A single gene encoding vitellogenin in the sea urchin *Strongylocentrotus purpuratus*: sequence at the 5' end

Ann-Bin Shyu⁺, Thomas Blumenthal and Rudolf A. Raff*

Institute for Molecular and Cellular Biology and Department of Biology, Indiana University, Bloomington, IN 47405, USA

Received August 17, 1987; Revised and Accepted November 10, 1987

Accession no. Y00513

ABSTRACT

The synthesis of vitellogenin (yolk protein precursor) in the sea urchin, *Strongylocentrotus purpuratus*, is unique in that both males and females produce a high level of the protein. In this paper we show that this organism also is unique in possessing only a single vitellogenin gene. Like the genes that encode analogous proteins in vertebrates, the sea urchin gene is large, about 19 kb in length. The sequence surrounding the 5' end of the gene revealed several other similarities to vertebrate vitellogenin genes: the signal sequence is exceptionally short and has a sequence similar to those from frog and chick; there is a canonical TATA box at -32; and there is a sequence closely resembling the estrogen-responsive element at -207.

INTRODUCTION

Small gene families encoding vitellogenin (yolk protein precursors) have been identified and characterized from *Drosophila*, locust, chickens, frogs and nematodes. The numbers of genes in each family range from two (*Locusta migratoria*) to six (*Caenorhabditis elegans*) (1-6). Vitellogenins and the genes that encode them appear to fall into two groups. The first group, found in *Drosophila*, comprises relatively small yolk proteins of about 45 kd, encoded by genes of about 1.5 kb in length (7). These genes do not appear to be related at the nucleotide sequence level to the members of the second group. This latter group is found in a surprising variety of organisms, from nematodes to chickens. These vitellogenins are large, between 185 and 220 kd (8-13). In spite of the fact that the nematode and vertebrate vitellogenins are quite clearly ancestrally related (14), the genes that encode them are structurally quite different. Whereas the nematode genes are only about 5 kb in length, with four very short introns (15; Nettleton, Spieth and Blumenthal, unpublished observations), the vertebrate genes are 16 to 20 kb long and with 34 introns each (2,16). On the other hand both the nematode and vertebrate genes produce transcripts characterized by having short 5' untranslated regions followed by unusually short and highly similar signal sequence coding regions (14,17).

The study of vitellogenin genes is important for the insights they

can provide into mechanisms of hormone regulation of gene activity during development. The Drosophila genes are expressed exclusively in the fat body and ovary of the adult female; the vertebrate genes in the adult female liver; and the nematode genes in the intestine of the adult hermaphrodite (18-20). The vitellogenin genes of Drosophila and vertebrate have been shown to be inducible by steroid-hormone (ecdysone and estrogen, respectively), and recent experiments have provided evidence for cis-acting sequences involved in the regulation (20-22). In the case of Drosophila, it has been demonstrated that the fat body and ovary-specific transcription is controlled by two separable elements, but the particular sites at which ecdysone receptor or other regulatory proteins act have not yet been identified.

Two heptameric sequences (Box 1 and Box 2) are present repeatedly upstream of the nematode genes (14). One of these, Box 2, is also present in the vertebrate promoters (where it has been called the "elegans box") and it has been implicated in estrogen-dependent transcription (23). In addition the site at which the estrogen receptor interacts with both frog and chicken promoters has been identified (23-25). This site is separable from, but sometimes adjacent to, the "elegans box".

We recently have reported the isolation of a cDNA clone for sea urchin vitellogenin and demonstrated that unlike all other species analyzed, the sea urchin gene is expressed in both sexes (13). The sites of expression are the adult intestine and gonad.

Here we show that, also in contrast to all previously described cases, sea urchin Strongylocentrotus purpuratus vitellogenin is encoded by a single gene. This gene is similar to those that encode vitellogenin in vertebrates. We also present sequence data that raises the possibility that the sea urchin vitellogenin gene is evolutionarily related to the vertebrate and nematode genes and that it may share regulatory elements with these distantly-related vitellogenin genes.

