Sarai, A., & Takeda, Y. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 6513–6517.
Schevitz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L., & Sigler, P. B. (1985) *Nature 317*, 782–786.
Senear, D. F., & Ackers, G. K. (1990) *Biochemistry* (in press).
Senear, D. F., Brenowitz, M., Shea, M. A., & Ackers, G. K. (1986) *Biochemistry 25*, 7344–7354.
Tsai, S. Y., Tsai, M.-J., & O'Malley, B. W. (1989) *Cell 57*, 443–448.
Turner, R., & Tjian, R. (1989) *Science 243*, 1689–1694.
Vershon, A. K., Liao, S.-M., McClure, W. R., & Sauer, R. T. (1987a) *J. Mol. Biol. 195*, 311–322.
Vershon, A. K., Liao, S.-M., McClure, W. R., & Sauer, R. T. (1987b) *J. Mol. Biol. 195*, 323–331.
Wang, J. C. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 200–203.
Wartell, R. M., & Adhya, S. (1988) *Nucleic Acids Res. 16*, 11531–11541.
Whitson, P. A., Olson, J. S., & Matthews, K. S. (1986) *Biochemistry 25*, 3852–3858.
Whitson, P. A., Hsieh, W.-T., Wells, R. D., & Matthews, K. S. (1987) *J. Biol. Chem. 25*, 4943–4946.
Zwieb, C., Kim, J., & Adhya, S. (1989) *Genes Dev. 3*, 606–611.

A Single mRNA Encodes Multiple Copies of the Egg Peptide Speract[†,‡]

Chodavarapu S. Ramarao,[§] Deborah J. Burks,[§] and David L. Garbers*[,§,∥]
Department of Pharmacology and Department of Molecular Physiology and Biophysics and Howard Hughes Medical Institute, Vanderbilt University School of Medicine, Nashville, Tennessee 37232
Received August 11, 1989; Revised Manuscript Received November 27, 1989

ABSTRACT: A complementary DNA clone (2.3 kb) that encodes the egg peptide speract (Gly-Phe-Asp-Leu-Asn-Gly-Gly-Gly-Val-Gly) has been isolated from an ovary cDNA library of the sea urchin *Strongylocentrotus purpuratus*. The DNA sequence predicts an open reading frame of 296 amino acids. The likely site of initiation, however, is a downstream in-frame translation initiation codon that would result in a polypeptide of 260 amino acids containing 10 decapeptides, each separated by a single lysine residue. Four of the peptides are speract, and six have the predicted structures of Gly-Phe-Ala-Leu-Gly-Gly-Gly-Gly-Val-Gly (occurs twice), Gly-Phe-Asn-Leu-Asn-Gly-Gly-Gly-Val-Gly, Gly-Phe-Ser-Leu-Thr-Gly-Gly-Gly-Val-Gly, Gly-Thr-Met-Pro-Thr-Gly-Ala-Gly-Val-Asp, and Ile-Asp-His-Asp-Thr-Leu-Ala-Ser-Val-Ser. The isolated cDNA insert hybridized to two species of ovarian mRNA (1.2 and 2.3 kb) obtained from species known to produce speract or speract-like peptides, but failed to hybridize to RNA from other species. Subsequently, a second ovarian cDNA clone (1.2 kb) was isolated and sequenced; this clone contained two additional potential decapeptides: Ser-Phe-Asp-Leu-Asn-Gly-Gly-Gly-Val-Gly and Ser-Thr-Met-Pro-Thr-Gly-Ala-Gly-Val-Asp. The various speract and speract-like peptides found in egg-conditioned media, therefore, reflect, in part, variable structures within a single copy of mRNA. The peptides are likely processed by a combination of trypsin and carboxypeptidase B activity; although the peptides containing a Gly-Gly-Gly-Val-Gly carboxyl-terminal sequence are known to possess sperm respiration-stimulating activity, the function, if any, of the other potential peptides remains to be determined.

The egg-conditioned media of sea urchins contain peptides that are capable of stimulating sperm metabolism and motility; these peptides may also act as chemoattractants for spermatozoa (Hansbrough & Garbers, 1981a; Suzuki et al., 1981; Garbers et al., 1982; Ward et al., 1985). One such peptide, speract (Gly-Phe-Asp-Leu-Asn-Gly-Gly-Gly-Val-Gly), was initially isolated from egg-conditioned media of *Strongylocentrotus purpuratus* sea urchins based on its ability to stimulate sperm respiration at acidic extracellular pH values (Hansbrough & Garbers, 1981a; Suzuki et al., 1981). Subsequently, speract was shown to have several effects on spermatozoa including increased motility (Hansbrough & Garbers, 1981a; Suzuki et al., 1981), elevated cyclic nucleotide concentrations (Garbers et al., 1982), proton efflux leading to intracellular alkalinization (Hansbrough & Garbers, 1981b), dephosphorylation of guanylate cyclase, leading to markedly decreased enzyme activity (Ramarao & Garbers, 1985), and transient elevations of Ca^{2+} (Schackmann & Chock, 1986; Lee & Garbers, 1986). The egg peptides are species-specific, and speract, for example, fails to stimulate spermatozoa from the sea urchin *Arbacia punctulata* (Ward et al., 1985). Egg-conditioned media of *A. punctulata* contain a different peptide named resact (Cys-Val-Thr-Gly-Ala-Pro-Gly-Cys-Val-Gly-Gly-Gly-Arg-Leu-NH_2) that stimulates spermatozoa (Suzuki et al., 1984; Ward et al., 1985). A diversity of peptide structures exist across the species, which accounts for the failure of egg-conditioned media of a particular species to necessarily cross-react with spermatozoa from another species. However, even within the same species, many variants of the same basic peptide structure have been isolated (Shimomura et al., 1986; Suzuki et al., 1988). From the egg-conditioned media of *Hemicentrotus pulcherrimus*, for example, the following peptides have been chemically isolated: Gly-Phe-Asp-Leu-Asn-Gly-Gly-Gly-Val-Gly, Gly-Phe-Asp-Leu-Thr-Gly-Gly-Gly-Val-Gly, Gly-Phe-Ser-Leu-Asn-Gly-Gly-Gly-

[†] Supported by NIH Grant HD10254 and by a grant from the Andrew W. Mellon Foundation.
[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02896.
* Address correspondence to this author at the Howard Hughes Medical Institute.
[§] Department of Pharmacology.
[∥] Department of Molecular Physiology and Biophysics and Howard Hughes Medical Institute.

