, transcription of this DNA and DNA containing a Xenopus laevis tRNA gene (a known RNA polymerase III gene) is resistant to 1 µg/ml of α-amanitin (lanes 4 and 8, respectively) and totally inhibited by 200 µg/ml of α-amanitin (lanes 5 and 9 respectively), demonstrating that the sea urchin sequence has the properties of an RNA polymerase III gene. The partial reduction of the transcription of the sea urchin sequence by 1 µg/ml of α-amanitin suggests that the properties of its promotor are not identical to those of the Xenopus laevis tRNA gene control. Because the tRNA gene is known to contain a particularly strong promotor, and because this repeated sequence family may be diverging in the absence of selective pressure to retain its promotors, it is not surprising that their responses to α-amanitin are slightly different.

The location of the RNA polymerase III transcription unit within the sea urchin sequence was mapped by hybridizing unlabelled in vitro transcripts from this region to RNA probes which were terminated at different positions. Three different ^{32}P-labelled, single-stranded RNA probes were prepared by transcribing DNA from the clone pGEM1-14 with SP6 RNA polymerase after cleaving the template with EcoRI, MspI and DdeI (see Figure 1A; the SP6 promotor is just to the right of the BamHI site shown on this map). When the probe which is complementary to the entire 1.117 kb of sea urchin sequence was used (template cleaved with EcoRI), two major fragments of 310 and 290 nucleotides were protected (Figure 3B, lane 3). The close correspondence in size of these fragments with the two most abundant transcription products seen in Figure 3A indicates that they are the full length transcripts from the sea urchin DNA. The larger fragment in Figure 3A (about 400 nucleotides) is probably due to transcription of the vector DNA sequence. When the probe which contains only the righthand 514 nucleotides was used (template cleaved with MspI), a single major fragment of 250 nucleotides was protected (Figure 3B, lane 4); when the probe containing only the righthand 376 nucleotides was used (template cleaved with DdeI), a single major 105 nucleotide fragment was protected (Figure 3B, lane 1). This analysis indicates that transcription by RNA polymerase III can initiate at two sites, approximately 540 and 560 nucleotides from the left end of the sea urchin DNA in pUC8-14, and that there is a single predominant 3' end for transcripts from this transcription unit, at approximately 850 nucleotides. The location of this RNA polymerase III transcription unit

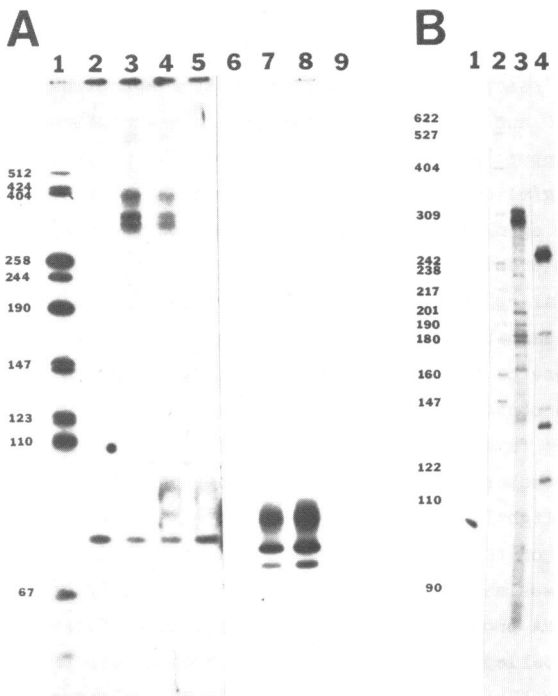

Figure 3. In vitro transcription by RNA polymerase III and mapping of SURF1-1 boundaries using a HeLa cell S-100 nuclear extract. A) α-amanitin sensitivity of transcription of sequence located upstream of muscle actin coding sequence. DNA from the plasmid clone pGEM1-14, containing the repeated sea urchin sequence (lanes 3 through 5) and pXltmet containing a Xenopus laevis methionine tRNA gene (lanes 7 through 9) were used as templates for in vitro transcription using α^{32}P-UTP as a precursor to label newly synthesized RNA. The products of transcription were electrophoresed on a 10% denaturing polyacrylamide gel and autoradiographed. Lane 1, molecular weight markers. Lanes 2,6; no template added to transcription system. Lanes 3,7; transcription with no α-amanitin. Lanes 4,8 and 5,9; transcription with 1 μg/ml and 200 μg/ml α-amanitin, respectively. Exposure for lanes 1-5 was about 20 hr; exposure for lanes 6-9 was 0.5 hr. B) Mapping of the boundaries of the in vitro transcription unit. DNA from the plasmid clone pSpG28 (containing the entire muscle actin coding sequence and 2.125 kb of 5' upstream sequence) was used as template for in vitro transcription, and the synthesized RNA (unlabeled) was hybridized to single stranded ^{32}P-labelled RNA synthesized by SP6 RNA polymerase using pGEM 1-14 DNA as a template. The hybridized RNAs were digested with RNase A (40 μg/ml), RNase T1 (2 μg/ml) in 0.3 M Na^+ and the resistant molecules were electrophoresed on a 10% polyacrylamide gel and autoradiographed. In this experiment the unlabelled RNA transcripts were hybridized to three different length single stranded, ^{32}P-labelled RNA probes complementary to pGEM-114 DNA. The probes begin at the BamHI site in pGem1-14 (see map in Figure 1) and extend to the nearest DdeI site (lane 1), to the nearest MspI site (lane 4) and the end of the sea urchin DNA (lane 3). Lane 2 contains molecular weight markers whose sizes are shown in nucleotides.

