Volume 10 Number 23 1982 Nucleic Acids Research

The 3' untranslated regions of two related mRNAs contain an element highly repeated in the sea urchin genome

Clifford D.Carpenter, Arthur M.Bruskin, Lisa M.Spain, Elizabeth D.Eldon and William H.Klein

Program in Molecular, Cellular, and Developmental Biology, Department of Biology, Indiana University, Bloomington, IN 47405, USA

Received 21 June 1982; Revised and Accepted 25 October 1982

ABSTRACT

Two closely related cDNA clones, pSpec1 and pSpec2, specifying two developmentally regulated tissue specific mRNAs from sea urchin embryos were used to probe a sea urchin genomic λ library. Screening 10,000 phage by plaque hybridization yielded several hundred positive signals. With more stringent wash procedures, only two to three phage were positive. Three of these phage, one isolated by stringent wash procedures and two isolated by standard wash procedures were further investigated by restriction analysis, RNA gel blots, and DNA sequencing. The phage isolated by the stringent wash procedure appears to be a gene coding for the Spec1 mRNA. The other phage contain only partial homology to pSpec1 and pSpec2, 150 to 200 base pairs of the 3' untranslated region of the Spec1 and Spec2 mRNAs. It is concluded that the Spec1 and Spec2 mRNAs contain a highly repetitive element near their 3' end. The element is present at 2000 to 3000 copies per genome and may be transcribed at some sites other than those coding for the Spec1 and Spec2 genes. The possible function and evolutionary origin of the repetitive element is discussed.

INTRODUCTION

Repetitive sequence elements, a few hundred base pairs in length, are known to be present in most eukaryotic genomes (see 1 for review). The sea urchin genome, for example, contains approximately 500,000 of these elements which fall into 3000 distinct families of sequences. Earlier studies using sea urchin genomic DNA have shown that repetitive elements from a given family can be very close in sequence, 97% homology or more, or widely divergent, 80% homology or less (2, 3). The function, if any, of the repetitive DNA families is obscure, though there have been a number of suggestions. These include controlling elements for transcriptional regulation (4, 5), transposable elements (6), or selfish DNA (7). Repetitive elements are known to be transcribed, but sequencing analysis suggests that they do not contain open reading frames (3) and thus do not appear in translated regions of mRNAs. In fact, it seems clear that most transcripts containing repetitive elements are confined to the nucleus.

0305-1048/82/1023-7829 $2.00/0

Heterogeneous nuclear RNA contains short repetitive elements, while mRNA in general seems to lack them (1). Several genes have now been shown to contain repetitive elements in their intervening sequences. These include the genes for silk fibroin (9), vitellogenin (10), CAD (11), serum albumin (12), and corticotropin-β-lipotropin (13). These examples all result in mRNAs that are devoid of repetitive sequences. In several species, however, a small number of mRNAs do contain repetitive elements (14, 15, 16). Recently, Calabretta et al. (17) reported that a large fraction of poly A^+ polysomal RNA from a human lyphoblastoid cell line contains the Alu repeat. In Dictyostelium, a repetitive element has been found at the 5' end of approximately 100 different mRNAs (18), and another repetitive element is associated with the developmentally regulated mRNAs appearing at the late aggregation stage (19). Tchurikov et al. (20) have found a gene in Drosophila melanogaster with a 3' exon that is present at the 3' ends of many different mRNAs.

In unfertilized sea urchin eggs at least 70% of the cytoplasmic maternal poly A^+ RNA contains repetitive elements (21). Embryonic polysomal mRNA is largely lacking these elements. These facts have led Davidson and his co-workers to hypothesize that many maternal RNAs represent stable unprocessed transcripts which do not function directly as mRNAs (22).

We now present evidence that functioning sea urchin embryonic mRNAs can indeed contain short repetitive elements. We have reported previously on two cDNA clones, pSpec1 and pSpec2, specifying two related sea urchin mRNAs. These mRNAs are absent or extremely rare in the unfertilized egg and increase in mass over 100 fold during the interval from early blastula to late gastrula stage (23, 24). The mRNAs are part of a small family of mRNAs that code for approximately ten low molecular weight acidic proteins actively synthesized in the embryonic ectoderm (24). Below we show that there are two classes of loci in the sea urchin genome which share sequence homology with the Spec1 and Spec2 mRNAs. Members of the first class are present at only a few sites in the genome and appear to contain the entire Spec1 or Spec2 mRNA sequence. These sites presumably are the genes for the ectoderm proteins. Members of the second class are present at 2000 to 3000 sites in the genome and contain only a portion of the Spec1 or Spec2 mRNAs; that portion is in the 3' untranslated region of the mRNAs. Although some of the sites representing the second class may be

transcribed, they do not represent genes coding for the ectodermal proteins.