0006-2960/90/0429-3383$02.50/0

Val-Ser, Ser-Phe-Ala-Leu-Gly-Gly-Gly-Gly-Val-Gly, and Gly-Phe-Ser-Leu-Ser-Gly-Ser-Gly-Val-Asp (Suzuki et al., 1988); all possess biological activity. Since egg-conditioned media have been collected from multiple sea urchins, it has not been clear whether or not the multiple peptide structures within a species are due to individual variation, a diverse number of peptides encoded within a single mRNA, or multiple genes.

Here, cDNA clones coding for speract have been isolated and the DNA sequences determined. After in vitro translation of the mRNA, antibodies produced to speract specifically precipitated a 28-kDa protein; the mass of this protein was equivalent to that predicted from the cDNA sequence, suggesting that speract is synthesized as a large molecular weight precursor. The deduced amino acid sequence of the putative precursor protein contained multiple speract and speract-like structures demonstrating that peptide diversity within a given species is explained, at least in part, by the synthesis of variants from a single mRNA.

Experimental Procedures

A *S. purpuratus* ovarian cDNA library constructed in λ Zap was a generous gift from Drs. Christer Hoog and Eric Davidson at the California Institute of Technology, Pasadena, CA.

S. purpuratus were purchased from Marinus Inc., Long Beach, CA. Restriction endonucleases, T4 DNA ligase, polynucleotide kinase, and other enzymes were from either New England Biolabs or Promega Inc. mCAP, mRNA capping kit, and rabbit reticulocyte in vitro express translation kit were from Strategene. The random-primed DNA labeling kits were from Boehringer Mannheim Biochemicals. Hybond N nylon membrane filters were purchased from Amersham. Radiolabeled nucleotides were from Amersham and Dupont/NEN. All other reagents and chemicals were of the highest purity available.

Oligonucleotide Probes. Two antisense probes were synthesized with an Applied Biosystems Model 380A nucleotide synthesizer: probe I, 5′-CCIAAICTIAAITTICCICCIC-CICAICCI-3′; probe II, 5′-CCIAAICTIGAITTICCICC-ICCICAICCI-3′. All codons contained 2-deoxyinosine at the third position according to the method of Ohtsuka et al. (1985).

Isolation of a cDNA Clone for Speract. Ovarian cDNA was transformed into *Escherichia coli* XL-1 cells. Approximately 5×10^5 clones were screened with a ^{32}P end labeled mixture of probes I and II. Prehybridization was performed at 45 °C for 16 h in a solution containing 6 × 150 mM NaCl/15 mM trisodium citrate, pH 7.0 (SSC),[1] 5 × Denhardt solution, 20 mM NaH_2PO_4, and 500 μg/mL denatured salmon sperm DNA. The filters were then incubated for 16 h at 45 °C in the presence of a solution containing 0.4% $NaDodSO_4$, 6 × SSC, 20 mM NaH_2PO_4, 500 μg/mL denatured salmon sperm DNA, and a mixture of ^{32}P end labeled probes I and II at 1×10^6 cpm/mL. Following hybridization, the filters were washed 3 times for 5 min each at room temperature and once for 30 min at 35 °C with a solution containing 6 × SSC and 0.1% $NaDodSO_4$. Twenty initial positive clones were isolated and the pBluescript plasmids were rescued by R408 helper phage as described (Russel et al., 1986). Restriction endonuclease fragments of one isolated clone were subcloned into M13 mp18 and mp19 (Yanishch-Peron et al., 1985) vectors and sequenced in both directions by Sanger's dideoxy chain termination method (Sanger et al., 1977) using the Sequenase kit (United States Biochemicals).

Southern Blot Analysis. Conditions for prehybridization, hybridization, and posthybridization washes were the same as described above. For genomic DNA, the final wash was performed at 65 °C in the presence of 0.1 × SSC containing 0.1% $NaDodSO_4$ for 15 min.

Northern Blot Analysis. Total RNA from the ovaries of *S. purpuratus*, *Strongylocentrotus franciscanus*, *Lytechinus pictus*, and *A. punctulata* was prepared by the LiCl precipitation method (Cathala et al., 1983). Polyadenylated RNA was obtained from these preparations by using oligo(dT)–cellulose (Collaborative Research) column chromatography (Davis et al., 1986). Three micrograms of RNA from each species was electrophoresed on 1% agarose gels that were equilibrated with 20 mM 3-(*N*-morpholino)propanesulfonic acid buffer, pH 5.5, containing 2% formaldehyde, 5 mM sodium acetate, and 1 mM EDTA. Prehybridization of RNA fixed onto nylon membranes was performed for 24 h at 45 °C with a solution containing 6 × SSC, 10 × Denhardt solution, and 500 μg/mL denatured salmon sperm DNA. The RNA on the membranes was then hybridized to the random-primed, [^{32}P]dCTP-labeled cDNA insert (2.3 kb) at 45 °C for 24 h. After the filters were washed 3 times with 6 × SSC and 1 × Denhardt solution at 45 °C for 30 min each, filters were further washed 2 times (30 min each) with 2 × SSC and 0.1% $NaDodSO_4$ at 45 °C. This was followed by a final wash with 0.1 × SSC and 0.1% $NaDodSO_4$ at 65 °C for 15 min.