RESULTS

Characterization of a partial vitellogenin cDNA clone

We previously showed that a lambda gt10 cDNA clone derived from a sea urchin intestinal cDNA library and designated clone 4, is a vitellogenin cDNA clone (13). The insert of clone 4 is approximately 980 bp in length. In order to determine whether this clone contains coding sequence, we sequenced the 5' and 3' ends of the cDNA insert. The 5'/3' orientation of clone 4 was determined by probing a RNA gel blot with M13 subclones from the clone 4 cDNA insert (13). In the 340 base sequence of the 5' end, we identified an open reading frame which persists for 240 bases. Immediately after the stop codon, the base composition of the remaining 100 bases becomes highly AT rich, suggesting that the open reading frame is the real coding region. In the 303 base sequence of the

Figure 1.

Structure of the cDNA clone. The top line indicates the amount of DNA sequenced from each end of the cDNA insert. Restriction fragments were subcloned into M13, mp18 or mp19 (32) for sequencing by the dideoxy chain termination method (33,34) as modified by Biggin et al. (35). The stippled box indicates the location of the open reading frame and the open box the 3' untranslated DNA. The AATAAA, presumably signalling the site of 3' end processing is also shown. Base composition is plotted below the map. All sequence analysis was performed using software developed by Pustell for IBI Inc.

3' end, we found an AATAAA polyadenylation signal which is 35 bases upstream from a short poly(A) sequence, and the whole region sequenced is highly AT rich (Fig. 1). Therefore, clone 4 most probably contains 240 bases of coding DNA and the DNA specifying an approximately 740 base 3' untranslated region.

A single gene encodes vitellogenin

In order to determine whether sea urchin vitellogenin is encoded by a gene family, as in previously characterized instances, we performed genomic DNA blot analysis. DNA preparations purified from sperm of five different individuals (a gift of William H. Klein), were digested with restriction enzymes, blotted, and probed with clone 4 cDNA. A simple hybridization pattern was detected (Fig. 2). When genomic DNA was cut with EcoRI or XbaI, only one or two bands were detected. There appears to be population polymorphism associated with one of the two bands but

Figure 2. Hybridization of the cDNA clone to genomic DNA. DNA preparations from the sperm of 5 male sea urchins (labeled a-e) were digested with the enzymes shown, electrophoresed on a 1% agarose gel, blotted, and probed with the insert fragment from clone 4, nick-translated (36). The bands are marked by arrows. Bands seen in all 5 animals are indicated with numbers (in kb).

Figure 3. Restriction map of the vitellogenin gene. Individual clones, the extents of which are shown by horizontal lines, were mapped, and the composite map prepared on the basis of overlaps. Restriction sites, and an insert, found on only one or two of the clones are indicated on the clones on which they were found. Enzymes used were: ▽, BamHI; ■, EcoRI; ●, HindIII; ▼, Kpn I; □, Sal I; ★, Xba I; ○, Xho I. The open box at the 3' end corresponds to the clone 4 cDNA. The solid box in GC1,5 represents a 300 bp insert in the Kpn I/HindIII fragment.

not with the other (4.0 kb band in EcoRI and 6.0 kb in XbaI digestions). Since the probe has no internal EcoRI or XbaI site, the simple pattern shown by the genomic Southern blot analysis suggests that the Vg gene is probably a single gene with two alleles present in each individual. The observed polymorphism is most likely due to polymorphic sites in the 3' flanking region of the Vg gene. The other site in each digest resides within the gene and is conserved among individuals examined (see below).

The result from the HindIII digestion supports this interpretation (Fig. 2). Since the clone 4 cDNA insert has an internal HindIII site, cleavage of the genomic DNA with HindIII yielded two or three bands that hybridized to the probe. One band, which has a size of 1.2 kb, is common to all individuals. In addition, wherever there are three bands detected, the 1.2 kb band shows more intense hybridization than the other two bands, which have about equal intensity to each other. The 1.2 kb band probably resides entirely within the gene and is conserved in all alleles examined. The cleavage of the two alleles of any individual with HindIII will yield two copies of the 1.2 kb HindIII band, and only one copy each of the two other bands containing allelic 3' flanking regions. Thus, the observed differences in hybridization intensity and restriction fragment length polymorphism are easily explained by the presence of only

a single gene. There appears to be considerable restriction site polymorphism in the 3' flanking region of the Vg gene, but much less within the gene itself.