MATERIALS AND METHODS

cDNA clones, λ clones, and plaque screens

The isolation and characterization of the Strongylocentrotus purpuratus pluteus cDNA clones, pSpec1 and pSpec2 has been described previously (23, 24). Nick-translated ^{32}P labeled pSpec1 DNA was used to screen a Charon 4 λ sea urchin genomic library by the method of Benton and Davis (25), except that the hybridization conditions were as described for the Southern hybridization in Bruskin et al (23). The library was a gift of Dr. Eric Davidson (California Institute of Technology) and was constructed by partially digesting sperm DNA from a single individual with HaeIII and inserting the DNA into Charon 4 using EcoR1 linkers. Final post-hybridization wash conditions included either 1 x SET (0.15 M NaCl, 0.03 M Tris, 0.0025 M EDTA pH 8.0), 1 x Denhardts (0.02% BSA, 0.02% Ficoll, 0.02% polyvinylpryolidone), 0.1% sodium pyrophosphate, and 0.025 M phosphate buffer at 68°C; or 0.2 x SET, 0.025 M phosphate buffer, 0.1% sodium pyrophosphate, and 0.1% SDS at 75°C.

RNA gel blots

Isolation of total cellular RNA by the guanidine-HCl method and subsequent purification of poly A^+ RNA was as described (23). For blotting purposes, either 3 µg of poly A^+ RNA or 10 µg of total cellular RNA was electrophoresed on formaldehyde-agarose gels, blotted to nitrocellulose, and hybridized with nick-translated probes as described (24).

DNA sequencing

Restriction fragments, labeled at either a unique 5' end with T4 polynucleotide kinase (BRL) or 3' end with AMV reverse transcriptase (courtesy of Dr. J.W. Beard, Life Sciences, Inc.), were sequenced according to the technique of Maxam and Gilbert (26).

RESULTS

Isolation of λ recombinants complementary to pSpec1 and pSpec2 cDNA clones

We have previously shown that pSpec1 and pSpec2 are closely related cDNA clones specifying what appears to be the 3' ends of two developmentally regulated tissue specific sea urchin mRNAs (23, 24).

These mRNAs appear in the ectoderm of the late stage embryo. We concluded that pSpec1 represents 0.5 kb of the 3' end of a 1.5 kb mRNA and pSpec2 represents 0.5 kb of the 3' end of a 2.2 kb mRNA. Because pSpec1 and pSpec2 have sequence homology, they also hybridize weakly to each other and to at least one other RNA 3.2 kb in size (see figure 3). When the pSpec1 or pSpec2 clones are used for selected translation of mRNAs, the resulting cell-free translation yields approximately ten low molecular weight acidic proteins (ranging in size from 14,000 to 17,000 daltons as seen on two-dimensional gels) (24). Both clones select the same set of mRNAs.

Based on these experiments, we screened a Charon 4 λ sea urchin DNA library with pSpec1 or pSpec2 with the expectation of finding a small number of positive genomic clones, each clone representing a gene for an ectoderm protein. A standard plaque hybridization screen of about 10,000 recombinants using nick translated pSpec1 DNA as a probe is shown in figure 1A. Many more recombinants hybridize than were expected. Several hundred positives are seen in such a screen. When random plaques scoring positive are purified, re-plated, and rehybridized with pSpec1, they again score as positive. Simple calculations from these data suggest that all or part of the pSpec1 and pSpec2 sequences are reiterated 2000 to 3000 times in the sea urchin genome.

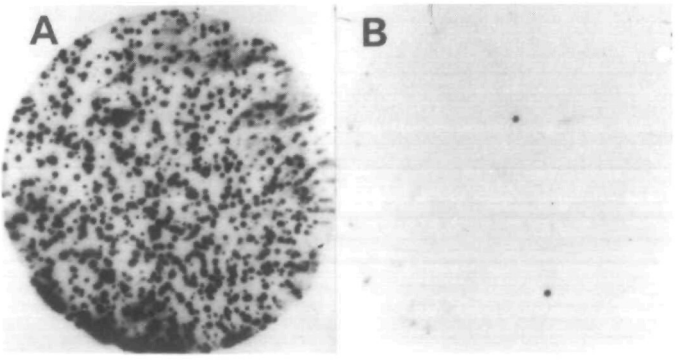

Figure 1: Plaque hybridization with pSpec1 and sea urchin genomic DNA-λDNA recombinants. Ten thousand plaques per plate, each containing phage with approximately 15 kb of inserted sea urchin sperm DNA, were lifted onto nitrocellulose and hybridized with 2 x 10^7 cpm of nick translated [^{32}P] pSpec1. A) Final wash was 0.15 M NaCl at 68°C. B) Final wash was 0.03 M NaCl at 75°C. The filter in panel A was exposed to film 5 times longer than the filter in panel B.