In Vitro Transcription and Translation. The pBluescript clone containing the 2.3-kb insert was used in these experiments. mRNA was synthesized from 2 μg of DNA using the mCAP kit from Stratagene and translated in a rabbit reticulocyte lysate in vitro express translation kit from Stratagene using [^{35}S]methionine.

Immunoprecipitation. Two microliters of the in vitro translation mixture was incubated overnight at 0–4 °C with 5 μL of antiserum to speract or with preimmune serum in a total volume of 50 μL containing 25 mM Tris, pH 7.5. *Staphylococcus aureus* cells, washed and resuspended in phosphate-buffered saline (50 μL), were added to the antibody–antigen complex and incubated on ice for 1 h. The mixture was microfuged, and the pellets were washed 3 times with 200 μL of phosphate-buffered saline. The pellets were solubilized in 50 μL of sample buffer (75 mM Tris, pH 6.8, 5% $NaDodSO_4$, 10% glycerol, 10% β-mercaptoethanol, and 0.2% bromophenol blue) by treating at 95 °C for 15 min. The tubes were centrifuged, and the supernatant fluid was applied to a 12.5% polyacrylamide gels as described by Laemmli (1980).

Speract Antibodies. Anti-speract antibodies were produced by East Acres Biologicals, Southbridge, MA. Speract was cross-linked to keyhole limpet hemocyanin and injected into rabbits. The preimmune and immune sera were tested, and the presence of antibodies in the immune sera was established by immunoblot analysis of whole egg extracts of *S. purpuratus* and by immunoprecipitation of GGG[Y^2]-speract. GGG-[Y^2]-speract is a synthetic analogue of speract that has comparable biological activity (Dangott & Garbers, 1984).

Results

Isolation of cDNA Clones. Antisense oligonucleotide probes I and II were used to screen a *S. purpuratus* ovarian cDNA library constructed in λ Zap. Twenty positive clones were selected and purified by further screening. Digestion of the clones with *Eco*RI yielded 1.2-kb inserts in 19 of the 20 clones; 1 clone contained a 2.3-kb insert. Inserts from all of the clones

[1] Abbreviations: SSC, 150 mM NaCl/15 mM trisodium citrate, pH 7.0; Denhardt solution, 0.02% poly(vinylpyrrolidone), 0.02% bovine serum albumin, and 0.02% ficoll 400; $NaDodSO_4$, sodium dodecyl sulfate.

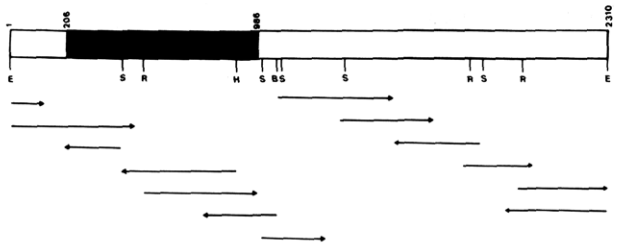

FIGURE 1: Partial restriction map of the 2.3-kb cDNA insert and the strategy for DNA sequencing. The entire cDNA is represented at the top with its coding (closed box) and noncoding (open box) regions. Horizontal arrows show the directions and extent of sequencing. The *Eco*RI cloning site sequence at either end of the insert was obtained by double-stranded sequencing. Several overlapping, sequenced clones are not shown. Restriction endonucleases used for subcloning were as follows: B, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; R, *Rsa*I; *Sau*3A1.

hybridized with the oligonucleotide probes by Southern blotting (data not shown), suggesting that all of them contained DNA sequences that coded for speract or speract-like peptides.

Characterization of 2.3-kb Insert. The restriction map and sequencing strategy for the largest insert (2.3 kb) are shown in Figure 1, and the nucleotide sequence and deduced amino acid sequence of the insert are shown in Figure 2. Open reading frames beginning with methionine at nucleotide 96 or 204 and extending through a termination signal at 984–986 represent the apparent precursor polypeptide. Within this reading frame, four sequences correspond to speract (Gly-Phe-Asp-Leu-Asn-Gly-Gly-Gly-Val-Gly). Two potential decapeptides with the sequence Gly-Phe-Ala-Leu-Gly-Gly-Gly-Gly-Val-Gly and two other decapeptides (Gly-Phe-Ser-Leu-Thr-Gly-Gly-Gly-Val-Gly and Gly-Phe-Asn-Leu-Asn-Gly-Gly-Gly-Val-Gly) with structures similar to speract exist within the deduced amino acid sequence of the apparent precursor. Each predicted peptide is separated by a single lysine residue. Two other decapeptides with little similarity to speract, Gly-Thr-Met-Pro-Thr-Gly-Ala-Gly-Val-Asp and Ile-Asp-His-Asp-Thr-Leu-Ala-Ser-Val-Ser, bordered by single lysine residues are also present in the deduced amino acid sequence. There are other potential peptides both longer and shorter than 10 amino acids in the deduced amino acid sequence that are flanked by lysine residues (Figure 2). It is not known if any of these potential peptides possess biological activity. A long nontranslated AT-rich region of 1324 nucleotides exists on the 3′ end. Two polyadenylation signal sequences AATAAA are present at nucleotide positions 1480–1485 and 1623–1628.