Isolation of genomic clones

A sea urchin genomic library (in EMBL 3, a gift of Paul Hardin and William Klein) was screened with nick-translated clone 4. The library was constructed using sperm DNA from a single animal (animal "a" of Fig. 2). Six clones were isolated, GC1 through GC6. Fragments containing transcribed regions were identified by DNA blot analysis using random-primed cDNA to poly(A)-containing intestinal RNA as probe. The six clones all share extensive restriction site homology with each other, and can be aligned as overlapping fragments from a single genomic region (Fig. 3).

Clones GC4 and GC3 have nearly identical restriction maps, with the exception of an EcoRI site at the right-most end of clone GC3, and a KpnI site present in clone GC4 but not clone GC3. When clone GC1 and clone GC3 maps are compared, three restriction site differences (a KpnI, an XhoI, and a BamHI site in clone GC1, missing from the others) and a restriction fragment length polymorphism are seen. The extra BamHI site of clone GC1 is in a 300 bp insert into the KpnI/HindIII fragment (Fig. 3, solid bar). The region that specifies the clone 4 cDNA sequence, ie., the 3' end of the gene, and 5'/3' orientation of the gene on the genomic clones were determined by DNA blot hybridization and DNA sequencing (data not shown).

The genomic fragments are predicted by the GC1 and GC3 maps

Since the genomic library from which GC1 to GC6 were isolated was constructed using sperm DNA from animal "a" of Fig. 2, the result from genomic DNA blot analysis of animal "a" was compared to the restriction maps of Fig. 3. The EcoRI, HindIII and XbaI fragments of animal "a" all have the sizes predicted by the assumption of two alleles represented by the clone GC1 and clones GC4 and GC3 maps. A single 4.0 kb EcoRI fragment is identical in the two maps; a 5.7 and a 6.0 kb XbaI band are predicted by the clones GC4 and GC3 and clone GC1 maps, respectively; and the two HindIII fragments, of 1.2 and 4.3 kb, are present in both sets of clones. The restriction fragments which correspond to those in the restriction maps of the cloned DNA, corresponded only to the restriction fragments which were hybridized to the probe at moderate stringency, indicating that there are no other genes closely related to the cloned vitellogenin gene.

The genomic DNA blots indicated the presence of only a single Vg gene. Therefore, we would have predicted only two classes of genomic clones, representing the two alleles, since the library was made from sperm DNA of a single animal. Since clone GC4 shares most of the restriction sites, even in the extensive 3' flanking region, with clone GC3, it appears likely that GC4 represents some sort of cloning artifact derived from the clone GC3 allele. This interpretation is supported by