The number of plaques scoring positive in such screens can be greatly reduced by increasing the wash criteria to 0.03 M NaCl at 75°C. A representative experiment is shown in figure 1B. Here only two positive clones per 10,000 are discernable. This experiment, together with that shown in figure 1A, suggests that most of the λ clones have only weak homology with the pSpec1 probe.

To determine the relationship between the different λ recombinants hybridizing with pSpec1, several different phage were isolated from the different screens. DNA was purified from these phage and restriction maps were constructed. Figure 2 shows the restriction maps of three of these clones, λSpec1 isolated using the higher criteria wash, and λrep1 and λrep2 isolated using the lower criteria wash. Figure 2 also shows the region of homology with pSpec1 (or pSpec2) determined by Southern analysis: a 0.9 kb EcoR1-Sal1 fragment in λSpec1, a 0.9 kb EcoR1-Hind3 fragment in λrep1, and a 0.6 kb BamH1-EcoR1 fragment in λrep2. Besides these small regions of homology, the clones have no other homology as judged both by the restriction maps shown in figure 2 and by cross

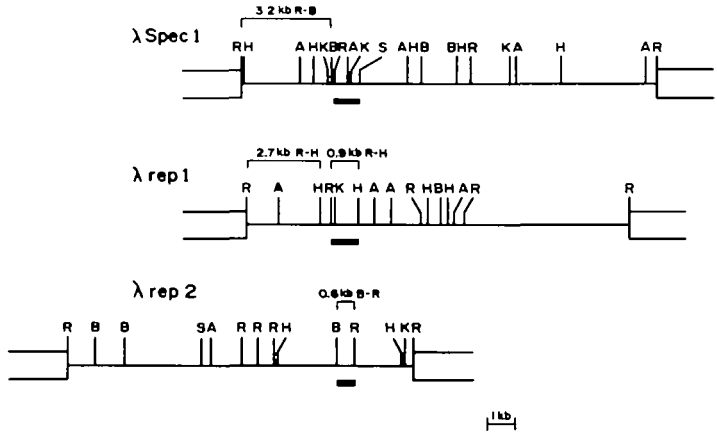

Figure 2: Restriction maps of λSpec1, λrep1, and λrep2. Symbols for restriction sites are: R, EcoR1; H, Hind3; A, Xba1; K, Kpn1; B, BamH1; S, Sal1. The EcoR1 sites bordering the inserted DNA derive from EcoR1 linkers ligated to partial HaeIII digested sperm DNA used in construction of the library. The thick dark bars underneath each map represent the smallest restriction fragment tested which hybridized to nick translated pSpec1 in a Southern analysis. The orientation of the gene on λSpec1 is 5' to 3' left to right on the figure. The restriction fragments shown in brackets have been isolated and subcloned into pBr322.

hybridization experiments. For example, when the 3.2 kb EcoR1-BamH1 fragment of λSpec1 (indicated in figure 2) is subcloned into pBr322, nick translated, and hybridized to blots of λrep1 or λrep2 DNA, there is no detectable hybridization (data not shown). However, several λ clones have thus far been isolated by hybridizing with pSpec1 and washing at high criteria and these show extensive sequence homology. Thus the λ clones hybridizing with pSpec1 fall into two classes: those which appear at high frequency in the genome and are weakly homologous to pSpec1, and those which are far fewer in number, probably fewer than ten different genomic loci and are closely related.

Hybridization of pSpec1, pSpec2, λspec1, and λrep1 with pluteus ectoderm and endoderm/mesoderm RNA

