J Mol Evol (1987) 25:126–133

Journal of
Molecular Evolution

Unusual Sequence Conservation in the 5′ and 3′ Untranslated Regions of the Sea Urchin Spec mRNAs

Paul E. Hardin[1] and William H. Klein[2]

[1] Program in Genetics, Indiana University, Bloomington, Indiana 47405, USA
[2] Department of Biochemistry and Molecular Biology, M.D. Anderson Hospital, University of Texas System Cancer Center, Houston, Texas 77030, USA

Summary. The Spec1 and Spec2 mRNAs (*Strongylocentrotus purpuratus* ectoderm mRNAs) represent a small gene family that encodes 10–12 members of the troponin C superfamily of calcium-binding proteins. These mRNAs and proteins accumulate in the aboral (dorsal) ectoderm of sea urchin embryos and larvae. Using genomic and cDNA clones, we have compared the sequences of four Spec mRNAs: Spec1, Spec2a, Spec2c, and Spec2d. The mRNAs all have at least 120 bases of 5′ untranslated leader, approximately 450 bases of open reading frame, and 900 bases (Spec1) or 1250 bases (Spec2a, 2c, 2d) of 3′ untranslated trailer. Unexpectedly, when long stretches of 5′ untranslated regions or 3′ untranslated regions are compared to one another, they are found to be less divergent than the protein-coding regions. Comparing Spec2d, the most divergent member of the family, with the other Spec mRNAs shows that while the protein-coding regions are 60–62% matched, the untranslated regions are greater than 80% matched. Comparisons among Spec1, Spec2a, and Spec2c demonstrate similar but less dramatic conservation of untranslated regions. Our data imply that the Spec gene family has evolved differently from most gene families, with mutations accumulating most rapidly in intron regions, less rapidly in protein-coding regions, and least rapidly in 5′ and 3′ untranslated regions.

Key words: Troponin C superfamily — Sea urchins — 5′, 3′ Untranslated regions — Spec genes

Offprint requests to: W.H. Klein

Introduction

In general, untranslated 5′ leaders and 3′ trailers of eukaryotic mRNAs accumulate mutations at much faster rates than the corresponding protein-coding regions. There are numerous cases for which this fact has been demonstrated and then exploited using the 5′ and 3′ untranslated regions (UTRs) to distinguish a specific gene from all other genes of a closely related gene family. For example, the β_1 and β_2 tubulin genes in chickens were distinguished using a 3′ UTR from a β_2 tubulin cDNA clone (Lopata et al. 1983); 3′ UTRs of genes encoding actins have been used as specific probes to investigate temporal and spatial aspects of expression in sea urchins (Shott et al. 1984; Cox et al. 1986) and *Drosophila* (Fyrberg et al. 1983). Although the mutation rates in untranslated regions may be high, these regions are probably under some constraints. Structural and functional roles have been suggested in several studies, and it is likely that specific 5′ and 3′ untranslated sequences are involved in a variety of processes including message processing (Birchmeier et al. 1984), translation (Thireos et al. 1984; McGarry and Lindquist, 1985), and message stabilization (Morris et al. 1986).

In this report we describe the strong sequence conservation in the 5′ and 3′ UTRs of the Spec mRNAs (*Strongylocentrotus purpuratus* ectoderm mRNAs). The Spec mRNAs are actively transcribed from 10–12 related but distinct genes during the postcleavage embryonic and larval periods of the sea urchin life cycle (Bruskin et al. 1982; Klein et al. 1985). These messages are restricted to aboral