gatttgattataatcatgggtgttttttttaaatcaaaactacgtgaac -497
 aataatgtacgtttaataatcaaggatagataaaggtaaccrcaanaataatctatcacctgcacccctcatattaagctctactgagaatcatggttcagaatccttcggtacatcgtttc -378
 caaaaaaacctgaaaaatatacaaaaaatacaaaaagaagaanaaataaactttgacaaaaacataatcaggaacccatggtcaaaaaacatcaatcttctgccaacagtacca -259
 aacttatccaccacagaccacccattctctttatcgaaactctctactgacgggtcaanttgacccctgactccaccccgccctaccaagatccaaccaatcagaacggggcagggtgtgga -140
 aacgtgacgatcgaaactttgagcaaatcgacacactggatgggagggcctggcggttctctgctgacgaatattgtgtggaanaatgaagataatcaaaagccatataaagcccc -21
 acatcttggataaattgttcATCTTCGCGCTTCAGACAGGCTTTGGCAGTGTCTGATTTGAGCATTTCBAGGACTTTGTTGCCGATTTTCATCAACTCAGCTCGAGGAAGTTGTATC -99
 ATG AGG GCT GCA ATT CTC TTT TGC CTG GTG GCC AGT TCT ATG GCC GTG CCA AGC GGC TCG TTG G gtaagttacatttaactcttctgtagtattc +194
 M R A A I L F C L V A S S H A V P S S L
 tacaccacaacttaaatgatcgtttgggtttctatggattctatggactgtattagaacttaattgcytcatatgctctgatatcagtgaaactgtgcaaatcataccacaagcgttc +313
 agttaataaggggataaatttgcattgttcagaaatcagatctaacagctttgtctcttacttagatgttttgatataatcatcttaattataatattacacctccagagagaca +432
 gatatacaatcttttagggagtagatctaaatgattttaattctatattgtttatctctctgtatttttaattattttaaatcaactttctaagaaaactgtatctagaaggtctaggc +551
 tattttgataggaaagtagagagtagcaaaattttaaacagaccccatgaggcgacgaacccgcgaatcataaagaagctaaagtgtaggctgctgcgcgatgcgctgcaataaacaggtt +670
 tgattttcaatgttctgaacgttttaacgtattgtatgggtattcaatgatcacataaattttgctaattgattcaacgtatgtaaaactctctatggatgtaggaaatgtgtgtattt +789
 atgtttttgtccatgccatgacctggctgtgtgtatgaacactgattgattgatgcccccaccactgaatttcgtgctgggggtttgttgtaatttatgttactgtttctgtgtgtctt +908
 tgtcttctgaacaaatcatagaccattcttaattttaaaccttagattttctgtatcatatacagcttagaatacagttgtgattgtagttcatatatttcttaantattgtatt +1027
 ttaaggaaatataaattgataaaggtaaaatcatataaattacaaatgagcaatcccgccagggcgtcaaaaactgaaactctacaaatcctttaaactatgcctccaaaagaaccatgc +1146
 atcaattttaaaatcagatctgtaggaagcgacacacactctagacccttagtgccttcaagaacgtctaaataccattctgagattctgagattctattctagaacgtcagtagaa +1265
 actagatctggatctataactttgaaatatagaatgcaggccactttttattgattcatcttaagtgtaaaaacgttgaaacttcaaatccataactctctctcttcttctgtctctt +1384
 gactaaataacatggggcgaagagtggttatagtgacttttcggtttaatacagttgaatcactacagtgctgactaagaagaattgacttagggcgcaagaagaatagttgcaataatt +1503
 gctgaattatctgaattatagctaaactactaactgagctgattgatgcaccccccacaaattagaatgtttacatagggctcgtgaaagaagactttatagagcaaat +1622
 (I)
 tgcaataaaccttataaattcattctctttatttaataatcag 6C TCG AGG CCC GGA ACT TGC CCA CCT CAA CCC BAT AGT CAG CTC ATG ATT BAA GCA +1721
 G S R P G T C C P P Q P S D Q V M I E A
 (II)
 ACA CAG TGC TCT TAG TGS TAT GGA TTA ACC TGG GAC TGG AAC TGC AAC AGC CAA GGA CAG BAA AAC TAC AAG TGC TGC CAA TAT GAG AAC +1811
 T R C S Y V Y G L T W M D W N C N S Q G Q E N Y K C C Q Y E N
 GAC ATC AGG ATC TGT GTC CCT CCA ATC CCA GCT G gtaagaataacttttaaaatatacaggaanaatgcaaggtatgcttttaagatgtaaaattgaaagtctctct +1916
 D I R I C V P P I P A
 tcttctatttgatttaatttaaaacagcaatctccgatctctctgatatccctgtatgaacatccacagtgaaacattaaaagaganaatgtgaacattcccatgtatgccttattgttg +2035
 cctatgccttaaccatctctctctgtatgaatgaactcagtgcaacataatgtcaacataaattgtgaactcatccagtttactgttagatgaagattgaaagttgcaaatgtgcta +2154
 ttgggaaactgagatttttactgttcaactggttgcacatatacagatcaaaagtgtcttattgttgccttttaccataaattccacacaacaaagtttttttgatgtttttatttcg +2273
 agtgttaactcttgcacaaactgaacaaaattatagcatttttctaccttttgcattgtaaaagcacacactattcacattgtaatttaactgtaatacattttctatcatattttcttta +2392
 gttttaaattcattgattttgtttgtttgtaagcgcttttgcagtgcttaaatgattggcgctataaaagaacaaattattattattatgattgaacccctattatcacccaag +2511
 (III)
 AT GTT GAT GTG GAG GTC GGT GTG GAG CAA CCA AGC CAG AGC GTG ACC AAG TCC GCC AGG CCA TCC AGA AGA CCC AGG ACT TCA TCC GCA +2600
 D V D V E V G V E Q P S Q S V T K S A R P S R R P R T S S A
 AGG TGG GTC TGT ACC CCG CCC CCG ATC AGC GCC GCC GCA CCA CCC CCA CCC CCG ACA CAG TCA GAT GGT GCG TCT CGA G +2679
 R M V C T P P P I S A A A P P P P P T Q S D G A S R