Assignment of Start Codon. Two potential start codons (ATG) at nucleotide positions 98–100 and 206–208 are present, but the one at 206–208 represents the likely initiation site. For the codon at positions 206–208, nucleotides G at positions +4 and −6 and C at −1, −5, and −9 are the most frequently occurring bases in those positions among a survey of 699 vertebrate mRNAs (Kozak, 1987). Also, nucleotide A at position −2 and G at −12 are the second most frequently occurring bases in those positions (Kozak, 1987). Resulting protein would have a molecular weight of 26960. Immediately following this ATG, a stretch of 22 amino acids with the characteristics of a signal peptide is present (Briggs & Gierasch, 1986). In contrast, the other potential start site (positions 98–100) has nucleotide G at position −9 as the most frequently occurring base in that position while nucleotides A at −10 and −11 and G at −12 are the second most frequently occurring bases in these positions (Kozak, 1987). The 15 amino acid region that immediately follows this potential start codon is relatively hydrophilic. From this start site, a 30651-Da protein would be synthesized. There are six other

```
GAATTCGGCTACTCGACTTTACGCTAGATCTCTACCGAGCTTTCTCGTCGAAAAGAACAACCGACCGACCCACGTCGTTGAAGATGAAGTTCATCGT  ATG CCT CCT GGC CCT CGT GGC GTA GCT TCG GGC AAG CCT GTC ATC  142
                                                                                                      Met Pro Pro Gly Pro Arg Gly Val Ala Ser Gly Lys Pro Val Ile
                                                                                                       1                                  10

AGC GGT CAG GAC CAG AAG TAC ACC CTT GCT GAT TCC AGC AAG GAT CTA GAG GGC ATT GCT CAC ATG GCT ATC GTC GAC TCA CTT TCC CCT CTA CAC ATC TCG CTC TCC TCT TTG GAA TCT  262
Ser Gly Gln Asp Gln Lys Tyr Thr Leu Ala Asp Ser Ser Lys Asp Leu Glu Gly Ile Ala His Met Ala Ile Val Asp Ser Leu Ser Pro Leu His Ile Ser Leu Ser Ser Leu Glu Ser
 16                 20                 30                 40                 50

GCG TGG AAC AAC CTC CTC AAC ATT GCT TCT CAG GAA GAC TAC TCT ATC CCG GAA CTT TCC ATA CCC AAA ATC GAC GTG AAG AGC ATC TTG TCA TGC AAG CCC AAG TAT TCT CCC AAG TAC  382
Ala Trp Asn Asn Leu Leu Asn Ile Ala Ser Gln Glu Asp Tyr Ser Ile Pro Glu Leu Ser Ile Pro Lys Ile Asp Val Lys Ser Ile Leu Ser Cys Lys Pro Lys Tyr Ser Pro Lys Tyr
 56                 60                 70                 80                 90

CCT GTG GTC TTG CAG TAC ATC TCT GAT CAT TAC GTC CAG GTC CAG GAC CAT ATC GCC AAC GCC AAG GAA CTT ACA GAA GGA CTG AAA TTC GTC TCG CAG CTC ATC ATG TAC AAG AAG ATC  502
Pro Val Val Leu Gln Tyr Ile Ser Asp His Tyr Val Gln Val Gln Asp His Ile Ala Asn Ala Lys Glu Leu Thr Glu Gly Leu Lys Phe Val Ser Gln Leu Ile Met Tyr Lys Lys Ile
 96                100                110                120                130

GAC CAC GAT ACC CTG GCC TCT GTC TCT AAA ATG TTG AGC AAT TAC TTG ACT GAC TAC GCA TCG ACA ATC TCA TCA CTT AAA TCA GTC GTC TGT CAA GAT CAA ACA GCG CCA TCT CAT CCT  622
Asp His Asp Thr Leu Ala Ser Val Ser Lys Met Leu Ser Asn Tyr Leu Thr Asp Tyr Ala Ser Thr Ile Ser Ser Leu Lys Ser Val Val Cys Gln Asp Gln Thr Ala Pro Ser His Pro
                140                150                160                170

ATG GAC GAG AGT TAC ATG GAC ACA CCC CTG TCA ATG ATC TTG AAA GGC ACC ATG CCT ACT GGC GCT GGA GTA GAC AAA GGC TTC GCA CTT GGT GGT GGT GGA GTT GGT AAA GGA TTC AAC  742
Met Asp Glu Ser Tyr Met Asp Thr Pro Leu Ser Met Ile Leu Lys Gly Thr Met Pro Thr Gly Ala Gly Val Asp Lys Gly Phe Ala Leu Gly Gly Gly Gly Val Gly Lys Gly Phe Asn
                180                190                200                210

TTG AAC GGC GGT GGA GTT GGC AAA GGC TTT GAC TTG AAC GGC GGT GGA GTT GGC AAA GGA TTC GAC TTG AAC GGC GGT GGA GTT GGC AAA GGC TTT GAC TTG AAC GGC GGT GGA GTT GGC  862
Leu Asn Gly Gly Gly Val Gly Lys Gly Phe Asp Leu Asn Gly Gly Gly Val Gly Lys Gly Phe Asp Leu Asn Gly Gly Gly Val Gly Lys Gly Phe Asp Leu Asn Gly Gly Gly Val Gly
                220                230                240                250

AAA GGC TTT GAC TTG AAC GGC GGT GGA GTT GGC AAA GGT TTC GCT TTG GGC GGT GGT GGA GTT GGC AAA GGA TTC AGC CTC ACC GGA GGA GGC GTC GGC AGG GAA GTC GAA ATC AAA GAC  982
Lys Gly Phe Asp Leu Asn Gly Gly Gly Val Gly Lys Gly Phe Ala Leu Gly Gly Gly Gly Val Gly Lys Gly Phe Ser Leu Thr Gly Gly Gly Val Gly Arg Glu Val Glu Ile Lys Asp
                260                270                280                290

TGG TAA  TCGACGTTCTGGCGCTCAATAGATCTGACATCAAATGCCATGCGCCCGTACAGGAATTTCATAATTGCCATTTTTTTTTACACTGATATCAATACTTTGGTCGTGCCTGTCCAATTATACAAATATGTATTAAATAATAGCATCAAATGC 1138
Trp End
 297

AAAAAAAAAAAAAAAAAAAAAAGGGTCTTGTTTTCTTGAGCTTGCAATGATCAACCATCATCCAGTGTCGTCATGATGGTCTGCGTCACAAGAGAGGCGATAAGCAAAATTTAAGACTCAAGATGGTATGTTAAACTTTGATCTTAAGGACGTTTTCTTT 1298

GCTTTTTAACGAGTCTTAAAACTAACATGAATAATGCATACTGTTATTAGAAAGGTCTTTAATTCCGTGTGTTTAAACTTATGCTGAGGCATTCGAGAAAAATATATGCCTGTATTAATCAGCAATACCTGCCTGCTATTCACTGACAGCGAGTTCTTCT 1458

GTTTCCTCATCAAATATTGAAAATAAATTTTTAACAAAAAAAATTACATATGCCTGTAAAATGGCTATTTCAGCATTTTTCTCTGTTTTCTACGATTCATATTTCTTTTTCATTGGAAAAACTAGAAAGAATATATAAAATATAAAATATTTTCAAAATA 1618

GTTTAATAAATTATTTCCAAAAGACAACTTTGTCAAAAAGAATTGAAAAAAGAAGAAAATGAATTGAAAGCAAAATTTTCTTTTCAAAATGTCAGCAACTGATAGTGATAGCCGCAGGTATTACATCAGGCAAATCAAACAGGTAAAGGTCATACAATGT 1778

ACCTTCTAGATGTTACTTTACAAGATTTTAGATCTACTTTGGTATCTTCTTTCCTTTTTCCTTTGATTGTAAACTCCTTTCAAGTAATACGAATTAATACCATAAGTTGCACCTGCCTAGCCTATACAGTCGACACTGCCCACAAACAAATCGAGAAGTA 1938

TAGAATAACGCAACATAAAAAAGGAAAGATTTTGGTAGCTGTGCTCTTTTCCTCTTTTCTTATCTTTCGAATGTTTACTCTCCTGTACTCATTTAAAATGGTAAATTAAAAAGAGAGCCAGATTTTTATTTGGAAGGTTGCAAATCGATTGACAGATAGG 2098

TCATCAGTGTAGAGCTTTGATACAAGTGACAAAAACGGTCCATACGCTTGTTGAGCAATATGGTTACTTCATGAAGGCCTTGTATATTTCACATAATGATAACAAATACTAGAAGTCAATGATGCATTGCTGCAAAGAATTTTGATAACGTGACTTTCTT 2258

TGTAGTCTGGTGCTTGGTAAAATCCAGAGACGCAATGAAAATACTTGAATTC 2310
```