If any of the λ clones discussed in the previous section contain genes for the 1.5 kb or 2.2 kb ectoderm mRNAs, they should hybridize with these RNAs in a Northern analysis. To test whether λSpec1, λrep1, and λrep2 contain genes coding for ectoderm enriched mRNAs, pluteus stage embryos were fractionated into ectoderm and endoderm/mesoderm and total RNA was extracted from each fraction. RNA gel blots hybridized with nick translated pSpec1, pSpec2, λSpec1, and λrep1 DNAs are shown in figure 3A. As noted above, pSpec1 and pSpec2 hybridize to RNAs greatly enriched in the pluteus ectoderm (figure 3A, lanes 1-4). pSpec1 hybridizes strongly to a 1.5 kb RNA and weakly to a 2.2 kb and 3.2 kb RNA. pSpec2 hybridizes strongly to a 2.2. kb RNA and weakly to a 1.5 kb and 3.2 kb RNA. Data similar to this have been published elsewhere (24). λSpec1 hybridizes with a pattern very similar to that of the pSpec1 cDNA clone (lanes 5, 6). We have also hybridized with the 3.2 kb EcoR1-BamH1 subclone of λSpec1 (indicated in figure 2) and shown that this probe also hybridizes strongly with the 1.5 kb RNA (data not shown). In addition, we have sequenced approximately 200 bp of λSpec1 corresponding to the pSpec1 homology and found it has exactly the same sequence as pSpec1 (see figure 4). This provides further evidence that λSpec1 is the site of transcription of a Spec1 mRNA. The simplest interpretation of these data is that λSpec1 contains a gene for a 1.5 kb ectoderm enriched mRNA.

λrep1 gives a much different pattern than λSpec1 (figure 3A, lanes 7, 8). λrep1 does not hybridize to any discrete RNAs but only to heterogeneous high molecular weight transcripts. We interpret this as hybridization to nuclear RNA. λrep1 is a 13.4 kb genomic clone which probably has several repetitive elements along its length, each of which

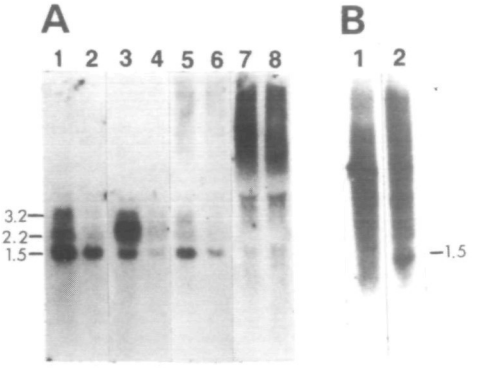

Figure 3: A) RNA gel blots of pluteus ectoderm and endoderm/mesoderm RNA hybridized with pSpec1, pSpec2, λSpec1, and λrep1. 10 μg of total pluteus ectoderm RNA (lanes 1, 3, 5, 7) or endoderm/mesoderm RNA (lanes 2, 4, 6, 8) were electrophoresed on 1.0% agarose gels, blotted to nitrocellulose filters, and hybridized with 2 x 10^7 cpm of nick translated [^{32}P] probe. Lanes 1 and 2, pSpec1; lanes 3 and 4, pSpec2; lanes 5 and 6, λSpec1; lanes 7 and 8, λrep1. The sizes of the transcripts in kb are indicated. The two interuptions seen in the pattern of hybridization with λrep1, lanes 7 and 8, correspond to the positions of the 28S and 18S ribosomal RNAs. The interuptions are due to the saturation of the nitrocellulose filters with ribosomal RNA in these regions, thereby preventing normal hybridization. No such interuptions are apparent when poly A$^+$ RNA is used for the blots rather than total RNA as can be seen in B. B) RNA gel blots of gastrula stage poly A$^+$ RNA with subcloned fragments of λrep1. 3 μg of poly A$^+$ RNA was electrophoresed on 1% agarose gels, blotted onto nitrocellulose, and hybridized with 2 x 10^7 cpm of nick translated [^{32}P] probe. Lane 1, hybridization using the 2.7 kb kb EcoR1-Hind3 fragment of λrep1; lane 2, hybridization using the 0.9 kb EcoR1-Hind3 fragment of λrep1. The band observed at 1.5 kb is indicated on the right.

may hybridize with a multitude of other sites in the genome. The hybridization of λrep2 with total RNA is identical to that of λrep1 (data not shown). Thus while λrep1 and λrep2 may be transcribed they do not contain genes for the Spec1 or Spec2 mRNAs.

If λrep1 hybridizes to pSpec1, and pSpec1 hybridizes to distinct ectoderm specific mRNAs, then λrep1 should also hybridize to the same set of RNAs. The fact that no hybridization of this type can be seen is explained by the weak homology between pSpec1 and λrep1. To determine if even weak homology with the Spec1 mRNA could be detected, the 0.9 kb EcoR1-Hind3 fragment of λrep1 containing the pSpec1 homology was subcloned into pBr322. We also subcloned the nearby 2.7 kb EcoR1-Hind3 fragment. To further increase the sensitivity of the hybridization reactions, the