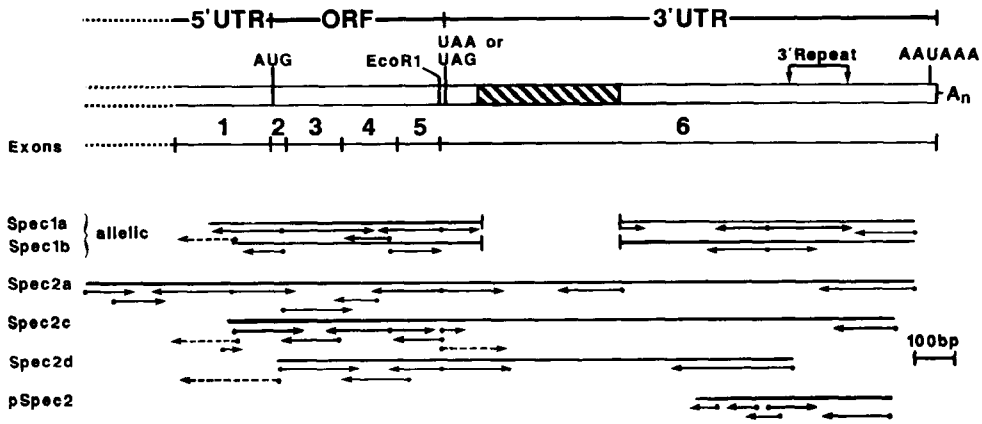

Fig. 1. Structure of Spec1, 2a, 2c, and 2d mRNAs. The first line depicts the 5′ untranslated region (5′ UTR), open reading frame (ORF), and 3′ untranslated region (3′ UTR) for Spec1a, 1b, 2a, 2c, and 2d. The transcriptional initiation site has been mapped for Spec1 but not for Spec2a, 2c, or 2d. The dashed line is to imply that the transcriptional initiation site for these latter messages may extend beyond that of Spec1. This is known to be true for Spec2a. The second line shows the placement of the translational initiation and termination codons as well as the EcoR1 site found in all the Spec genes at the start of exon 6. The cross-hatched area depicts the region in the 3′ UTR that is absent in Spec1a and Spec1b messages. The region designated 3′ repeat is a repetitive element found in all the Spec mRNAs and also in 2000–3000 other places in the *S. purpuratus* genome. The AAUAAA and -An are to represent the poly A processing signal and poly A tail found in the Spec messages. The third line shows the exon boundaries for the Spec genes. Exon 1 is untranslated except for the ATG; exon 6 is untranslated except for the 5′-most 26 bases (29 bases for Spec2d). The Spec cDNA clones are depicted as thick lines. The break in the Spec1 cDNA clones indicates the missing 350 bases of 3′ UTR when compared to the Spec2 sequence. pSpec2 is a 3′ UTR cDNA clone described by Bruskin et al. (1981, 1982). By sequence it is similar to Spec2a and Spec2c. The thin lines with arrows below each cDNA clone indicate regions that have been sequenced. Dashed lines indicate sequences derived from the corresponding exons of genomic clones

(dorsal) ectoderm cells and encode a family of calcium-binding proteins belonging to the troponin C superfamily (Lynn et al. 1983; Carpenter et al. 1984). We have analyzed the structure and expression of four Spec messages: Spec1, which encodes a 1.5-kb mRNA that accumulates to its maximal level between the early blastula and gastrula stages, and Spec2a, 2c, and 2d, which encode 2.2-kb mRNAs that accumulate to their maximal levels during the postgastrula stages. Hybridization and sequence analyses of the open reading frames (ORFs) from Spec cDNA clones have shown that the Spec2a and Spec2c proteins are more closely related to each other than to the Spec1 protein (Carpenter et al. 1984). We now show that the Spec2d protein is highly divergent from the other Spec proteins, and, unexpectedly, that long stretches of the 5′ and 3′ UTRs of the Spec mRNAs are far more conserved than the protein-coding regions. Our data suggest that the Spec gene family has evolved differently from most gene families, with mutations accumulating more slowly in untranslated regions than in those encoding proteins.

Materials and Methods

cDNA and Genomic Clones. The Spec1a, 1b, 2a, 2c, and 2d cDNA clones were isolated from a λgt10 library made from gastrula poly A^+ RNA (Carpenter et al. 1984; and unpublished data). The pSpec2 cDNA was isolated from a gastrula cDNA library made in pBR322 (Bruskin et al. 1981). The Spec1a and Spec1b cDNAs represent transcripts from two alleles of the Spec1 gene (Hardin et al. 1985) and encode the same protein product. A Spec2d genomic clone, X-13, was used to determine the 5′ UTR sequence and to generate the 5′ UTR probe used in Fig. 2. Two other genomic clones, Z-9 (Hardin et al. 1985) and 5A-2 (unpublished data), were used to determine the sequence of the 5′ UTRs from Spec1a and Spec2c, respectively; a third genomic clone, PK1 (Hardin et al. 1985), was used to determine a partial sequence of the Spec2c 3′ UTR. These four genomic clones were isolated from λEMBL3 libraries made from Sau III A partialed genomic DNA.

Southern Blots and Hybridization. Spec cDNAs were digested with EcoR1, electrophoresed on a 1% agarose gel, and transferred to nitrocellulose. M-13 clones containing the 5′ UTR probe (a 0.5-kb SalI–EcoR1 fragment from X-13), the 3′ UTR probe (a 0.9-kb EcoR1 fragment from the Spec2d cDNA clone), and the ORF probe (a 0.4-kb EcoR1 fragment from the Spec2d cDNA clone) were prepared as follows. "Plus" strand M-13 DNA was isolated, annealed with the hybridization probe primer (New England Biolabs), and elongated with the large fragment (Klenow) of *Escherichia coli* DNA polymerase I in the presence of [α-^{32}P]dCTP, dATP, dTTP, and dGTP for 30 min at 30°C. The labeled probes had a specific activity of 10^8 cpm/μg. Prehybridization and hybridization were performed according to Bruskin et al. (1981). The filters were washed in solutions containing 0.15 M NaCl at 68°C (Bruskin et al. 1981).

Sequencing. pUC8 subclones were made from the Spec1a, Spec2a, and Spec2c cDNA clones, and from the Spec1a genomic clone. The appropriate regions of these subclones were sequenced by the chemical-cleavage sequencing technique of Maxam and Gilbert (1980). M-13 subclones were made from the Spec2d cDNA and genomic clones, and from the Spec2c genomic clone. The appropriate regions of these subclones were sequenced by the chain termination method of Sanger et al. 1977. The sequenced

regions of the Spec cDNA and genomic clones used in this study are indicated in Fig. 1.