Figure 4.

Sequence of the 5' end region of the vitellogenin gene. The sequence of the 5'-most two fragments shown in the map of Fig. 4 was determined by the "dideoxy" method. Universal sequencing primers were used except in cases where suitable restriction sites did not exist to allow overlapping coverage of inserts. In such cases custom synthesized oligonucleotides were used to prime from within the insert. Oligonucleotides were synthesized using an Applied Biosystems 380 DNA Synthesizer, following the instructions of the manufacturer. The primers were further purified by polyacrylamide gel electrophoresis. Exons are shown in upper-case letters. The TATA box is boxed; the sites to which the primers hybridized are thick-underlined; two potential CAAT-box like sequences and the region homologous to the estrogen receptor binding site are also indicated by underlining.

an experiment in which clone 4 cDNA was hybridized to a blot containing DNA of animal "a", the animal from which the genomic library was made, digested with BamHI and XhoI (data not shown). Two bands were detected; one corresponding to the 13.5 kb fragment of clones GC3 or GC4, and another band of approximately 20.0 kb, presumably from the clone GC1

Figure 5.

Localization of the 5' end of the vitellogenin gene by primer extension. Three oligonucleotides (I, II, and III) capable of annealing to the three positions shown in Fig. 4 were synthesized and used to prime RNA-directed DNA synthesis according to the method of Singh et al. (37). Poly(A) containing female intestinal RNA was used as template. Markers are pBR322 cut with *Sau* 3A.

allele (see map). Since only two bands were detected, we conclude that clone GC4 is most likely artifactual.

Cloning and identification of the 5' end

We used the left-most fragment of clone GC3 to isolate genomic clones containing further upstream sequences. Two clones, GC12 and GC13, did not extend farther in the 5' direction (Fig. 3). However one, GC8, does carry an insert of 15 kb of which only about 8.5 kb overlaps clone GC3. Digestion of GC8 with *Sal*I and *Xho*I cleaved the region of 5' extension into four fragments, each of which was used to probe a blot containing intestinal RNA. Each probe detected a single band corresponding in size to the 5.1 kb Vg message, indicating that all, or almost all, of GC8 contains Vg-coding sequences (data not shown). The left-most 2.2 kb *Sal*I/*Xho*I fragment was hybridized to an RNA blot

containing RNAs from different tissues and embryos. The result (not shown) showed that only a single band of 5.1 kb was detected in the intestine of both sexes. Since Vg message is very abundant in the intestine, the observed weak hybridization suggested that this left-most fragment must contain only a small amount of exon.

To identify the 5' exon or exons, we determined the DNA sequence of the left-most 2.2 kb SalI-XhoI fragment and the 1.1 kb Xho fragment immediately 3' of the SalI/XhoI fragment. In this entire region we observed only two open reading frames of significant length (Fig. 4). The first of 180 bp, is at the extreme 5' end of the 1.1 kb fragment, while the second, of at least 168 bp, begins at the extreme 3' end of this fragment and extends in the 3' direction beyond the region analyzed. Both open reading frames are bracketed by consensus intron/exon boundaries, indicating that they might be coding exons of the Vg gene. Based on the sequences of these two open reading frames, three complementary synthetic 16-base oligomers were made, (primers I, II, III, Fig. 4, underlined regions). All three hybridized to 5.1 kb mRNA in the female intestinal RNA preparation. Therefore, they are all complementary to Vg mRNA, and presumably the two putative reading frames are Vg coding regions.