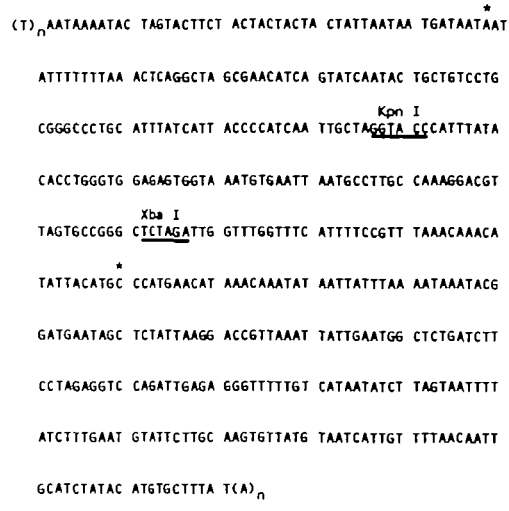

Figure 4: Nucleotide sequence of pSpecl. In order to facilitate comparison to figure 6, the sequence of pSpecl is oriented in an unconventional manner. The displayed strand is the transcribed strand and is 5' to 3' left to right on the figure. The mRNA sequence is the complementary strand and the direction of transcription is right to left on the figure. This direction was determined by labeling pSpecl DNA at the Xba site indicated in the sequence with [^{32}P]α dCTP and reverse transcriptase. The labeled DNA was then digested with HhaI yielding two fragments labeled at the 3' end of opposite strands. These fragments were isolated from polyacrylamide gels and then used as probes in RNA gel blot hybridizations. The orientation was determined by the fragment that hybridized to the RNA. The stars bound the portion of the pSpecl sequence shown in Figure 6.

probes were hybridized with poly A$^+$ RNA instead of total RNA (the Specl and Spec2 mRNAs are polyadenylated, unpublished results). The results of these experiments are shown in figure 3B (lanes 1,2). Besides the hybridization to heterogeneous high molecular weight transcripts, the 0.9 kb EcoRl-Hind3 fragment shows a discernable band at 1.5 kb (figure 3B, lane 2). The 2.7 kb EcoRl-Hind3 fragment does not yield such a band, but only the heterogeneous hybridization (figure 3B, lane 1). These experiments suggest that λrepl contains sequences that are only weakly homologous to the pSpecl sequence.

Nucleotide sequence of pSpecl and evidence for its location in the 3' untranslated region of the Specl mRNA.

The nucleotide sequence of the pSpecl cDNA clone is shown in figure 4. It is 470 bp of sea urchin sequence, which corresponds to about one-third of the total mRNA length. The sense and anti-sense strands of

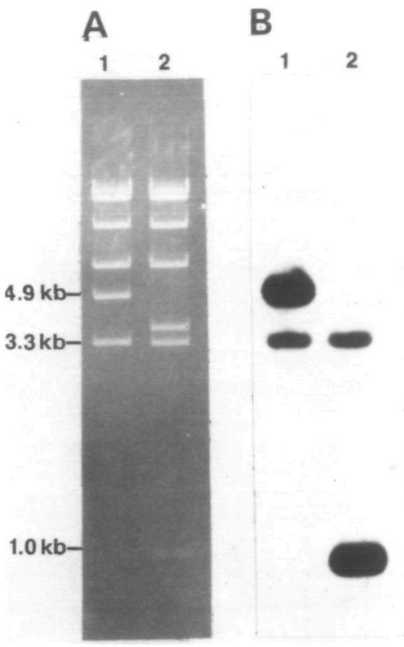

Figure 5: Restriction fragments of λSpecl hybridizing with [^{32}P] polyA$^+$ RNA. DNA from λSpecl was digested with EcoRl (lane 1) or EcoRl-Sall (lane 2), electrophoresed on a neutral agarose gel, and transferred to nitrocellulose. The filter was hybridized with 8 x 10^7 cpm of [^{32}P] polyA$^+$ RNA labeled with γ[^{32}P] ATP using polynucleotide kinase. Post hybridization wash conditions were as described by Bruskin et al. (23). A) Ethidium bromide stained gel. B) Autoradiograph of hybridized filter.

pSpecl have been determined as detailed in the legend to figure 4 and as described previously (24). Inspection of the pSpecl sequence reveals no open translational reading frames extending toward the 5' end of the mRNA. Since the pSpecl cDNA clone was constructed by cloning cDNA-RNA hybrids in pBr322 after reverse transcription of oligo dT-primed polyA$^+$ RNA it is likely that pSpecl represents a sequence at or near the 3' end of the mRNA. To determine approximately how close the pSpecl sequence is to the 3' end of the mRNA, we have made use of transcription mapping with λSpecl and embryonic polyA$^+$ RNA.