FIGURE 2: Complete nucleotide sequence and deduced amino acid sequence of the 2.3-kb insert. The two potential start codons are marked with arrows. Speract and other decapeptides are boxed, and the lysine residues are shaded. Two polyadenylation signals toward the 3′ end are underlined, and termination codons are boxed. Numbers to the right are nucleotide positions, and numbers below each line are amino acid residue numbers starting with Met (1) at nucleotide positions 98–100.

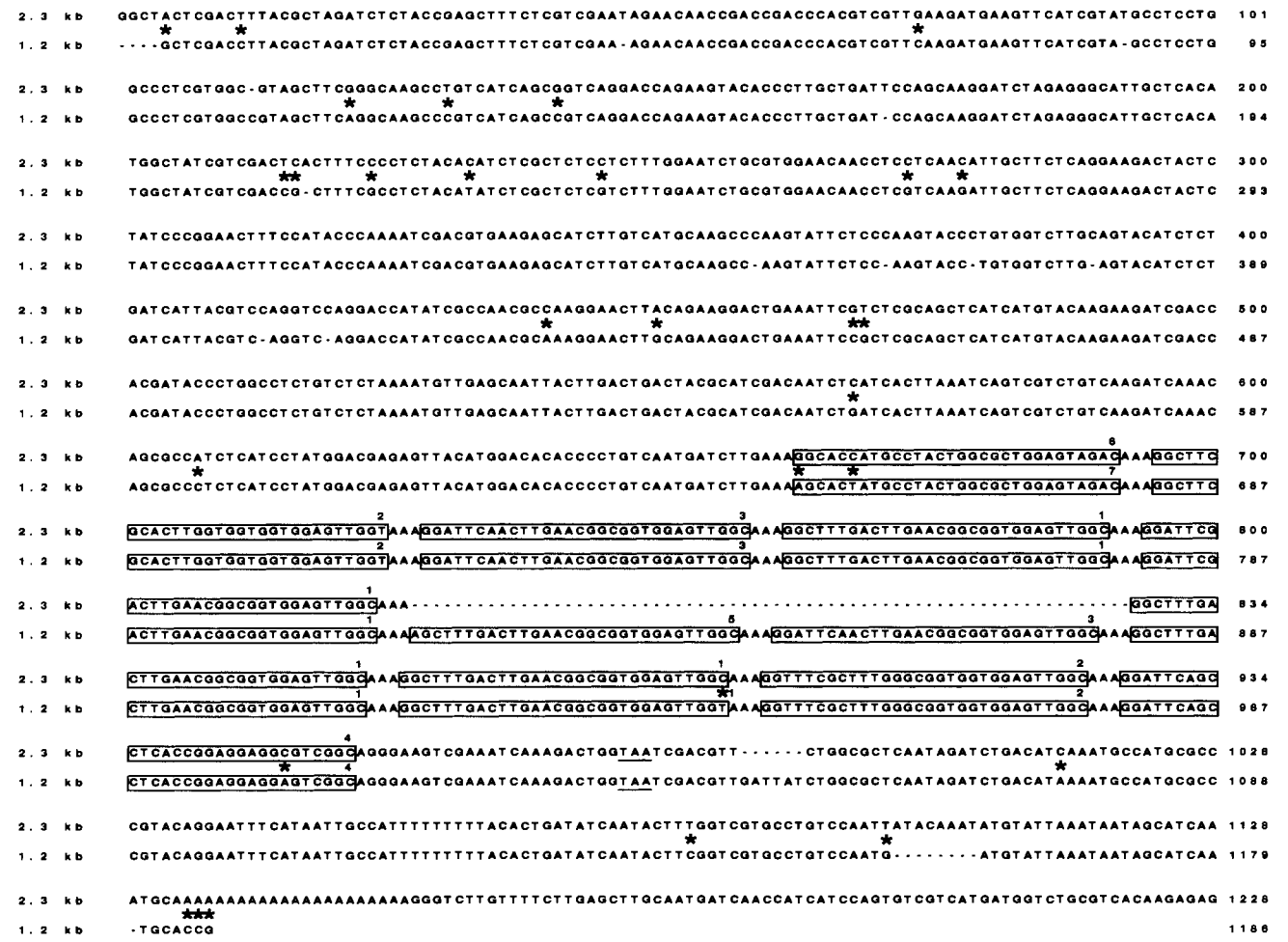

FIGURE 3: Alignment of the nucleotide sequences of 2.3- and 1.2-kb cDNA clones. Asterisks denote differences in the sequences. Gaps introduced for best alignment are represented by hyphens. Numbers to the right are nucleotide positions. Nucleotide sequences that code for speract and other decapeptides are boxed. Numbers on top of each sequence refer to the peptides in Table I. Termination codons at the 3′ end of the messages are underlined. The 2.3-kb clone extends approximately 1000 bp further on the 3′ end (Figure 2).

ATG codons within the same open reading frame at nucleotide positions 497–499, 533–535, 623–625, 638–640, 656–658, and 674–676 (Figure 2).

Characterization of 1.2-kb Insert. The clones containing the 1.2-kb insert had a restriction endonuclease digestion pattern similar to the 2.3-kb insert (data not shown), and the various 1.2-kb clones mapped identical with each other. A comparison of the nucleotide sequence of a 1.2-kb clone with that of the 2.3-kb insert is shown in Figure 3. The bases that differ are marked with asterisks. Sequence analysis of the 1.2-kb clone showed the presence of DNA encoding one additional speract analogue (Ser-Phe-Asp-Leu-Asn-Gly-Gly-Gly-Val-Gly). The predicted peptide, Gly-Phe-Asn-Leu-Asn-Gly-Gly-Gly-Val-Gly, was repeated twice whereas it occurred only once in the longer clone. In addition, the predicted peptide, Gly-Thr-Met-Pro-Thr-Gly-Ala-Gly-Val-Asp, present in the 2.3-kb insert was replaced by the sequence Ser-Thr-Met-Pro-Thr-Gly-Ala-Gly-Val-Asp in the 1.2-kb insert. The smaller insert also differed from the 2.3-kb insert in that it lacked approximately 1000 bp from the 3′-untranslated region.