Based on the above analysis, we carried out primer extension experiments. The three primers were hybridized with intestinal poly(A)⁺ RNA and extended with reverse transcriptase. Each yielded single extended fragments (Fig. 5). The size difference between the primer I-extended and primer II-extended products coincided with the distance between the sites to which they hybridize. The relative size differences between the extended products of primers I and III and between those of primers II and III are also consistent with the distance deduced from the DNA sequence, assuming the intron-exon boundaries shown in Fig. 4. The 5' end of Vg mRNA is approximately 191 bases upstream from the primer I site. In order to determine the precise location of the 5' end, we performed primer extension, accompanied by dideoxy sequencing, using primer I. Each reaction yielded the same 191 base extended product beyond which no other extended products were observed (data not shown). By comparison of the sequence from primer extension with the sequence from DNA sequencing, the 5' end location shown in Fig. 4 was determined. There is a canonical TATA box 32 base pairs upstream from the transcription initiation site. The splice sites defined by the cDNA sequencing experiment defined above match canonical splice site sequences and create a single open reading frame beginning at nucleotide +100 of exon 1 (the first AUG in the mRNA) and continuing through exons 2 and 3.

DISCUSSION

S. purpuratus contains only a single vitellogenin gene

The experiment of Fig. 1 provides evidence that S. purpuratus contains only a single gene encoding vitellogenin. Recently, Hardin et

Estrogen Receptor Binding Site			
GGTCACCAATGACC	-333	Xenopus	A1
GGTCACAGTGACC	-330		A2
<u>A</u> GTCACTGTGACC	-332		B1
<u>A</u> GTCACTTTGACC	-333		B2
GGTCAACAT <u>A</u> ACC	-346	Chick	
GGTCAGCGTGACC	-625		
GGTCAGACTGACT <u>T</u>	-166	VLDL II	
GGTCAAT <u>_</u> TGAT <u>C</u>	-207	Sea Urchin	

Figure 6.
Comparison of the sea urchin palindromic sequence with estrogen receptor binding sites. The estrogen receptor binding sites are from Walker et al. (38).

al. (26) reported that the SpecI gene of S. purpuratus, which encodes a major calcium-binding protein in the embryonic ectoderm is a single polymorphic gene. They showed that different alleles of the SpecI gene are distinct from each other in their 3' untranslated regions, demonstrating that sea urchin genes can be highly polymorphic not only in the flanking sequences but also in the non-essential regions of the gene. These observations suggest that the sea urchin Vg clones shown in Fig. 3, which have nearly identical restriction maps even in the 3' flanking region represent the same allele. The very simple patterns from all three digestions of five different individual animals examined indicate that Vg is encoded by a single gene with two alleles present in any individual animal. Thus, the single gene hypothesis not only reflects the polymorphic feature of the 3' flanking region of sea urchin genes but also is consistent with the observed simple pattern revealed by genomic Southern analysis. Finally, clone GC4, which was isolated only once, differs from GC2 and GC3 at only a single site. This clone is most likely artifactual, since the novel KpnI site is not present in the genome of the animal from which the library was prepared (unpublished observation).

The S. purpuratus Vg gene is about 19 kb in length. It appears to represent a structure containing multiple short exons separated by long introns, since cDNA primed with random sequence oligonucleotides hybridized to almost every restriction fragment of the gene (data not shown). The relatively long, AT-rich 3' untranslated region is typical of sea urchin genes.

The 5' end of the vitellogenin gene

We determined the sequence of approximately 3.2 kb of DNA containing the 5' end of the Vg gene. The first three exons are 163, 180, and at least 168 bp in length. The first two introns (1502 and 666 bp long)

contain canonical splice sites at their 5' and 3' boundaries. The location of the 5' end of the mRNA was determined to be at an A residue surrounded by pyrimidines. In the 546 bp upstream sequence, a canonical TATA box is present 32 bp upstream from the transcription start site and a CCAAT-like sequence is at position -66. In addition there is a consensus CCAAT box at -165 bp.