λSpecl was digested with EcoRl or EcoRl and Sall and used for a Southern analysis with radioactive polyA$^+$ RNA. The RNA probe was prepared by hydrolyzing the RNA in base to 200 nucleotides and labeling the 5' ends with γ[^{32}P] ATP and polynucleotide kinase. Figure 5A shows the ethidium bromide stained agarose gel of λSpecl digested with EcoRl or EcoRl-Sall. There are three EcoRl insert fragments: 6.5 kb, 4.9 kb, and 3.3 kb. The 4.9 kb EcoRl fragment contains the pSpecl sequence (see figure 2). This fragment is cleaved by Sall to a 3.9 kb fragment and a 1.0 kb fragment which contains the entire 470 bp pSpecl sequence (figure 5A, lane 2). The hybridization of these fragments with the RNA probe is shown in figure 5B.

Hybridization is seen with the 4.9 kb and 3.3 kb EcoRl fragments (figure 5B, lane 1). Since λSpecl appears to contain a gene for a 1.5 kb ectoderm enriched RNA, it would be expected to show hybridization further upstream from the 4.9 kb EcoRl fragment. The adjacent 3.3 EcoRl fragment presumably contains such hybridizing sequences. As mentioned above, the 3.2 kb EcoRl-BamHl fragment of λSpecl hybridizes strongly to a 1.5 kb RNA in a Northern analysis. Figure 5B, lane 2, shows that after digestion with EcoRl and Sall the 3.3 kb EcoRl fragment hybridizes, as does the 1.0 kb EcoRl-Sall fragment. There is no hybridization to the right of the Sall site on λSpecl.

Since the Sall site is approximately 250 bp from the end of the pSpecl sequence, these data, together with the sequencing analysis, demonstrate that pSpecl is 470 bp of untranslated sequence mapping no more than 250 bp from the 3' end of the Specl mRNA. Thus, while it is not certain whether pSpecl is at the very 3' end of the Specl mRNA, it clearly appears to be somewhere in the 3' untranslated region. We assume that an analogous situation is true for pSpec2 and its mRNA.

Sequence analysis of the homologous regions of pSpecl, pSpec2, λrepl, and λrep2.

In order to ascertain the relationship between λrepl, λrep2, and pSpecl and pSpec 2, we have sequenced the appropriate regions of these DNAs. The phage DNA sequences were determined by subcloning the fragments indicated in figure 2. A summary of this sequence data is shown in figure 6. When λrepl and λrep2 are compared, a short sequence homology of about 150 bp of both λrepl and λrep2 with the pSpecl and pSpec2 sequences can be clearly seen. This homology centers around a Kpnl site present in pSpecl and λrepl. The Kpnl site is 135 bp from the 3' end of pSpecl. In the case of λrepl, the homology with pSpecl is about 80% and diverges away sharply about 90 bp from the left side of the Kpnl site and 70 bp from the right side. In the case of λrep2, the homology is also about 80%; it diverges from pSpecl about 60 bp from the left side of the Kpnl site. The divergence to the right of the Kpnl site has not been mapped since the sequences are still reasonably homologous up to the EcoRl site at the end of the insert. However, because the adjacent sequences (e.g., the 2 kb EcoRl fragment) to the right of the 0.5 kb BamHl-EcoRl fragment do not hybridize with pSpecl DNA (see figure 2), it seems likely the homology must cease somewhere between 75 and 150 bp from the Kpnl site. We conclude from these experiments that the weak hybridization of λrepl and

Figure 6: Partial sequences of pSpec1, pSpec2, λrep1, and λrep2. pSpec1 has been completely sequenced (470 bp) and pSpec2 has been partially sequenced; the regions homologous to λrep1 and λrep2 are shown. The subcloned 0.9 kb EcoR1-Hind3 fragment of λrep1 and the 0.6 kb BamH1-EcoR1 fragment of λrep2 were partially sequenced and their homologous regions are shown. The shaded areas are regions of homology. (o) denotes a gap in the sequencing ladder (this usually implies a methylated cytosine). The boxed region denotes a Kpn1 site found in pSpec1 and the 0.9 kb EcoR1-Hind3 subclone. The displayed strand is 5' to 3', left to right. λrep1 and pSpec1 (corresponding to λSpec1) are oriented opposite to the maps shown in figure 2. The λrep2 sequence is oriented in the same way as the map in figure 2. The vertical arrows mark the boundaries of homology for λrep1 and λrep2 with pSpec1. No boundary for λrep2 to the right of the Kpn1 site has been found.