(shown in Figure 1) therefore agrees closely with the location of the repeated sequence as determined by Hickey et al., (27). When the full-length probe and the probe which is truncated at the MspI site were used several additional minor bands were evident (Figure 3B, lanes 3 and 4). These bands are probably due to premature termination sites because the size of the four minor bands in lane 4 plus the distance between the transcription start site and the MspI site are roughly the same as the sizes of the four most distinct minor bands in lane 3.

The Sequence of the RNA Polymerase III Transcription Unit Contains the Features of a Retroposon

The complete nucleotide sequence of the sea urchin DNA in the plasmid clone pUC8-14 was determined (Figure 4). Examination of the sequence in the region of the RNA polymerase III transcription unit revealed that this sequence contains the major features associated with repetitive sequence elements which have moved in the genome via RNA intermediates (retroposons). First, the transcription unit is bounded by a short direct repeated sequence, TATATA (nucleotides 529-534, 868-873). Second, a 21 nucleotide simple A + T-rich sequence is located at the 3' end (nucleotides 845-865). Third, a sequence which is closely related to the B box of an internal RNA polymerase III promotor (31) is present (nucleotides 620-630) beginning at 59 and 79 nucleotides downstream from the two sites of initiation. While the transcription unit contains no particularly good match with a consensus A box sequence, which can be part of a split internal RNA polymerase III promotor, it has been shown that the A box sequence is very loosely defined and may not be required for transcription of some Alu repeat sequences (4,5,32,33). Because of the remarkable similarity of features of this sequence with those of other short retroposons, it is likely that this sequence has become repeated in the genome by amplification through RNA intermediates. This is the first evidence of this category of sequence in the sea urchin genome. We will refer to this family of sequences as sea urchin retroposon family 1 (SURF1); the particular cloned member described here will thus be called SURF1-1.

Contiguous Regions of the SURF1-1 Sequence are Related to tRNA and to a Previously Reported Sea Urchin Repeated Sequence

To determine whether any sequence in the clone pUC8-14 is related to other known sea urchin sequences or other known RNA polymerase III genes, we compared the sequence shown in Figure 4 to all sea urchin DNA sequences and most RNA polymerase III gene sequences in the BIONET database. Two interesting relationships emerged from this analysis. First, a 79 nucleotide region (from

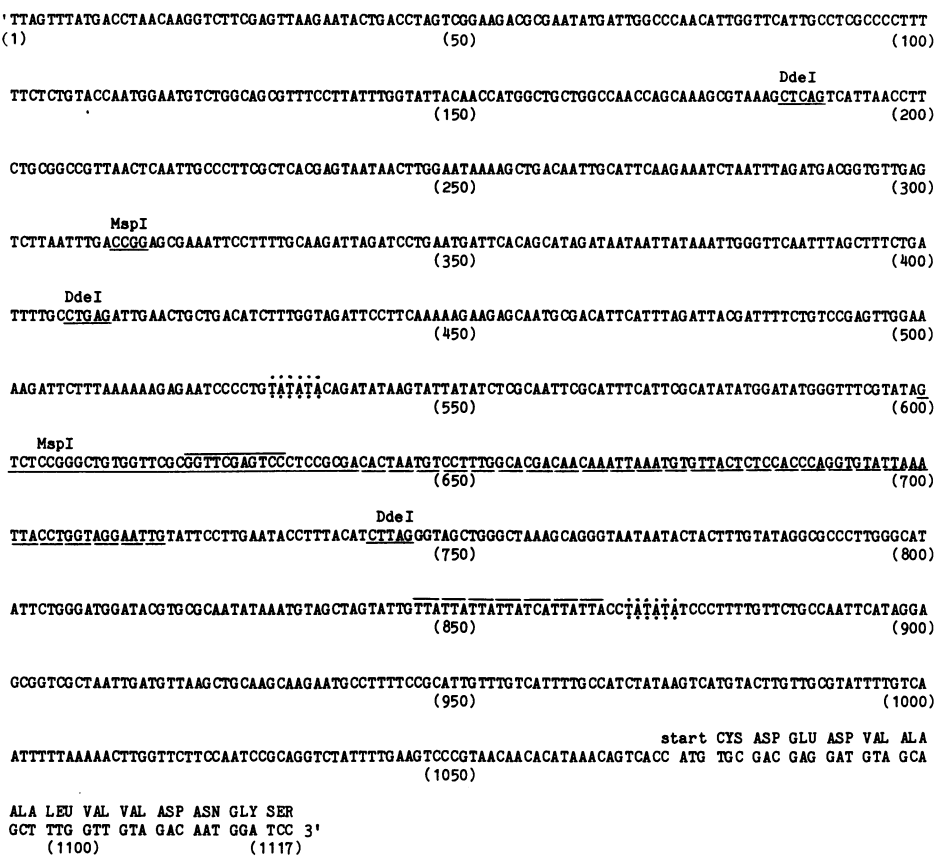

Figure 4. Sequence of the sea urchin DNA in the plasmid clone pUC8-14. The strand shown contains the sequence of the muscle actin gene mRNA and of the RNA polymerase III transcripts of SURF1-1. The last 45 nucleotides of the displayed sequence codes for the first 14 amino acids of the muscle actin; the inferred protein sequence is displayed directly above the nucleotide sequence. Important features of this sequence are indicated as follows: the short direct repeats on either side of the SURF1-1 sequences have dots above and below them; the tRNA homologous region is underlined with a solid line; the sequence with homology to the B box portion of the internal RNA polymerase III promotor is overlined with a solid line; the Spec-repeat homologous region is underlined with a dashed line; the A+T-rich simple sequence is overlined with a dashed line. The MspI and DdeI sites are shown.