Comparison of the 5' flanking region of the four *Xenopus* Vg genes, one chicken Vg gene, and the chicken VLDL II gene revealed a conserved palindromic sequence, GGTCANNNTGACC, which is present in one to three copies upstream of each gene (23,24,25). Recently, by using a vitellogenin-CAT hybrid gene in transient transfection experiments, this palindromic sequence was found to be present in a 35 bp estrogen-responsive element (ERE) as an essential element for estrogen control (23,24). Sea urchins and other echinoderms produce estrogen, and estrogen appears to affect oogenesis in starfish (27,28). Interestingly, at position -207 of the sea urchin Vg gene, we found an ERE-like palindromic sequence, GGTC AATTGATC (Fig. 6). Instead of having three bases between the two 5 bp inverted repeats, it contains only two bases. Since this 12 base ERE-like sequence is not a perfect match to the ERE essential element of vertebrates and since the Vg gene in sea urchin is not known to be estrogen-regulated, we do not know if and how this element might be involved in the regulation of sea urchin vitellogenin gene expression. This is an especially intriguing question, because expression of the sea urchin Vg gene is not sex-limited (13). The hypothesis that this sequence may be a cis-acting element is further supported by another finding. The first 7 bases, GGTC AAT, bear striking homology to a *C. elegans* heptamer, TGTC AAT, a conserved sequence repeated within the first 200 bp upstream of each of the six vitellogenin genes (14).

In the DNA sequence downstream from the mRNA cap site, we predict that the translation begins at the first AUG at position +100. This assumption is supported by two observations. First, this assignment yields a continuous open reading frame for the first three exons, which code for approximately 137 amino acids. Second, the predicted N-terminus for this protein is compatible with the fact that Vg is a secreted protein; that is, the initiation methionine is followed by an arginine and then a block of 13 contiguous non-polar amino acids. A similar exceptionally short 15 amino acid signal sequence is a characteristic of the vertebrate and nematode vitellogenins, which previously have been shown to be homologous proteins (14). The fact that the short sea urchin signal sequence is quite closely related to the nematode/vertebrate sequences suggests that the sea urchin vitellogenin gene may be homologous to the others. On the other hand, the sea urchin Vg mRNA has a 5'-untranslated sequence of 99 bp. This is quite long compared to all other Vg genes studied. The 5'-untranslated sequences of the vertebrate, nematode and locust genes are only 9 to 28 bp long and

those of the three Drosophila genes are about 55 bp in length (14,17,29,30,31). Thus, a final answer to the question of an evolutionary relationship between the sea urchin gene and the Vg genes of other evolutionarily distant organisms must await further sequence analysis.

ACKNOWLEDGEMENTS

We thank Chyi-Ying Chen for help with the primer extension experiments, and W. Klein for gifts of sea urchin DNA samples and a genomic library prepared in his laboratory. The oligonucleotide synthesis facility at Indiana University is funded by a contract from the Indiana Corporation for Science and Technology. This study was supported by NIH grant GM30870 to T.B. and NIH grant HD21986 to R.A.R.