λrep2 with pSpec1 and pSpec2 is due in both cases to a 150–200 bp sequence homology of about 80%.

DISCUSSION

We have found two classes of λ recombinants homologous to the pSpec1 and pSpec2 cDNA clones. One class hybridizes weakly to pSpec1 (or pSpec2), and the other class hybridizes strongly. From the data presented here, we suggest that a repetitive sequence family whose elements are

approximately 150 to 200 base pairs in length is present at 2000 to 3000 copies per haploid genome and that one member of this family is present in the 3' untranslated region of the Spec1 mRNA and another is in the 3' untranslated region of the Spec2 mRNA.

A comparison of the sequence homology of two representative genomic clones with the pSpec1 and pSpec2 cDNA clones suggests that the exact boundaries of the repetitive element are ill-defined. Nevertheless, since λrep1 and λrep2 were chosen at random and show a large degree of homology to each other, it seems unlikely that more than one "block" of the Spec1 (or Spec2) mRNA is reiterated to the extent seen in the λ screens. The simplest conclusion is that all 2000 to 3000 sites carry essentially the same repetitive element as shown in figure 6. This, of course, is not the only interpretation. We estimate the 3' untranslated region of the Spec1 mRNA to be almost 1 kb in length (Bruskin, Carpenter, and Klein, unpublished results), and therefore other repetitive elements could be associated with it.

As judged by Northern analysis and sequence homology, λSpec1 is much different than λrep1 or λrep2, and most likely is a gene for a Spec1 1.5 kb mRNA. We have also isolated a λ recombinant phage which appears to be a gene for a Spec2 2.2 kb mRNA (Bruskin, Carpenter, and Klein, unpublished results). Hybrid selection and cell-free translation experiments with either pSpec1 or pSpec2 DNAs yield approximately ten low molecular weight acidic peptides (24). This implies the existence of several more mRNAs and perhaps several more genes in this family.

It is not known how much more sequence homology beyond the repetitive element is in the 3' untranslated region of Spec1 and Spec2 mRNAs. A comparison of the pSpec1 and pSpec2 sequences suggests that this homology extends beyond that contained in either λrep1 or λrep2. Because of the physical relatedness of the proteins encoded by these mRNAs (all low molecular weight and similar in isoelectric point), it seems reasonable to suggest that additional homology will be seen when the genes coding for the Spec1 and Spec2 mRNAs are further analyzed.

λrep1 and λrep2 both hybridize with embryonic RNA. We have looked at four other genomic λ clones which hybridize with pSpec1 or pSpec2 under the lower stringency wash procedures, and these also yield RNA hybridization identical to those shown by λrep1 in figure 3A. Because of the complex nature of these probes, however, we cannot say with certainty how many of these sites are actually transcribed. The heterogeneous high

molecular weight of the transcripts suggests the λ clones are hybridizing with nuclear RNAs. These transcripts could represent precursors to mRNAs or low levels of constitutively expressed RNAs of unknown function (27). The intensity of the hybridization is probably the result of the multitude of sites represented by the probe. If a given λ clone with a 15 kb insert had five different repetitive elements on it and each element were reiterated 100 times in the genome, this would represent 500 different sites in the genome. If roughly one third of these sites were transcribed as nuclear RNA (28) and each site represented 0.01% of the nuclear RNA, this would amount to approximately 0.5% of the non-ribosomal RNA of the embryo. The fact that no discrete bands of mRNA size are seen in the six clones we have investigated suggests that most of the sites containing the pSpec1 homology do not code for mRNAs which accumulate to any significant extent (0.05% or so) in the embryonic cytoplasm. This is not unreasonable given that at most only 5% of the genome codes for mRNAs (29) and that the pSpec1 homology is randomly placed with respect to mRNA coding sequences.

Although no direct evidence has been obtained as yet, we find no reason to believe this repetitive element has a function. Several other mRNAs are known to have long 3' untranslated tails (30, 31), but the significance of these tails is obscure. We hypothesize that one repetitive element of the family discussed here inserted into a region of the genome between the ancestral Spec amino acid coding sequence and its transcriptional termination site. Subsequent amplification of the Spec genes led to a family of genes coding for embryonic ectoderm proteins with a highly repetitive element toward the 3' end of the genes. This simple model suggests an evolutionary origin for the repeat element in the mRNA without requiring any function for the repeat in transcription, processing, or translation.

ACKNOWLEDGEMENTS

We thank Angela Tyner for excellent technical assistance. This work was supported by Grant PHS HD14182 (to W.H.K.) from the National Institutes of Health. C.D.C. and A.M.B. were supported by a predoctoral training grant from the National Institutes of Health (GM7227).