Sequence Comparisons. Spec sequences were aligned using the Nucaln program to maximize homology (Lipman and Wilbur, Mathematical Research Branch, NIADDK, NIH, Bldg. 31, Rm. 4B-54). The dot matrices were generated using the PCS sequence analysis program (Lagrimini et al. 1984).

Results

Hybridization of Spec cDNA Clones with the 5′ UTR, ORF, and 3′ UTR of Spec2d

The structure of the Spec genes and mRNAs provides a convenient way of determining sequence similarities among the various messages. Figure 1 portrays a Spec mRNA, its exon components, and six cDNA clones used for hybridization and sequence analyses. The first exon of all the Spec genes we have investigated is untranslated except for the final three bases, which constitute the translational initiator codon. The final exon, exon 6, always has an EcoR1 site as its 5′ boundary, followed by the final 26–29 bases of the open reading frame and 900 (Spec1) or 1250 bases (Spec2a, 2c, 2d) of untranslated sequences (Fig. 1; Hardin et al. 1985; Hardin and Klein, unpublished data). Thus, the first and last exons can be used as 5′ and 3′ UTR probes, respectively. We have previously used an exon 6 probe to demonstrate that two of our cDNA clones, Spec1a and Spec1b, share greater sequence conservation with each other than with two other cDNA clones, Spec2a or Spec2c (Carpenter et al. 1984). Subsequently we showed that Spec1a and 1b are allelic, whereas Spec2a and 2c are represented by separate genes (Hardin et al. 1985; Hardin and Klein, unpublished data).

In the course of our investigations, another Spec gene and cDNA clone, Spec2d, was isolated. As we show below, this gene is far more divergent than other Spec family members previously characterized. Hybridization reactions carried out under mild stringencies (final wash conditions 0.15 M NaCl, 68°C), using Spec2d 5′ UTR (obtained from a genomic Spec2d clone), ORF, or 3′ UTR probes (obtained from the Spec2d cDNA clone), show a peculiar pattern. The 5′ and 3′ UTR probes hybridize readily with Spec1a, 1b, 2a, and 2c (Fig. 2), but no hybridization is observed when the ORF probe is used (Fig. 2). The lack of hybridization is due to sequence differences, because the size of the hybridizing portion of the ORF probe is 420 bases (vs 150 bases for the 5′ UTR probe and 900 bases for the 3′ UTR probe) and length effects would be negligible. The absence of hybridization of the 5′ UTR probe with the Spec2d cDNA clone (Fig. 2) is because the cDNA clone is not full length and does not contain this region. The results presented in Fig. 2 imply that the Spec2d gene is more divergent in its protein-coding region than in its 5′ and 3′ UTRs when compared to Spec1a, 1b, 2a, and 2c.

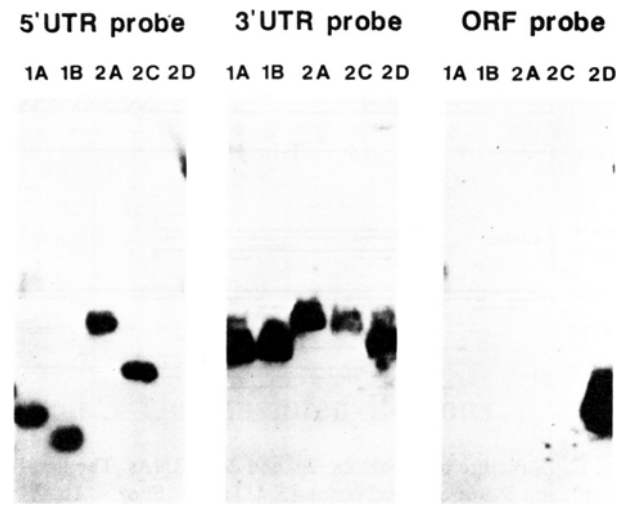

Fig. 2. Hybridization of Spec1a, 1b, 2a, 2c, and 2d cDNA clones with the 5′ UTR, 3′ UTR, and ORF regions of Spec2d. The 5′ UTR probe was a 0.5-kb Sal1–EcoR1 genomic fragment containing the entire exon 1 of Spec2d. The 3′ UTR probe was a 0.9-kb EcoR1 fragment containing the 5′-most 900 bases of exon 6 from the Spec2d cDNA clone. The ORF probe was a 0.4-kb EcoR1 fragment from the Spec2d cDNA clone containing the complete reading frame except for the first six bases at the 5′ end and the last 29 bases at the 3′ end. The fragments were inserted into M-13 and labeled with [α-^{32}P]dCTP and Klenow fragment using the reverse primer procedure described in Materials and Methods. Spec cDNA clones were digested with EcoR1, electrophoresed on a 1% agarose gel, transferred to nitrocellulose, and hybridized with the indicated probes. The Spec2d cDNA is missing exon 1 (see Fig. 1) and therefore does not hybridize with the 5′ UTR probe