nucleotide 639 to 717 in Figure 4) was found within the transcription unit that is 72% homologous with a repeated sequence which is present within the 3' untranslated region of the mRNAs for the Spec1 and Spec2 genes of S. purpuratus (Figures 4 and 5). No meaningful homology was found with the Spec

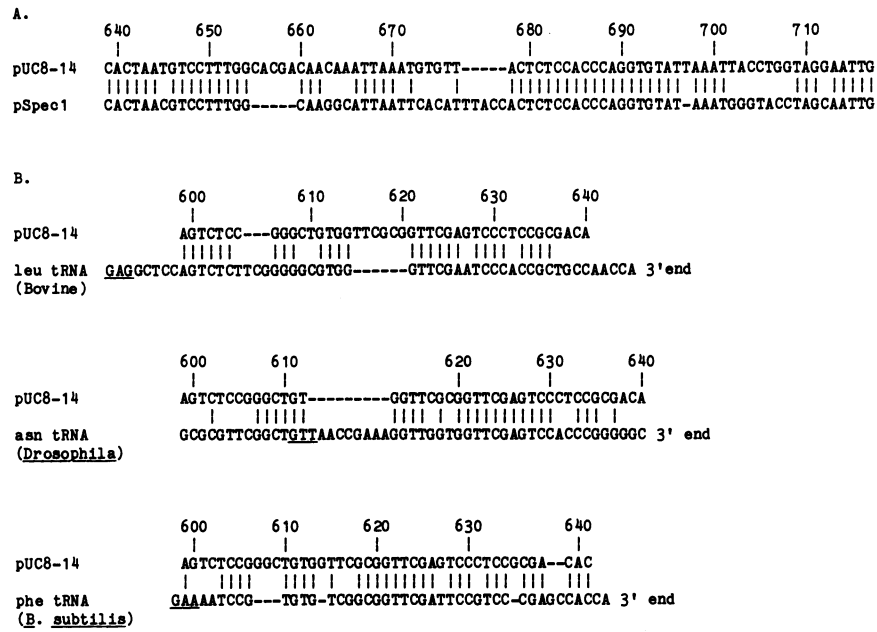

Figure 5. Homology of the SURF1-1 sequence with the S. purpuratus Spec repeat, and with three tRNA sequences. In each case the transcript strand of the sea urchin sequence is shown in the 5' to 3' orientation (left to right) and is numbered as in Figure 4. (A) The related portions of the pUC8-14 and pSpec1 sequences are aligned to produce the maximum match. The strand of the Spec repeat which is shown is the mRNA strand of pSpec1 (1). (B) The homology between the portion of the pUC8-14 sequence from nucleotides 599 to 640 and three different tRNA sequences is displayed. In each case the tRNA anticodon triplet is underlined, and the sequence is shown to the 3' end of the tRNA.

repeat beyond this region, although the repeated sequence in the Spec genes extends in both directions to a total length of 150-200 nucleotides (1).

A second set of homologies was also found when the pUC8-14 sequence was compared to RNA polymerase III genes. In this case a region within the transcription unit which directly abuts the Spec-homologous region, from nucleotides 599 to 640 (Figure 1), showed approximately 70 to 75% homology with several tRNA sequences. Figure 5 depicts the homology with three tRNAs from distantly related organisms. An equivalent amount of homology was also found with other tRNAs, but it could not be determined whether this sequence was even more closely related to sea urchin tRNA sequences because none have been reported. Also, the extent of homology is not sufficiently higher with any particular tRNA to conclude which tRNA this sequence might derive from. In

each case the homology falls mostly between the anticodon and the 3' end of the tRNA sequence, including the B box portion of the internal promotor, and does not extend further in either direction. Because no homology was found to the more 5' portion of any tRNA, it is most likely that if this sequence derives from a progenitor tRNA, that part of it has been lost. Such a loss could be the result of a deletion of this region, or of nucleotide substitutions over a long evolutionary period. If the loss of this region was due to nucleotide substitution, then strong selective pressure must have maintained the more 3' sequence; selection for a functional internal promotor sequence might account for this level of conservation.

Number of Copies of SURF1 in the Sea Urchin Genome

It was previously shown, by genome blotting experiments, that the repeated sequence in the clone pUC8-14 is dispersed throughout the genome and must be present in at least several hundred copies; it also appeared that the portion of this sequence which lies to the left of the right-most DdeI site in Figure 1 is repeated many more times than that lying to the right of this site (27). To determine the number of copies of this sequence in the genome, and to ask whether different parts of it are present a different number of times, we performed a reconstruction hybridization experiment with two different single-stranded RNA hybridization probes spanning different lengths of the cloned fragment. Sea urchin genome DNA and different amounts of DNA from the plasmid clone pUC8-14 were dotted onto nitrocellulose filters. These filters were then hybridized with the two probes: the 376 nucleotide region between the BamH1 site and the nearest DdeI site which contains the 3' 105 nucleotides of the *in vitro* transcription unit (hereafter referred to as the Dde probe), and the 514 nucleotide region between the BamH1 site and the nearest MspI site which contains all but the 5' 45 to 65 nucleotides of the transcription unit, including all of the Spec repeat homology and most of the tRNA homology (referred to as the Msp probe)(see Figure 1). Comparison of the hybridization of these probes to sea urchin and plasmid DNA demonstrates that each is complementary to sequences present multiple times within the genome and that the Dde probe hybridizes substantially less than the Msp probe (Table 1). The approximate number of copies detected in the genome was calculated to be 38 for the Dde probe, and 787 for the Msp probe. Because the Dde probe contains the 3' 105 nucleotides of the *in vitro* transcription unit, there can be only about 38 copies of this unit with the same 3' ends within the genome. The Msp probe contains most of the remainder of the transcription unit (including the tRNA and Spec-repeat homologous regions) and is present nearly 800 times within the

Table 1. Number of copies per genome of sequences present on pUC8-14.