REFERENCES

1. Davidson, E.H. and Posakony, J.W. (1982). Nature (London), in press.
2. Klein, W.H., Thomas, T.L., Lai, C., Scheller, R.H., Britten, R.J., and Davidson, E.H. (1978). Cell 14, 889-900.

3. Scheller, R.H., Anderson, D.M., Posakony, J.W., McAllister, L.B., Britten, R.J., and Davidson, E.H. (1981). J. Mol. Biol. 149, 15-39.
4. Britten, R.J. and Davidson, E.H. (1969). Science 165, 349-358.
5. Davidson, E.H. and Britten, R.J. (1979). Science 204, 1052-1059.
6. Haynes, S.R. and Jelinek, W.R. (1981). Proc. Natl. Acad. Sci. USA 78, 6130-6134.
7. Doolittle, W.F. and Sapienza, C. (1980). Nature (London) 284-601.
8. Posakony, J.W., Scheller, R.H., Anderson, D.M., Britten, R.J., and Davidson, E.H. (1981). J. Mol. Biol. 149, 41-67.
9. Pearson, W.J., Mukai, T., and Morrow, J.F. (1981). J. Biol. Chem. 256, 4033-4041.
10. Wahli, W., Dawid, I.B., Ryffel, G.U., and Weber, R. (1981). Science 212, 298-304.
11. Padgett, R.A., Wahl, G.M., and Stark, G.R. (1982). Molec. and Cellular Biol. 2, 302-307.
12. Sargent, T.D. (1981). Dissertation (Calif. Inst. Tech., Pasadena, Calif.).
13. Tsukada, T., Watanabe, Y., Nakai, Y., Imura, H., Nakanishi, S., and Numa, S. (1982). Nucleic Acids Res. 10, 1471-1479.
14. Klein, W.H., Murphy, W., Attardi, G., Britten, R.J., and Davidson, E.H. (1974). Proc. Natl. Acad. Sci. USA 71, 1785-1789.
15. Dina, D., Meza, I., and Crippa, M. (1974). Nature (London) 248,486-490.
16. Campo, M.S. and Bishop, J.O. (1974). J. Mol. Biol. 90, 649-663.
17. Calabretta, B., Robberson, D.L., Maigel, A.L., and Saunders, G.F. (1981). Proc. Natl. Acad. Sci. USA 78, 6003-6007.
18. Kimmel, A.R. and Firtel, R.A. (1979). Cell 16, 787-796.
19. Zuker, C. and Lodish, H.F. (1981). Proc. Natl. Acad. Sci. USA 78, 5386-5390.
20. Tchurikov, N.A., Naumova, A.K., Zelentsova, E.S., and Georgiev, G.P. (1982). Cell 28, 365-373.
21. Costantini, F.D., Britten, R.J., and Davidson, E.H. (1980). Nature (London) 287, 111-117.
22. Thomas, T.L., Posakony, J.W., Anderson, D.A., Britten, R.J., and Davidson, E.H. (1981). Chromosoma 84, 319-335.
23. Bruskin, A.M., Tyner, A.L., Wells, D.E., Showman, R.M., and Klein, W.H. (1981). Develop. Biol. 87, 308-318.
24. Bruskin, A.M., Bedard, P.A., Tyner, A.L., Showman, R.M., Brandhorst, B.P., and Klein, W.H. (1982). Develop. Biol. 92, in press.
25. Benton, W.D. and Davis, R.W. (1977). Science 196, 180-182.
26. Maxam, A.M. and Gilbert, W. (1980). Methods in Enzymol. 65, 499-560.
27. Wold, B.J., Klein, W.H., Hough-Evans, B.R., Britten, R.J., and Davidson, E.H. (1978). Cell 14, 941-950.
28. Hough, B.R., Smith, M.J., Britten, R.J., and Davidson, E.H. (1975). Cell 5, 291-299.
29. Galau, G.A., Klein, W.H., Davis, M.M., Wold, B.J., Britten, R.J., and Davidson, E.H. (1976). Cell 7, 487-505.
30. Moormann, R.J.M., van der Velden, H.M.W., Dodemont, H.J., Andreoli, P.M., Bhoemendal, H., and Schoenmakers, J.G.G. (1981). Nucleic Acids Res. 9, 4813-4823.
31. Scheller, R.H., McAllister, L.B., Crain, W.R., Durica, D.S., Posakony, J.W., Thomas, T.L., Britten, R.J., and Davidson, E.H. (1981). Molecular and Cellular Biol. 1, 609-628.