Sequence Comparisons of Spec1a and Spec2d

To support the assertion that the protein-coding portions of the Spec genes are less conserved than the untranslated regions, we compared the sequences of Spec1a and Spec2d. Spec1a (and Spec1b) encode a 152-amino acid calcium-binding protein that is the predominant Spec protein in the sea urchin early pluteus larva (Carpenter et al. 1984). Although the ORF of the Spec2d message has not been identified with a specific protein in vivo, a cDNA clone with precisely the same exon sequences as the Spec2d gene has been identified, suggesting that the Spec2d ORF encodes a bona fide sea urchin protein (Hardin and Klein, unpublished data). There are several Spec proteins of lower abundance than the Spec1 protein at the pluteus stage; these are encoded by the 2.2-kb Spec2 mRNAs (Klein et al. 1983). Figure 3A and B shows that although the ORFs of Spec1a and Spec2d are clearly related, they are dis-

tinct sequences. Comparison of the nucleotide sequences of Spec1a and Spec2d shows an overall match of 62%, with large regions of divergence, including deletions and insertions throughout the ORFs (Fig. 3A). The changes occur with nearly equal frequency in all three codon positions. In one very divergent region, between bases 51 and 270, the Spec1a ORF shares only 87 bases with Spec 2d. The match appears somewhat higher at the beginning and end of the reading frame. This low overall sequence similarity between the ORFs explains the lack of hybridization seen in Fig. 2C. The insertions and deletions observed when comparing Spec2d with Spec1a result in four frame shifts with respect to the Spec2d ORF. In each case when a frame shift occurs, it is countered by a second frame shift that keeps the Spec2d ORF in the original frame. Although very minor sequencing errors may exist in these data, both strands have been sequenced for many bases. In addition, exons 3, 4, 5, and 6 of the Spec2d genomic clone have been sequenced and match precisely with the corresponding sequences in the cDNA clone. Thus, while the divergence between these sequences is unusual, we are confident that the differences are real.

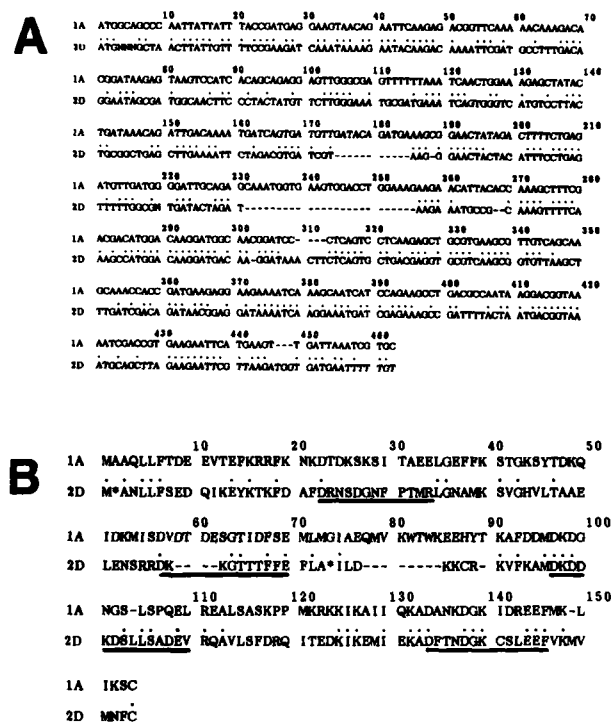

Fig. 3. Comparison of Spec1a and Spec2d open reading frames. **A** Nucleotide sequence. Dashed lines indicate bases absent from the sequence. The N at position 220 of the Spec2d sequence signifies an ambiguous base. **B** Protein sequence. The asterisks indicate residues not sequenced (position 2) or an ambiguity (position 74). The region of each calcium-binding domain that interacts with calcium ion is indicated as a solid line underneath the sequence

Comparison of the protein sequences encoded by the Spec1a and Spec2d mRNAs shows an overall similarity of only 34% (Fig. 3B). As with comparisons of other troponin C-related proteins, most of the conserved amino acid residues between Spec1a and Spec2d reside in or near the regions of the calcium-binding domains that interact with calcium ions (underlined residues in Fig. 3B), although the first calcium-binding domain does not appear to be conserved. Conservation of the other three domains implies that functional constraints are operating in at least some parts of the molecule. Because the other Spec genes we have isolated are much more closely related to Spec1a (see below), we conclude from this analysis that Spec2d must be a distantly related member of the Spec gene family.

Given the distant relationship, it would be expected that the untranslated portions of the Spec2d message would share little similarity with Spec1a. The hybridization experiments described above (Fig. 2) and the sequence comparison of the 5′ and 3′ UTRs of Spec1a and Spec2d (Figs. 4 and 5) demonstrate that this is not the case.