DNA on filter	cpm (x 10^{-3}) hybridized with two different probes[a]	
pUC8-14 DNA[b] (copies/genome)	DdeI	MspI
500	23.5	36.1
2,000	111.9	142.9
8,000	290.9	432.1
sea urchin genomic DNA[c]		
2.50 µg	1.8	56.8
0.25 µg	0.18	6.8
copies/genome[d]	38	787

[a] Two different single-stranded ^{32}P-labelled RNA probes were hybridized to nitrocellulose filters containing a total of 2.5 µg of DNA each. The probes were constructed by transcribing the sea urchin sequence present in the clone pGEM1-14, after cleaving the DNA with either DdeI or MspI. The probes are designated by the restriction enzyme names, and are 0.376 kb (DdeI) and 0.514 kb (MspI). In each case a total of 3.8×10^6 cpm of probe was hybridized to a single filter. Each number is the average of duplicate filters. The increase in hybridized counts when the plasmid DNA is increased 4-fold (from 500 to 2,000 copies) of 4.8- and 4.0-fold for the two probes (average is 4.4), and the decrease in hybridized counts when the genomic DNA is dropped 10-fold (2.50 to 0.25 µg) of 10.0- and 8.4-fold (average is 9.2) indicate that in these ranges the probe sequences are present in sufficient excess to hybridize to all complementary sequences on the filters. It is thus possible to accurately measure the number of copies/genome of these sequences with this data, and to make linear extrapolations over a range of at least 40-fold. The hybridizations of each DNA-containing filter were carried out in 0.030 ml of a buffer containing 50% formamide, 50 mM $NaPO_4$ (pH 6.5), 0.8 M NaCl, 1 mM EDTA, 2.5 x Denhardt's, 250 µg/ml denatured salmon sperm DNA and 500 µg/ml of yeast tRNA at 55°C overnight. The filters were washed at 50°C with 50 mM NaCl, 20 mM $NaPO_4$ (pH 6.5), 1 mM EDTA and 0.1% SDS.

[b] Different amounts of DNA from the clone pUC8-14, which contained amounts of sea urchin sequence corresponding to 500, 2,000 and 8,000 copies per haploid genome, were mixed with mouse genomic DNA to total 2.5 µg and loaded onto nitrocellulose filters. Filters containing 2.5 µg of only mouse DNA were also hybridized with each probe and counted, and these amounts (approximately 100 cpm for each) have been subtracted from the numbers reported here.

[c] The filters with only 0.25 µg of sea urchin genomic DNA also contained 2.25 µg of mouse genomic DNA as carrier.

[d] The number of copies per haploid genome of each fragment was calculated by multiplying the ratio (cpm hybridized to 2.5 µg of sea urchin DNA/cpm hybridized to 500 copies-worth of plasmid DNA) by 500; for DdeI, $1.8 \times 10^3/23.5 \times 10^3$ (500) = 38; for MspI, $56.8 \times 10^3/36.1 \times 10^3$ (500) = 787.

genome. It was reported by Carpenter et al., (1) that there are 2-3,000 copies of the Spec repeat in the sea urchin genome; the high number of genomic sequences detected by the Msp probe is thus probably due to the Spec-repeat-homologous region. The Spec repeat elements studied by Carpenter et al. were 150-200 nucleotides. The Spec-repeat-homologous portion of this sequence spans only 79 nucleotides, and is 72% homologous to the Spec1 sequence; the lower number of copies detected here (787 as compared with 2-3,000) is probably because this particular element does not hybridize with all members of the heterogenous Spec-repeated sequence family under the hybridization conditions used here.

Expression of SURF1 Sequences During Embryonic Development

Examination of the RNA synthesized by isolated nuclei has demonstrated that SURF1 sequences are transcribed throughout embryonic development with a peak of activity at 11 hours (128 cells) after fertilization (27). Attempts to identify transcripts from this region by Northern blotting of total embryo RNA did not reliably detect hybridization bands; we therefore used a more sensitive assay for the presence of transcripts, protection of single-stranded RNA probes from RNase digestion after hybridization. In the experiment shown in Figure 6, total RNA from eggs, 11 hour, blastula, gastrula, and pluteus stage embryos, was hybridized to a ^{32}P-labelled RNA probe complementary to the entire 1.117 kb sea urchin fragment in the clone pGEM1-14. Several protected bands ranging from about 65 to 140 nucleotides are evident, with the highest levels of these bands being detected in RNA from 11 hour embryos. Longer autoradiographic exposure of this gel (and other experiments) showed that each of these bands is present in all stages through pluteus. We have carried out similar RNase protection experiments using a variety of RNase digestion conditions (varying both the RNase concentration and the Na^+ concentration) and have found many protected bands in each case. The conditions used in the experiment presented in Figure 6 (2.5 µg/ml RNase A, 0.125 µg/ml RNase T1, 0.9 M Na^+) yielded the largest protected fragments. The observation that we were unable to find RNase digestion conditions which yielded a single homogeneous protected fragment is consistent with the in vivo transcription of many related, but non-identical, copies of this repeated sequence. Whether the 140 nucleotide fragment, seen most prominently in 11 hour embryo RNA, is a full length transcript of some members of this family cannot be determined from this analysis, but it does set a lower limit on the size of these transcripts. Furthermore, when an RNA probe which includes only sequences between the MspI site at nucleotide 604 and the BamHI site in