RNA Hybridization. To determine the size of the mRNA in *S. purpuratus* and to determine whether or not a similar message existed in related species of sea urchins, poly(A+) RNA from *S. purpuratus*, *L. pictus*, *S. fransciscanus*, and *A. punctulata* was analyzed by Northern blot hybridization, using the 2.3-kb insert as a probe. Three species (*S. purpuratus*, *L. pictus*, and *S. franciscanus*) contained two hybridizing bands of mRNA of approximately 2.3 and 1.2 kb (Figure 4). However, the probe did not detectably hybridize to the mRNA from *A. punctulata* (Figure 4). Spermatozoa from *A. punctulata* do not respond to speract (Ward et al., 1985; Shimomura et al., 1986), and egg-conditioned media of *A. punctulata* do not stimulate *S. purpuratus* or *L. pictus* spermatozoa (Shimomura et al., 1986). *L. pictus* spermatozoa, in contrast, respond to speract, and *L. pictus* egg-conditioned media are known to contain speract-like peptides (Hansbrough & Garbers, 1981a).

Analysis of cDNA-Expressed Products. When the 2.3-kb cDNA was digested and transcribed as described under Experimental Procedures, the resulting mRNA was determined to be 2.3 kb as judged by Northern blot analysis (data not shown). mRNA synthesized in the proper orientation (*Bam*HI cleavage) resulted in an approximately 28-kDa protein that specifically reacted with antibodies to speract (data not shown). Neither mRNA made in the opposite orientation (*Xho*I cleavage) nor RNA transcribed in vitro with T7 RNA polymerase (antisense strand) resulted in detectable immunoprecipitation. The apparent molecular weight of translated protein roughly agrees with the start site at nucleotide positions 206–208. These results indicate that the cDNA encodes a 2.3-kb mRNA which is translated into a speract precursor.

Discussion

We have described the isolation and sequence of two cDNA clones that encode speract, a peptide that stimulates sper-

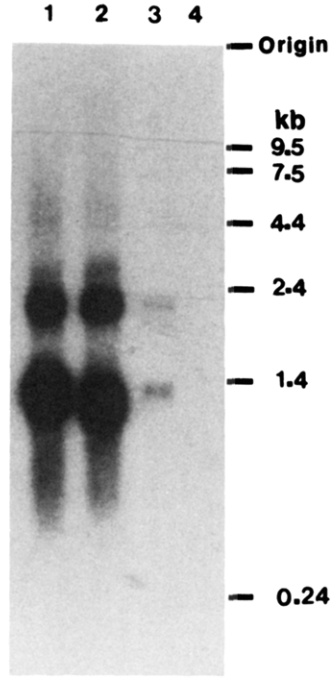

FIGURE 4: Analysis of ovarian mRNA from various sea urchin species by Northern blot hybridization. Lane 1, *S. purpuratus*; lane 2, *S. fransciscanus*; lane 3, *L. pictus*; lane 4, *A. punctulata*. Three micrograms of poly(A+) RNA was loaded per lane. Preparation of mRNA and hybridization and washing conditions are described under Experimental Procedures. Numbers to the right correspond to the RNA standards.