The 5′ UTRs are 84% matched for 136 bases immediately upstream from the AUG initiator codon and then diverge completely (Fig. 4). In the first 42 bases preceding the initiator codon, there are only three mismatched bases. This is in striking contrast to expectation, and clearly demonstrates a stronger conservation of sequence in an untranslated region than in a protein-coding region. The initiation of transcription of the Spec1a gene is 217 bases upstream from the AUG initiator codon (Hardin et al. 1985), but the Spec2d transcriptional initiation site has not yet been mapped. The site of divergence between the two sequences maps to the middle of the 5′ UTR of the Spec1a message, and it is rea-

```
                10         20         30         40         50         60         70         80
1A  ACTCTTTCTT GAATAATACT CTTCGTCAGC ACTCCACCCT TACCACTCCT TGAAAGCAAG TCGACTAGAA GAGAAGTATA
    .  . ......... ... . .. . ... ..... .  ....  .......... .. .. . .. .... .....
2D  TTATATGTCC GTATAATACT CTTTGACATC ACTGCACCCT TGTCACTTT- TGAAAGCAAG TCAACCATAA GAGAGGTATA

                90         100        110        120        130        140        150
1A  GG-----TAC CTGAGCTCCA GAGGAA---- -TTTTCTATC TTCAATTTTG CCGGATCTAT CCCAACTTCC AAAG
    ..   ...  .....  ....  ...  ..        ...  ....  ..........  ..........  ..........  ....
2D  GGTAGGCTAC CTGAGATCCA GAGAAATTTT GCTTTTTATC TTCAATTTTG CCGGATCTAT CCCAATTTCT AAAG
```

Fig. 4. Comparison of Spec1a and Spec2d 5′ UTRs. The three bases immediately after position 154 are the ATG translational initiation codon. The similarity between the two sequences begins at position 11. The Spec1 transcriptional initiation site is 217 bases from the initiation codon. The Spec2d transcriptional initiation site has not been determined

Table 1. Comparison of Spec mRNA 5′ UTRs, ORFs, and 3′ UTRs

Spec message[b]	Percent matched[a]				
			3′ UTR[d]		
	5′ UTR[c]	ORF	A	B	C
1a–2a	91 (206)	73 (449)	91 (65)		84 (200)
1a–2c	93 (115)	74 (449)	91 (65)		79 (143)
1a–2d	84 (136)	62 (419)	86 (65)	80 (179)	
1a–2p				80 (319)	
2a–2c	93 (121)	86 (449)	90 (73)		83 (145)
2a–2d	84 (140)	60 (417)	82 (73)		
2c–2d	89 (119)	60 (419)	82 (73)		
2d–2p				83 (153)	

[a] Percent matched was calculated by dividing the total number of matches by the total number of bases. Gaps of any size are counted as one mismatch and one base. Undetermined bases are not counted. The number of total bases compared is indicated in parentheses

[b] Sequences were obtained from either Spec cDNA clones or in the cases of Spec1a, 2c, and 2d, from sequences of exonic regions of Spec genes where full-length cDNA clones have not been isolated (see Fig. 1). The Spec message termed "2p" is pSpec2, a cDNA clone depicted in Fig. 1 and described by Bruskin et al. (1981, 1982)

[c] The sequence matches represent bases immediately preceding the AUG codon and continuing upstream until the point of divergence between the two sequences

[d] Three regions 3′ to the translational termination codon are compared. The region termed A begins immediately after the termination codon and extends to the upstream breakpoint of Spec1 (see Fig. 1). The region termed B begins either immediately upstream from the 3′ repeat element indicated in Fig. 1 and continues upstream (in the case of the 1a–2d or 2d–2p comparison), or begins approximately 70 bases upstream from the 3′ repeat element and continues downstream for 300 bases (including the 3′ repeat element) (in the case of the 1a–2p comparison). The region termed C is the 3′ end of the message. In the case of Spec1a and Spec2a, this includes the AAUAAA processing site. For Spec2c, the sequence begins about 25 bases upstream from this site

sonable to assume the divergence point in Spec2d is also in, or at the end of, the 5′ UTR.

The 3′ UTRs of Spec1a and Spec2d also show strong sequence conservation. All the Spec2 3′ UTRs are about 1250 bases, whereas the Spec1a 3′ UTR is about 900 bases, the length difference being accounted for by an absence of 350 bases in Spec1a, with a break point 67 bases downstream from the translational termination codon (Fig. 1). When the sequences upstream from this break point are compared, strong similarities between Spec1a and Spec2d are found beginning 2 bases and 15 bases after the Spec1a and Spec2d termination codons, respectively, and continuing until the break point (Fig. 5A). In this region there are only nine mismatches out of 66 bases. This conservation is remarkable considering there is not a single 66-base stretch in the ORFs of these two mRNAs that has fewer than 14 mismatches.

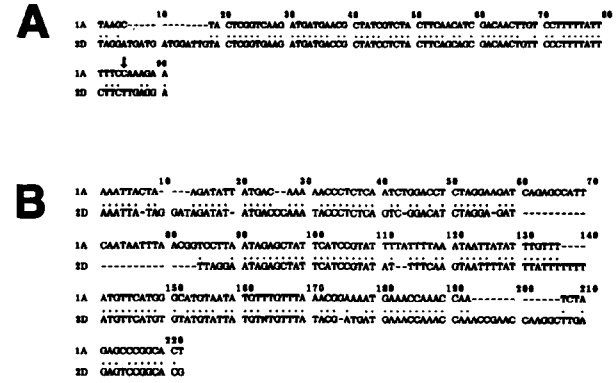

Fig. 5. Comparison of Spec1a and Spec2d 3′ UTRs. **A** The region from the translational termination codon to approximately 100 bases downstream. The deletion breakpoint in the Spec1a UTR occurs at position number 85 and is indicated by an arrow. **B** The region starting approximately 350 bases (Spec1a) or 700 bases (Spec2d) from the translational termination codon

The sequence conservation appears to extend throughout the length of the 3′ UTR. Figure 5B shows the sequence of a region that begins about 350 and 700 bases downstream from the end of translation of Spec1a and Spec2d, respectively. There is clearly strong sequence conservation even at this distance beyond the translational termination. This region is immediately upstream from a 150-base repetitive element found in the 3′ UTRs of all Spec mRNAs and also in 2000–3000 other places in the sea urchin genome (Carpenter et al. 1982). The 3′ repeat elements (their position on Spec messages is shown in Fig. 1) show a sequence conservation of 80%. Other than their association with Spec genes, these repeat elements appear to be randomly dispersed in the genome (Carpenter et al. 1982).