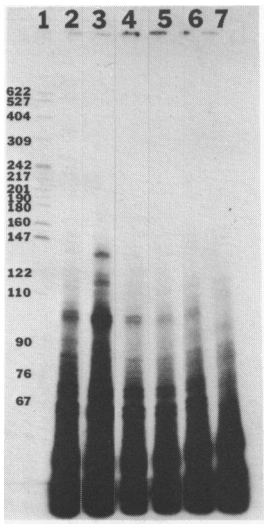

Figure 6. Expression of SURF1 sequences during embryonic development. Total RNA (5µg) from egg (lane 2), 11 hour embryos (lane 3), blastulae (lane 4), gastrulae (lane 5) and plutei (lane 6) was hybridized to ^{32}P-labelled RNA complementarty to the 1.117 kb of sea urchin sequence present in pGEM1-14 (anti-message strand). After digestion with RNase A (2.5 µg/ml) and RNase T1 (0.125 µg/ml) at 0.9 M Na^+, the resistant material was electrophoresed on a 10% polyacrylamide gel and autoradiographed. RNase digested probe, which was hybridized with no RNA, is shown in lane 7. Lane 1 contains molecular weight markers whose sizes are indicated in nucleotides.

pUC8-14 is used, the 140 nucleotide fragment is reduced in size by about 20 to 40 nucleotides (data not shown). The 5' end of this 140 nucleotide *in vivo* transcript thus falls very close to the site(s) of *in vitro* transcription initiation, and the transcript includes not only the entire tRNA related sequence but also virtually all of the Spec-repeat related sequence.

To further identify the boundaries of the *in vivo* transcription of this sequence, an RNase protection assay was performed on RNA from 11 hour and pluteus stage embryos with four different hybridization probes (Figure 7). These probes span either the last 105 nucleotides of the *in vitro* transcription unit (Dde probe), or all but the first 45 to 65 nucleotides of the *in vitro* transcription unit (Msp probe), or the entire transcription unit (Eco and 7.2 probes). The RNase digestion conditions in this experiment were more stringent (40 µg/ml RNase A, 2 µg/ml RNase T1, 0.3 M Na^+) than those used in Figure 6, and resulted in more and smaller repeat-transcript-protected fragments. About 14 protected fragments which derived from SURF1

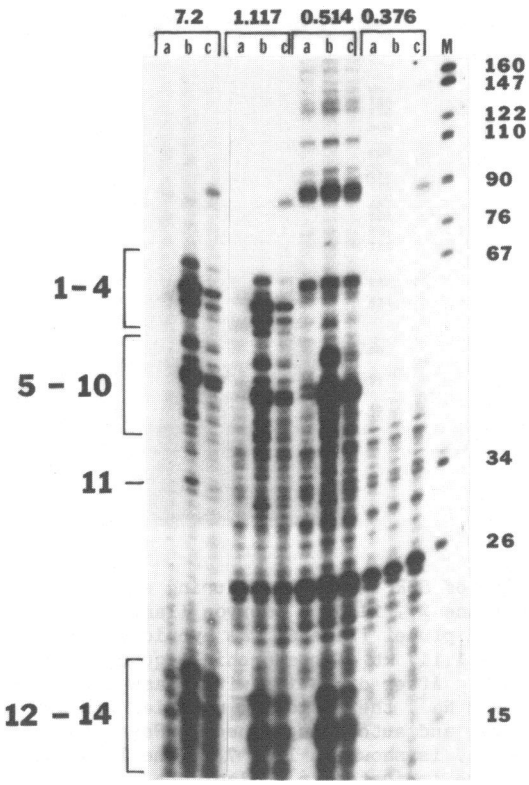

Figure 7. RNase protection mapping of the boundaries of *in vivo* transcription of the SURF1 sequences and the muscle actn gene. Total RNA from 11 hour (b lanes) and pluteus stage (c lanes) embryos was hybridized with ^{32}P-labelled riboprobes which extended various distances upstream from the BamHI site at codons 13 and 14 in the muscle actin gene (see Figures 1 and 4). The probes extended 7.2 kb (riboprobe was synthesized from the clone pGEM4-7.6), 1.117 kb (from pGEM1-14 cleaved with EcoRI), 0.514 kb (from pGEM1-14 cleaved with MspI), and 0.376 kb (from pGEM1-14 cleaved with DdeI). RNase digestion was with RNase A (40 µg/ml) and RNase T1 (2 µg/ml) at 0.3 M Na^+. Lanes marked a contain digested riboprobe which was not hybridized to sea urchin RNA. The fourteen most prominent SURF1-specific fragments are numbered on the left. Lane M contains molecular weight markers with the sizes shown in nucleotides.