*To whom correspondence should be addressed

*Present address: Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA, USA

REFERENCES

1. Wahli, W., Dawid, I.B., Wyler, T., Jaggi, R.B., Weber, R., and Ryffel, G.U. (1979) *Cell* **16**, 535-549.
2. Arnberg, A.C., Meijlink, F.C.P.W., Mulder, J., Bruggen, E.F.J. van, Gruber, M., and Geert, A.B. (1981) *Nucleic Acids Res.* **9**, 3271-3286.
3. Barnett, T., Pacht, C., Gergen, J.P., and Wensink, P.C. (1980) *Cell* **21**, 729-738.
4. Wyatt, G.R., Locke, J., Bradfield, J.Y., White, B.N., and Deeley, R.G. (1981) in *Juvenile Hormone Biochemistry*, Pratt, G.E. and Brooks, G.T. Eds., pp. 299-307, Elsevier/North Holland Biomedical Press, Amsterdam.
5. Blumenthal, T., Squire, M., Kirtland, S., Cane, J., Donegan, M., Spieth, J., and Sharrock, W.J. (1984) *J. Mol. Biol.* **174**, 1-18.
6. Spieth, J. and Blumenthal, T. (1985) *Mol. Cell Biol.* **5**, 2495-2501.
7. Bownes, M. and Hames, B.D. (1977) *J. Expt. Zool.* **200**, 149-156.
8. Sharrock, W.J. (1983) *Dev. Biol.* **96**, 182-188.
9. Chen, T.T., Strahlendorf, P.W., and Wyatt, G.R. (1978) *J. Biol. Chem.* **253**, 5325-5331.
10. Wiley, H.S. and Wallace, R.A. (1978) *Biochem. Biophys. Res. Commun.* **85**, 153-159.
11. Wang, S.-Y. and Williams, D.L. (1980) *Biochemistry* **19**, 1557-1563.
12. Wang, S.-Y., Smith, D.E., and Williams, D.L. (1983) *Biochemistry* **22**, 6206-6212.
13. Shyu, A.-B., Raff, R.A., and Blumenthal, T. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3865-3869.
14. Spieth, J., Denison, K., Kirtland, S., Cane, J., and Blumenthal, T., (1985a) *Nucleic Acids Res.* **13**, 5283-5295.
15. Spieth, J., Denison, K., Zucker, E., and Blumenthal, T. (1985b) *Nucleic Acids Res.* **13**, 7129-7138.
16. Wahli, W. and David, I.B. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1437-1441.
17. Walker, P., Brown-Luedi, M., Germond, J.-E., Wahli, W., Meijlink,

- F.C.P.W., Van het Schip, A.D., Roelink, H., Gruber, M., and AB, G. (1983) *EMBO J.* **2**, 2271-2279.
18. Kimble, J. and Sharrock, W.J. (1983) *Dev. Biol.* **96**, 189-196.
19. Brennan, M.D., Weiner, A.J., Goralski, T.J., and Mahowald, A.P. (1982) *Dev. Biol.* **89**, 225-236.
20. Wahli, W., Dawid, I.B., Ryffel, G.U., and Weber, R. (1981) *Science* **212**, 298-304.
21. Garabedian, M.J., Hung, M.-C., and Wensink, P.C. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1396-1400.
22. Garabedian, M.J., Shepherd, B.M., and Wensink, P.C. (1986) *Cell* **45**, 859-867.
23. Klein-Hitpass, L., Schorpp, M., Wagner, U., and Ryffel, G.U. (1986) *Cell* **46**, 1053-1061.
24. Seiler-Tuyns, A., Walker, P., Martinez, E., Merillat, A.-M., Givel, F., and Wahli, W. (1986) *Nucleic Acids Res.* **14**, 8755-8770.
25. Jost, J.-P., Seldran, M., and Geiser, M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 429-433.
26. Hardin, S.H., Carpenter, C.D., Hardin, P.E., Bruskin, A.M., and Klein, W.H. (1985) *J. Mol. Biol.* **186**, 243-255.
27. Schoenmakers, H.J.N. and Dieleman, S.J. (1981) *Gen. Comp. Endocrinol.* **43**, 63-70.
28. Schoenmakers, H.J.N., van Bohemen, Ch. G., and Dieleman, S.J. (1981) *Dev. Growth and Differen.* **23**, 125-135.
29. Locke, J., White, B.N., and Wyatt, G.R. (1987) *DNA*, In press.
30. Hung, M.-C. and Wensink, P.C. (1983) *J. Mol. Biol.* **164**, 481-492.
31. Yan, Y.L., Kunert, C.J., and Postlethwait, J.H. (1987) *Nucleic Acids Res.* **15**, 67-85.
32. Messing, J. (1983) *Meth. Enzymol.* **101**, 20-78.
33. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
34. Smith, A.J.H. (1980) *Meth. Enzymol.* **65**, 560-580.
35. Biggin, M.D., Gibson, T.J., and Hong, G.F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963-3965.
36. Southern, E.M. (1975) *J. Mol. Biol.* **98**, 503-517.
37. Singh, K., Carey, M., Saragosti, S., and Botchan, M. (1985) *Nature* **314**, 553-556.
38. Walker, P., Germand, J.-E., Brown-Luedi, M., Givel, F., and Wahli, W. (1984) *Nucleic Acids Res.* **12**, 8611-8626.