Comparisons of Other Spec mRNAs

The strong conservation of sequence in the UTRs of Spec1a and Spec2d is quite dramatic because of the corresponding weak conservation of the protein-coding region. This phenomenon can also be seen when Spec2d is compared with Spec2a or Spec2c (Table 1, rows 6 and 7). However, the conservation of Spec UTRs can also be seen when the other three less divergent messages are compared. Table 1 shows that among ORFs, Spec2a and 2c are closest in sequence, followed by Spec1a. When 5′ or 3′ UTRs are compared, in all cases they show at least as much and usually more similarity than the ORFs. Comparison of Spec1 and Spec2a or 2c shows that the reading frames are significantly lower in matches than the UTRs (Table 1, rows 1 and 2) and even in the Spec2a–Spec2c comparison, where the proteins show very strong conservation, stronger conservation is seen outside the protein-coding regions (Table 1, row 5). That the UTR sequence similarity

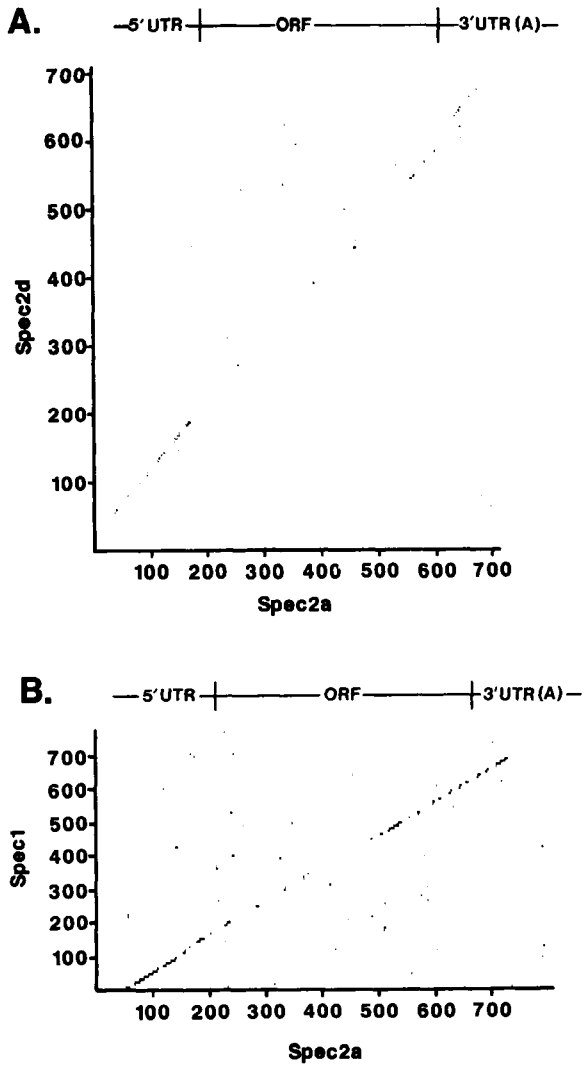

Fig. 6. Dot matrix comparison of Spec mRNA sequences. The sequences are computer scanned base by base and seven identical bases in a row is scored as a dot. **A** Comparison of Spec2a vs Spec2d; **B** comparison of Spec1a vs Spec2a

extends to the very end of the Spec mRNAs can be seen when Spec1a–Spec2a, Spec1a–Spec2c, or Spec2a–Spec2c are compared (Table 1).

Graphic displays of sequence comparison between Spec2a and Spec2d and between Spec1a and Spec2a are shown in Fig. 6A and B, respectively. This is a fairly stringent comparison where homology is scored only if there are seven matched bases in a row. In both cases the conservation of 5′ and 3′ UTRs with respect to the ORFs is readily apparent. Note that the analysis extends only about 90 bases into the 3′ UTRs. Beyond this point, Spec2a and Spec2d remain matched, but Spec1a does not, due to the break in sequence discussed above. Another feature emerging from this analysis is that the last one-third to one-fourth of the ORF is more conserved than the remaining upstream ORF. Close inspection of the Spec1a–Spec2d amino acid sequence comparison (Fig. 3B) shows that the C-terminal portion of the Spec proteins is more conserved than the remainder of the sequences.