transcription and which ranged from 15 to 65 nucleotides were detected with the Eco and 7.2 probes; in every case the bands protected by RNA from 11 hour embryos were more intense than those protected by pluteus RNA. In addition, a fragment of about 85-90 nucleotides was protected by the muscle actin mRNA from pluteus stage embryos. When the Msp probe was used the three largest

repeat protected fragments were eliminated, further confirming that the *in vivo* transcription begins to the left of the MspI site at nucleotide 604 in pUC8-14 (see Figure 1). Interestingly, when the Dde probe was used, the RNA from these two stages protected none of the bands that correspond to transcripts of the SURF1 sequences; although the 85 nucleotide muscle-actin-mRNA fragment is protected by pluteus RNA. The lack of detectable transcripts of this repeated sequence family beyond the DdeI site beginning at nucleotide 740 might be explained in several different ways. First, and simplest, is that the *in vivo* transcription unit does not extend beyond the DdeI site at nucleotide 740. If this is the case, then the *in vivo* and *in vitro* transcription units are not entirely coincident because the *in vitro* transcription unit includes about 105 nucleotides beyond this site. Second, it is possible that *in vivo* transcription of the SURF1 sequence family extends over the entire distance mapped *in vitro*, but that most members contain different sequences at their 3' ends and that these sequences are therefore less abundant in RNA. We have demonstrated that there are nearly 800 genomic copies of the sequence between the MspI site beginning at nucleotide 604 and the DdeI site at nucleotide 740, and only about 40 copies of the sequence to the right of this DdeI site. If all 800 copies were transcribed equally then transcripts of this 3' sequence would be about 20-fold less abundant than the 5' part of the transcripts, making them more difficult to detect. Another possibility is that not all of the members of the SURF1 sequence family are transcriptionally active at these stages, and that those with 3' sequences related to the one described here are among the inactive copies.

Relationship of the SURF1-1 Sequence to the Muscle Actin Gene

The proximity of the SURF1-1 sequence to the muscle actin gene is of interest because of the possibility of a functional relationship between the two. The RNase protection experiment shown in Figure 7 demonstrates that a fragment of approximately 85-90 nucleotides is protected by pluteus stage RNA when probes extending as little as 0.333 kb or as far as about 7.2 kb upstream of the protein-coding sequence are used. When the protected fragments from similar experiments using RNA from several additional developmental stages are analyzed we find that no fragment greater than 85-90 nucleotides is protected, and that this fragment is not protected by RNA from stages before pluteus (data not shown). The appearance of this fragment at late embryonic stages indicates that it represents the portion of these probes which is protected by the muscle actin mRNA. Examination of the sequence in this region indicates that this fragment includes the first 45 nucleotides of protein-coding

sequence and about 40 nucleotides of 5' untranslated mRNA sequence (see Figures 1 and 4). Initiation of transcription of the muscle actin gene could begin at this position, but several observations suggest that this is not the case. First, the nucleotide sequence 40 nucleotides upstream of the actin-coding sequence (nucleotide 1033 in Figure 4) is similar to the splice acceptor site found at the 3' end of introns. Second, the nearest sequence resembling a TATA box is another 60 nucleotides upstream, and thus is not in a typical location to support transcription initiation at this site. Third, preliminary primer extension experiments indicate that the 5' untranslated portion of the muscle actin mRNA is about 180 nucleotides and not 40 nucleotides (unpublished observations). These considerations lead us to conclude that the SURF1-1 sequence probably lies in an intron within the 5' untranslated mRNA sequence of the muscle actin gene. Since no additional pluteus-RNA-specific protected fragments have been detected as far upstream as 7.2 kb from the protein-coding region, we cannot yet identify the site of initiation of transcription of the muscle actin gene.

DISCUSSION

A large part of the repeated sequence component of mammalian genomes is comprised of short interspersed elements (SINES, Ref. 2) which derive from RNA polymerase III genes, and move about the genome by means of an RNA intermediate (4,5,9). The distinguishing features of these elements are: (1) the presence of short direct repeated sequences at each end, which are duplications of the target site sequence, (2) the presence of an internal RNA polymerase III promotor sequence, (3) a simple sequence at the 3' end of the transcription unit which is usually A-rich, and (4) homology with an RNA polymerase III gene. Repeated sequences with these features have also been found in the genomes of the tortoise and the newt (12), salmon (13) and the silk moth (14). The experiments presented here describe a sea urchin sequence with the features that are characteristic of the polIII/SINE retroposons reported in other organisms, but has no detectable sequence homology with those sequences. This particular sea urchin element, SURF1-1, probably lies within the first intron of the single muscle actin gene and the family of these elements is present about 800 times in the genome.

The boundaries of the SURF1-1 sequence were defined both by sequence features and *in vitro* transcription by RNA polymerase III. Mapping of the *in vitro* transcription unit indicated that it is 300 nucleotides long and falls approximately between nucleotides 550 and 850 on Figure 4, in the region

previously shown to contain a repeated sequence (27). Examination of the DNA sequence in this region revealed that the sequence features which are diagnostic of retroposons are present and appropriately positioned on the transcription unit. The sequence TATATA is repeated on either side of the *in vitro* transcription unit; a 21 nucleotide simple A+T-rich sequence is present at the 3' end of the transcription unit; a consensus B box sequence of an internal RNA polymerase III promotor is located 60-80 nucleotides downstream from the sites of initiation of transcription; a region of 40 nucleotides is homologous with several tRNA sequences. On the basis of this analysis we conclude that the SURF1-1 sequence is probably a short retroposon.