Table I: Amino Acid Sequences of Potential Decapeptides Predicted from the cDNA Sequence of the 1.2- and 2.3-kb Inserts Isolated from an Ovary cDNA Library

		no. of copies	
no.[a]	sequence	2.3-kb insert	1.2-kb insert
1	GFDLNGGGVG	4	4
2	GFALGGGGVG	2	2
3	GFNLNGGGVG	1	2
4	GFSLTGGGVG	1	1
5	SFDLNGGGVG		1
6	GTMPTGAGVD	1	
7	STMPTGAGVD		1

[a] Numbers correspond to peptide sequences in Figure 3.

matozoan motility. When transcribed and translated in vitro, each cDNA directs the synthesis of a 28-kDa protein precursor that is immunoprecipitated by anti-speract antiserum. As shown in Figure 3, the coding and noncoding regions of the 1.2- and 2.3-kb clones are very similar. The amino acid sequence deduced from the clones predicts a precursor protein containing four sequences that correspond to speract, each flanked by a single lysine residue. Six other potential decapeptides separated by single lysine residues exist in the 2.3-kb insert. The clones differ in that the 1.2-kb clone contains seven speract-like molecules, in addition to the four speract sequences (Table I, Figure 3). The 1.2-kb clone also lacks approximately 1000 bp that are present in the 3′-untranslated region of the larger clone.

Northern blot analysis of *S. purpuratus* ovary mRNA confirms the existence of two transcripts (1.2 and 2.3 kb) that are recognized by the speract cDNA probe. Several explanations could be offered for the presence of two different messages. They could be products of two nonallelic genes, or they may represent genetic variation among sea urchin individuals. Another explanation for the existence of the 2.3- and 1.2-kb clones could be alternate mRNA splicing. Alternate splicing as a mechanism for the generation of these two mRNAs is suggested by the two polyadenylation signals present in the 3′ region of the 2.3-kb insert. The splicing of different exons not only could account for the variation in size of the mRNAs but also might explain the presence of additional speract analogues in the 1.2-kb sequence (nucleotides 813–879).

Of the peptide structures predicted from the DNA sequence, Gly-Phe-Ala-Leu-Gly-Gly-Gly-Gly-Val-Gly isolated from the ethanol extracts of *S. purpuratus* and *Pseudocentrotus depressus* egg jelly (Suzuki et al., 1988) and Gly-Phe-Ser-Leu-Thr-Gly-Gly-Gly-Val-Gly purified from *S. purpuratus* egg jelly (Suzuki et al., 1988) have been chemically identified. Eight of the 10 peptide sequences in the 2.3-kb cDNA clone and 10 of the 11 deduced peptide sequences in the 1.2-kb cDNA clone conform to a formula devised by Suzuki et al. (1988) for biologically active speract-like peptides: X_1-Phe-X_2-Leu-X_3-Gly-X_4-Gly-Val-X_5, where X_1 = Gly or Ser, X_2 = Asp, Ala, or Ser, X_3 = Asn, Ser, Ala, or Gly, X_4 = Ala or Gly, and X_5 = Gly, Ser, or Asp. Other sequences, Ile-Asp-His-Asp-Thr-Leu-Ala-Ser-Val-Ser, Gly-Thr-Met-Pro-Thr-Gly-Ala-Gly-Val-Asp (from 2.3 kb), and Ser-Thr-Met-Pro-Thr-Gly-Ala-Gly-Val-Asp, do not fit this formula, and they have yet to be synthesized and tested for biological activity.

The apparent evolutionary pressure to retain multiple active peptide structures within a single mRNA remains unclear. In the case of *H. pulcherrimus* and *Anthocidaris crassispina*, the chemically isolated analogues of speract appear to possess equivalent biological potency (Suzuki et al., 1988). Therefore, it appears that mutations of speract tend to be in the direction of molecules with equivalent activity. This implies that one speract molecule within a single mRNA would not be sufficient for effective fertilization, and strong pressure, therefore, is exerted to retain multiple copies of the active peptide. An alternative explanation is that the proposed precursor is cleaved between the single dibasic (lysine) residues at amino acid positions 133 and 134. It could be argued, then, that the complete structure from amino acid 135 to amino acid 288 is the important biologically active principle and that speract and the associated peptides represent degradation products containing activity. Possibly the large fragment also would possess additional activity to that of speract. Yamaguchi et al. (1988), for example, recently have shown that speract can potentiate the effect of a jelly coat macromolecular complex on the induction of an acrosome reaction. Could the larger fragment containing multiple copies of speract play an even greater role in the acrosome reaction? On the other hand, a critical concentration of speract or active speract analogues may depend on the formation of multiple copies of the peptides from a single mRNA. Presently, it is not known when or where the precursor mRNA is formed, or when the polypeptide is synthesized. In *Orthopyxis*, Miller (1978) found that sperm cells became immotile in the acellular matrix surrounding the egg until "active" molecules were released from the egg; this activation was coincident with formation of the second polar body. In the sea urchin, the second polar body forms prior to the spawning of gametes. With the cDNA now in hand, it should be possible to express the precursor in cells that do not normally produce speract; if an active molecule(s) is (are) released, site-directed mutagenesis may reveal the significance of multiple speract and speract-like peptides within the same polypeptide.

A feature of the apparent precursor molecule is the separation of speract and speract analogues by single lysine residues. If a trypsin-like protease is present, then the precursor could be processed by tryptic digestion, followed by carbox-

ypeptidase B activity that specifically hydrolyzes the CO_2H-terminal lysine and arginine residues. Trypsin-like proteases are well documented as being present in eggs (Sawada et al., 1984) and cortical granules (Alliegro & Schuel, 1985, 1988) of *S. purpuratus* and other species of sea urchins.