Discussion

Our results show an unusual sequence conservation in long stretches of the untranslated regions of the Spec gene family. It is likely that conservation of the Spec coding regions is due to functional constraints, but it is more difficult to interpret conservation in noncoding regions. The Spec proteins accumulate during embryonic development specifically in the aboral (dorsal) ectoderm. Although their exact function is unknown, by analogy to other members of the superfamily it seems reasonable to assume their function involves binding calcium ions, thereby mediating a specific cellular process or event (Bruskin et al. 1982; Lynn et al. 1983; Carpenter et al. 1984; Klein et al. 1985). The open reading frames of Spec1a and Spec2d show greater conservation of amino acid sequence in the calcium-binding domains and the C-terminal regions than in other portions of the molecule. Levels of sequence conservation at least as great as those in the ORF are seen over long stretches of the untranslated regions of the message. The overall divergence in the Spec1a–Spec2d ORFs is 38%. Given that accumulation of each percent divergence for freely mutating regions in sea urchin genomes corresponds to approximately 1.4 million years in time (Britten 1986), the minimum time these genes have been evolving from a common ancestor is about 50 million years. Since it is unlikely the sequences are freely mutating, this is probably an underestimate. This evolutionary time predates the origin of the family Strongylocentrotidae, and suggests that the Spec genes are very ancient. Conservation of sequences tens or hundreds of base pairs upstream and downstream from the protein-coding regions implies that strong functional constraints are operating on these regions.

Most reported cases of conservation in 5′ and 3′ UTRs involve very small regions of fewer than a dozen bases (Birchmeier et al. 1984; McGarry and Lindquist 1985; Morris et al. 1986). These presumably represent various translational and processing signals. There are instances where longer stretches of sequence are conserved. Several actin gene isotypes are unusually well conserved in their 3′ UTRs (Ponte et al. 1983). The various actin isotypes (α, β, γ) of humans and mice have very divergent 3′ UTRs, but, for example, the 3′ UTR of a γ gene from humans is homologous to the corresponding 3′ UTR of a γ gene in mice (Ponte et al. 1983). The 3′ UTRs of some non-cell cycle-regulated histone genes have also been conserved. A histone H3 vari-

ant in humans has 527 bases of 3′ UTR that is 85% matched to the corresponding chicken variant. This is a far greater similarity than would be expected given the evolutionary distance between these two species. In this case, the 3′ UTR has been implicated in conferring stability on the mRNA in the non-S phase of the cell cycle (Wells, personal communication). The 5′ UTR of the human H3 mRNA has also been implicated in mRNA stability (Morris et al. 1986).

The Spec gene family shows quite different properties from previously reported gene families. First, large stretches in both 5′ and 3′ UTRs are conserved, and it is likely from our results, although not proven, that the conservation extends over most of the length of the UTR. Second, unlike the human–mouse actin gene conservation cited above, all the members of the Spec family we have thus far investigated show UTR conservation within the species. There is no detectable conservation of introns among the Spec genes (Hardin et al., unpublished data), and it is therefore unlikely that the conservation of UTRs is maintained by any standard correction mechanism involving gene conversion over the entire gene length. It is possible, however, that the 5′ and 3′ UTRs could be corrected without affecting the introns or ORFs. Spec1 and Spec2c are closely linked genes—only 7 kb of DNA separate them (Hardin et al. 1985)—whereas Spec2a and Spec2d do not appear to be tightly linked to other Spec genes (i.e., within 10 kb). If gene conversion mechanisms were operating on the Spec genes, it might be expected that the Spec1 and Spec2c UTRs would maintain greater sequence similarities with each other than with other Spec UTRs. As can be seen in Table 1, this is not the case. We favor the hypothesis that the Spec gene family has been diverging in a non-uniform manner, with mutations accumulating most rapidly in the introns, less rapidly in the ORF regions, and least rapidly in 5′ and 3′ UTRs. As would be expected from a gene family evolving from a common ancestor, the relationships derived from comparisons of the ORF sequences and the UTR sequences are the same; Spec2a and Spec2c are closely related, Spec1 is less related, and Spec2d is least related.

The polymorphic nature of the Spec1 mRNA provides additional insight into UTR conservation. Three alleles have been characterized for Spec1: Spec1a, Spec1b, and Spec1c (Carpenter et al. 1984; Hardin et al. 1985). These alleles have virtually identical ORFs (Carpenter et al. 1984). A comparison of the 3′ UTRs of Spec1a and 1b shows that the first 100 bases or so after the termination codon are conserved as much as the ORF, but that the remainder of the 3′ UTRs are approximately 90% matched (Carpenter et al. 1984). There is less similarity between the Spec1a and Spec1c 3′ UTRs (Carpenter et al. 1984). Thus, in the case of the Spec1 alleles, the protein-coding region is more conserved than most of the 3′ UTR. Presumably the function of the Spec1 protein does not allow for appreciable variation. The amount of divergence in the 3′ UTRs of the Spec1 alleles is about the same as or perhaps slightly higher than with other Spec genes (Spec2a, 2c, 2d), and suggests that while these sequences show strong conservation, some variation can and does occur. This variation is consistent with the idea that the untranslated regions may play a gross structural role that could include interactions with other cellular components, for example, the cytoskeleton or nuclear matrix. This could be the basis of selective message transport, localization, or stability.