The copy number analysis of the SURF1 family demonstrated that the sequence on the 3' end is repeated about 40 times in the genome and that the 5' portion is present nearly 800 times. This observation might be explained in two ways. First, it is possible that there are about 800 copies of this element, but that they contain many different 3' ends. The following set of events could have produced this situation. An initial retrotransposition of a shorter sequence (<200 nucleotides at the 5' end) might have inserted multiple copies of the sequence into many different locations within the genome. Because these copies would carry their own internal RNA polymerase III promotors each might then have been further amplified through subsequent retrotransposition events. If in some (or many) of these cases the new transcripts extended to include some of the adjacent, and unrelated, genomic sequence then each newly created retroposon would carry its own unique 3' sequence, thus creating many hybrid elements with a common 5' portion. It has been suggested that transcription of retroposon elements terminates at fortuitous oligo(T) tracts located downstream in flanking genomic DNA (4); this would produce transcripts with different 3' ends depending on their chromosomal locations. The second possible explanation is that there are only 40 members of this retroposon family and that they all contain essentially the same sequence, but that during the original retrotransposition event a portion of a sequence already repeated in the genome (Spec repeat) was included in the retroposon. Again, transcription by RNA polymerase III into adjacent sequences could explain the addition of new sequence to the original transcript. We cannot exclude either of these possibilities, but it is interesting to note that *in vivo* transcripts related to the 5' portion of SURF1-1 are readily detected in RNA from 11 hour and pluteus stage embryos while none related to the 3' portion are seen (Figure 7). Furthermore, it was previously demonstrated that newly transcribed RNA from 11 hour embryos which hybridizes to this region is

mostly synthesized by RNA polymerase III (27). It thus appears that either the *in vitro* transcription unit mapped for SURF1-1 does not exactly coincide with the *in vivo* transcription of members of this family, or that the transcription of those members with the SURF1-1 3' end is too low to detect.

The SURF1-1 sequence reported here contains a 79 nucleotide region with homology to part of a sequence repeated in the *S. purpuratus* genome 2-3,000 times which has been shown to be present within the 3' untranslated portion of the mRNAs from Spec1 and Spec2, and which is 150-200 nucleotides long (1). No homology is detected beyond this 79 nucleotides with any of four cloned versions of the Spec repeat, and none of these sequences have features of retroposons. The evidence presented here also demonstrates that the Spec-repeat-related sequence is contained within SURF1 *in vivo* transcripts which are at least 140 nucleotides long. It is clear then that some members of the Spec-repeat family are included in autonomous RNA polymerase III transcription units, and some are contained within RNA polymerase II transcription units. However, much of the transcription of these sequences detected by the SURF1-1 sequence in early embryos (Figure 2 and Ref. 27) is by RNA polymerase III and is of only one strand suggesting that there is relatively little expression of those copies which are attached to RNA polymerase II transcription units. Considering the sequence heterogeneity of the Spec-repeat family [about 80% homology among the copies reported by Carpenter *et al.*, (1), and only 72% homology of the SURF1-1 sequence with the pSpec1 sequence] it is likely that the SURF1-1 sequence hybridizes only with a subset of the Spec-repeat family. This is further suggested by the different number of genomic copies detected by Carpenter *et al.* (1), 2-3,000, and reported here, about 800. It is therefore possible that the SURF1-1 sequence preferentially detects those members which are in the SURF1 retroposon family. We cannot be certain of the relationship between the Spec-repeat family and the SURF1 retroposon family or the origin of the SURF1 family, but we suggest the following simple scenario. Initially, the Spec sequence became amplified and dispersed throughout the genome by DNA mediated transposition mechanisms such as recombination or transposons, and one copy of this sequence was inserted just downstream of a tRNA gene. Subsequently, a transcript of this tRNA gene which extended into the Spec sequence was copied by reverse transcriptase, and multiple copies of these retroposons were then inserted into dispersed regions of the genome. After this, some of these SURF1 retroposons carrying the Spec repeat sequence were further amplified and dispersed, picking up a variety of different 3' ends in the process.

The question of what function, if any, this or other RNA polymerase III transcribed repeated sequences might have is puzzling and unresolved. The Alu family repeats of human and the B2 family of mouse are present both on RNA polymerase II transcripts and as autonomous RNA polymerase III transcripts (reviewed in 4 and 5), as is likely the case for the SURF1 family described here. Furthermore, it has been shown that transcription of the mouse B2 family decreases during fetal development and during the differentiation of EC and EK cell lines (34), and that RNA polymerase III transcription of B2 sequences increases substantially in NIH 3T3 cells after transformation by SV40 (35). As a result of these studies, it has been suggested that the transcription of SINES is generally higher in undifferentiated tissues than in differentiated tissues (5). Expression of the SURF1 sequences also follows this general pattern during embryogenesis, that is, expression is highest during the cleavage stages and decreases as development proceeds during which time there is increasing differentiation of a variety of cell types. Whether the apparent relationship between expression of these sequences and differentiation state reflects a cause and effect relationship is not known. The particular copy of the SURF1 family which we have characterized here, SURF1-1, was initially discovered because it lies about 200 nucleotides upstream from the protein-coding sequence of the muscle actin gene. Our current analysis suggests that this is within a large intron in the 5' untranslated mRNA sequence of this gene. However, since we have been unable to locate the predicted first exon within the first 7.2 kb upstream, we cannot be certain of the structure of this part of the muscle actin gene. Nevertheless, the close proximity of the SURF1-1 element, which contains a functional RNA polymerase III promotor, to the muscle actin gene raises the question of whether it might affect expression of its actin gene neighbor. This could be tested by assessing the effect of deletion of this sequence on expression of the cloned muscle actin gene injected into sea urchin embryos.