Many neuropeptides and hormones are initially synthesized as large precursor molecules. The classical cleavage site in the proteolytic processing of a precursor is a sequence of two basic amino acids originally described by Steiner et al. (1967). Since then, a large number of precursors that are proteolyzed at dibasic residues have been described (Loh & Gainer, 1983). However, processing after a single basic amino acid residue is also known to occur. A list of precursors in which monobasic processing is found includes those for vasopressin and atrial natriuretic peptide, growth factors, neuropeptides, substance P, proenkephalins A and B, interleukin 3, the frog skin peptides xenopsin and PGLa, yeast killer toxin, and *Aplysia* egg-laying hormone precursor (Schwartz, 1986). While a majority of monobasic cleavage sites contain arginine, lysine is known to occur in at least two precursors. *Aplysia* neuroactive peptide Phe-Met-Arg-Pro-NH_2 is synthesized as a polyvalent precursor that contains at least 19 copies of the peptide (Schaffer et al., 1985). It has been proposed that monobasic cleavage takes place in at least 15 sites, followed by removal of the carboxyl-terminal lysine residue from each peptide (Schaffer et al., 1985). The cholecystokinin precursor also may be processed after a single lysine residue, giving rise to cholecystokinin 22 (Eng et al., 1983; Zhou et al., 1985).

Acknowledgments

We thank Dr. Christer Hoog and Dr. Eric Davidson for the *S. purpuratus* ovary cDNA library. The secretarial assistance of Penny Stelling and Cindy Reeder is appreciated.

References

Alliegro, M. C., & Schuel, H. (1985) *Biochemistry 24*, 3926–3931.

Alliegro, M. C., & Schuel, H. (1988) *Dev. Biol. 125*, 168–180.

Briggs, M. S., & Gierasch, L. M. (1986) *Adv. Protein Chem. 38*, 109–180.

Cathala, G., Savouret, J.-F., Mendez, B., West, B. L., Karin, M., Martial, J. A., & Baxter, J. D. (1983) *DNA 2*, 329–335.

Dangott, L. J., & Garbers, D. L. (1984) *J. Biol. Chem. 259*, 13712–13716.

Davis, L. G., Dibner, M. D., & Battey, J. F. (1986) in *Basic Methods in Molecular Biology*, pp 99–101, Elsevier, New York.

Eng, J., Shiina, Y., Pan, Y.-C. E., Blacher, R., Chang, M., Stein, S., & Yalow, R. S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 6381–6385.

Garbers, D. L., Watkins, H. D., Hansbrough, J. R., Smith, A., & Misono, K. S. (1982) *J. Biol. Chem. 257*, 2734–2737.

Hansbrough, J. R., & Garbers, D. L. (1981a) *J. Biol. Chem. 256*, 1447–1452.

Hansbrough, J. R., & Garbers, D. L. (1981b) *J. Biol. Chem. 256*, 2235–2241.

Kozak, M. (1987) *Nucleic Acids Res. 15*, 8125–8148.

Laemmli, U. K. (1980) *Nature 227*, 680–685.

Lee, H. C., & Garbers, D. L. (1986) *J. Biol. Chem. 261*, 16026–16032.

Loh, Y. P., & Gainer, H. (1983) in *Brain Peptides* (Krieger, D. T., Brownstein, M. J., & Martin, J. B., Eds.) pp 79–116, Wiley, New York.

Miller, R. (1978) *J. Exp. Zool. 20*, 203–211.

Ohtsuka, E., Matsuki, S., Ikehara, M., Takahashi, Y., & Matsuhara, K. (1985) *J. Biol. Chem. 260*, 2605–2608.

Ramarao, C. S., & Garbers, D. L. (1985) *J. Biol. Chem. 260*, 8390–8396.

Russel, M., Kidd, S., & Kelly, M. R. (1986) *Gene 45*, 333–338.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Sawada, H., Miura, M., Yokosawa, H., & Ishii, S. (1984) *Biochem. Biophys. Res. Commun. 121*, 598–604.

Schackmann, R. W., & Chock, B. P. (1986) *J. Biol. Chem. 261*, 8719–8728.

Schaffer, M., Picciotto, M. R., Kreiner, T., Kaldany, R.-R., Taussing, R., & Scheller, R. H. (1985) *Cell 41*, 457–467.

Schwartz, T. W. (1986) *FEBS Lett. 200*, 1–10.

Shimomura, H., Suzuki, N., & Garbers, D. L. (1986) *Peptides 7*, 491–495.

Steiner, D. F., Cunningham, D. D., Spigelman, L., & Aten, B. (1967) *Science 157*, 697–700.

Suzuki, N., Kajiura, H., Nomura, K., Garbers, D. L., Yoshiono, K., Kurita, M., Tanaka, H., & Yamaguchi, M. (1988) *Comp. Biochem. Physiol. 89B*, 687–693.

Suzuki, N., Nomura, H., Ohtake, H., & Isaka, S. (1981) *Biochem. Biophys. Res. Commun. 99*, 1238–1244.

Suzuki, N., Shimomura, H., Radany, E. W., Ramarao, C. S., Ward, E. G., Bentley, J. K., & Garbers, D. L. (1984) *J. Biol. Chem. 259*, 14874–14879.

Ward, G. E., Brokaw, C. J., Garbers, D. L., & Vacquier, V. D. (1985) *J. Cell Biol. 101*, 2324–2329.

Yamaguchi, M., Niwa, T., Kurita, M., & Suzuki, N. (1988) *Dev., Growth Differ. 30*, 159–167.

Yanisch-Peron, C., Vieira, J., & Messing, J. (1985) *Gene 33*, 103–119.

Zhou, Z. Z., Eng, J., Pan, Y.-C. E., Chang, M., Holmes, J. D., Raufman, J.-P., & Yalow, R. S. (1985) *Peptides 6*, 337–341.