Acknowledgments. We thank Clifford Carpenter and Susan Hardin for sequencing some of the Spec genomic and cDNA clones. We also thank Bruce Brandhorst, Susan Hardin, Matthew Keast, Elizabeth Eldon, Craig Tomlinson, and Alice Rubacha for comments on the manuscript and Kathy Oleszkowicz and Mike Perry for technical assistance. This work was supported by NIH grant HD 22619 to W.H.K. W.H.K. is the recipient of an NIH Research Career Development Award, HD 735.

References

Birchmeier C, Schumperli C, Sconzo G, Birnstiel ML (1984) 3′ editing of mRNAs: sequences, requirements and involvement of a 60-nucleotide RNA in maturation of histone mRNA precursors. Proc Natl Acad Sci USA 81:1057–1061

Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1393–1398

Bruskin AM, Tyner AL, Wells DE, Showman RM, Klein WH (1981) Accumulation in embryogenesis of five mRNAs enriched in the ectoderm of the sea urchin pluteus. Dev Biol 87:308–318

Bruskin AM, Bedard PA, Tyner AL, Showman RM, Brandhorst BP, Klein WH (1982) A family of proteins accumulating in ectoderm of sea urchin embryos specified by two related cDNA clones. Dev Biol 91:317–324

Carpenter CD, Bruskin AM, Spain LM, Eldon ED, Klein WH (1982) The 3′ untranslated regions of two related mRNAs contain an element highly repeated in the sea urchin genome. Nucleic Acids Res 10:7829–7842

Carpenter CD, Bruskin AM, Hardin PE, Keast MJ, Anstrom J, Tyner AL, Brandhorst BP, Klein WH (1984) Novel proteins belonging to the troponin C superfamily are encoded by a set of mRNAs in sea urchin embryos. Cell 35:663–671

Cox KH, Angerer LM, Lee JJ, Davidson EH, Angerer RC (1986) Cell lineage-specific programs of expression of multiple actin genes during sea urchin embryogenesis. J Mol Biol 188:159–172

Fyrberg EA, Mahaffey JW, Bond BJ, Davidson N (1983) Transcripts of the six *Drosophila* actin genes accumulate in a stage- and tissue-specific manner. Cell 33:115–123

Hardin SH, Carpenter CD, Hardin PE, Bruskin AM, Klein WH (1985) The structure of the Spec1 gene encoding a major calcium-binding protein in the embryonic ectoderm of the sea urchin, *Strongylocentrotus purpuratus*. J Mol Biol 186: 243–255

Klein WH, Spain LM, Tyner AL, Anstrom J, Showman RM, Carpenter CD, Eldon ED, Bruskin AM (1983) A family of mRNAs expressed in the dorsal ectoderm of sea urchin embryos. In: Malacinski GM, Klein WH (eds) Molecular aspects of early development. Plenum Press, Inc., New York, pp 131–140

Klein WH, Carpenter CD, Philpotts LE, Brandhorst BP (1985) The sea urchin Spec family of calcium binding proteins: characterization and consideration of possible role in larval development. In: Sawyer RH, Showman RM (eds) The cellular and molecular biology of invertebrate development, number 15. University of South Carolina Press, Columbia, pp 275–295

Lagrimini LM, Brentano ST, Donelson JE (1984) A DNA sequence analysis package for the IBM personal computer. Nucleic Acids Res 12:605–614

Lopata MA, Havercroft JC, Chow LT, Cleveland DW (1983) Four unique genes required for β tubulin expression in vertebrates. Cell 31:713–724

Lynn DA, Angerer LM, Bruskin AM, Klein WH, Angerer RC (1983) Localization of a family of mRNAs in a single cell type and its precursors in sea urchin embryos. Proc Natl Acad Sci USA 80:2656–2660

Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol 65: 499–560

McGarry TL, Lindquist S (1985) The preferential translation of *Drosophila* hsp 70 mRNA requires sequences in the untranslated leader. Cell 42:903–911

Morris J, Maraski F, Weber L, Hickey E, Greenspan D, Bonner J, Stein J, Stein G (1986) Involvement of the 5′ leader sequence in coupling the stability of a human H3 histone mRNA with DNA replication. Proc Natl Acad Sci USA 83:981–985

Ponte P, Gunning P, Blau H, Kedes L (1983) Human actin genes are single copy for α-skeletal and α-cardiac actin but multicopy for β- and γ-cytoskeletal genes: 3′ untranslated regions are isotype specific but are conserved in evolution. Mol Cell Biol 3:1783–1791

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467

Shott RJ, Lee JJ, Britten RJ, Davidson EH (1984) Differential expression of the actin gene family of *Strongylocentrotus purpuratus*. Dev Biol 101:295–306

Thireos G, Penn MD, Greer H (1984) 5′ untranslated sequences are required for the translational control of a yeast regulatory gene. Proc Natl Acad Sci USA 81:5096–5100

Received October 1, 1986/Revised November 24, 1986

Note added in proof: The protein sequence presented in Fig. 3B for Spec 2d has four residues mistakenly translated from the nucleotide sequence. Position 34 should be F rather than R; position 58, R rather than K; position 69, P rather than F; and position 103, K rather than S. In addition, a dot signifying a match of residues between Spec 1a and Spec 2d was omitted at position 40. These changes do not alter the conclusions drawn in the text.