ACKNOWLEDGEMENTS

We thank David S. Durica for the gift of the clone, pGEM4-7.6, Gary Kunkel for the HeLa cell S100 extract and Barbara Lee for the graphic artwork. We also thank Drs. William Klein and Paul Hardin for their comments on the manuscript and Sandra Johnson and Carol Savage for typing the manuscript. This research was supported by NIH Grants GM24620 and CA12708. P.E.N. is a trainee on an institutional NRSA, HD07312.

*To whom correspondence should be addressed

+Present address: Northeastern University School of Law, 400 Huntington Avenue, Boston, MA 02115, USA

REFERENCES

1. Carpenter,C.D., Bruskin,A.M., Spain,L.M., Eldon,E.D. and Klein,W.H. (1982) Nucl. Acids Res. 10, 7829-7842.
2. Singer,M.F. (1982) Cell 28, 433-434.
3. Deininger,P.L. and Daniels,G.R. (1986) TIG 2, 78-80.
4. Rogers,J. (1985) Int. Rev. Cytol. 93, 187-279.
5. Weiner,A.M., Deininger,P.L. and Efstratiadis,A. (1986) Ann. Rev. Biochem. 55, 631-661.
6. Jagadeeswaran,P., Forget,B.G. and Weissman,S.M. (1981) Cell 26, 141-142.
7. Ullu,E., Murphy,S. and Melli,M. (1982) Cell 29, 195-202.
8. Ullu,E. and Tschudi,C. (1984) Nature 312, 171-172.
9. Daniels,G.R. and Deininger,P.L. (1985) Nature 317, 819-822.
10. Lawrence,C.B., McDonnell,D.P. and Ramsey,W.J. (1985) Nucl. Acids Res. 13, 4239-4252.
11. Sakamoto,K. and Okada,N. (1985) J. Mol. Evol. 22, 134-140.
12. Endoh,H. and Okada,N. (1986) Proc. Natl. Acad. Sci. USA 83, 251-255.
13. Matsumoto,K., Murakami,K. and Okada,N. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 3156-3160.
14. Adams,D.S., Eickbush,T.H., Herrera,R.J. and Lizardi,P.M. (1986) J. Mol. Biol. 187, 456-478.
15. Graham,D.E., Neufeld,B.R., Davidson,E.H. and Britten,R.J. (1974) Cell 1, 127-137.
16. Klein,W.H., Thomas,T.L., Lai,C., Scheller,R.H., Britten,R.J. and Davidson,E.H. (1978) Cell 14, 889-900.
17. Cohen,J.B., Hoffman-Liebermann,B. and Kedes,L. (1985a) Mol. Cell. Biol. 5, 2804-2813.
18. Johnson,S.A., Davidson,E.H. and Britten,R.J. (1984) J. Mol. Evol. 20, 195-201.
19. Posakony,J.W., Scheller,R.H., Anderson,D.M., Britten,R.J. and Davidson.E.H. (1981) J. Mol. Biol. 149, 41-67.
20. Yager,L.N., Kaumeyer,J.F., Lee,I. and Weinberg,E.S. (1987) J. Mol. Evol. 24, 346-356.
21. Hoffman-Liebermann,B., Liebermann,D., Kedes,L.H. and Cohen,S.N. (1985) Mol. Cell. Biol. 5, 991-1001.
22. Liebermann,D., Hoffman-Liebermann,B., Weinthal,J., Childs,G., Maxson,R., Mauron,A., Cohen,S.N. and Kedes,L. (1983) Nature 306, 342-347.
23. Cohen,J.B., Liebermann,D. and Kedes,L. (1985b) Mol. Cell. Biol. 5, 2814-2825.
24. Crain,W.R., Boshar,M.F., Cooper,A.D., Durica,D.S., Nagy,A. and D. Steffen (1987) J. Mol. Evol. 25, 37-45.
25. Durica,D.S., Schloss,J.A. and Crain,W.R. (1980) Proc. Natl. Acad. Sci. USA 77, 5683-5687.
26. Weil,P.A., Segall,J., Harris,B., Ng,S. and Roeder,R.G. (1979) J. Biol. Chem. 254, 6163-6173.
27. Hickey,R.J., Boshar,M.F. and Crain,W.R. (1987) Dev. Biol., in press.
28. Soiero,R. and Darnell,J.E. (1969) J. Mol. Biol. 126, 783-802.
29. Maxam,A.M. and Gilbert,W. (1980) Methods Enzymol. 65, 499-560.
30. Weil,P.A., Luse,D.S., Segall,J. and Roeder,R.G. (1979) Cell 18, 469-484.
31. Fowlkes,D.M. and Shenk,T. (1980) Cell 22, 405-413
32. Fuhrman,S.A., Deininger,P.L., LaPorte,P., Friedmann,T. and Geiduschek,E.P. (1981) Nucl. Acids Res. 9, 6439-6456.
33. Perez-Stable,C., Ayres,T.M. and Shen,C.J. (1984) Proc. Natl. Acad. Sci. USA 81, 5291-5295.
34. Murphy,D., Brickell,P.M., Latchman,D.S., Willison,K. and Rigby,P.W.J. (1983) Cell 35, 865-871.
35. Singh,K., Carey,M., Saragosti,S. and Botchan,M. (1985) Nature 314, 553-556.