Cloning of the mRNA for the protein that crosslinks to the egg peptide speract

(sperm/sea urchin)

LAWRENCE J. DANGOTT*†‡§, JAN E. JORDAN†, RAE ANN BELLET*, AND DAVID L. GARBERS†‡¶

*Howard Hughes Medical Institute and Departments of †Pharmacology and ‡Molecular Physiology and Biophysics, Vanderbilt University Medical Center, Nashville, TN 37232

Communicated by Stanley Cohen, December 5, 1988

ABSTRACT An apparent receptor for the egg peptide speract (Gly-Phe-Asp-Leu-Asn-Gly-Gly-Gly-Val-Gly) was identified by covalently coupling a radiolabeled speract analogue to intact spermatozoa and was then purified by DEAESephrose chromatography and preparative gel electrophoresis after solubilization with Lubrol PX. The purified, crosslinked protein was digested with Staphylococcus aureus V8 protease and a resultant peptide, purified from polyacrylamide slab gel slices, was shown to have the peptide originated from the apparent receptor. Clone containing a 2.5-kilobase insert was subsequently isolated from a sea urchin testis cDNA library that contained DNA sequences encoding an open reading frame of 532 amino acids that included the above peptide sequence. The deduced amino acid sequence suggests that the protein contains a 26-residue amino-terminal signal peptide, a large extracellular domain relatively rich in cysteine (5%) that includes a four-fold repeat of about 115 amino acids, a single membrane-spanning region, and only 12 amino acids residues extending into the cytoplasm. Analysis of total RNA from Strongylocentrotus purpuratus testis by Northern blot revealed a 2.5-kilobase RNA. Preliminary data show the presence of hybridizing RNA of the same apparent size in other sea urchin species, including Arbacia punctulata, which does not respond to speract.

Spermatozoa contain binding sites for specific molecules associated with the egg (includes the acellular matrices surrounding the egg); these interactions may result in substantial morphological or behavioral changes in the spermatozoa. One such group of molecules is represented by certain egg peptides that can stimulate sperm metabolism and motility and can cause a chemotactant response of sperm cells when present as chemical gradients (1–5). The structures of the peptides vary between sea urchin species; however, the major biochemical responses to each of the peptides, which include net H+ efflux (6), increased intracellular Ca2+ (7), increased cyclic AMP and cyclic GMP (1, 3, 4), and activation of guanylate cyclase activity (8), appear to occur in all of the species. Speract (Gly-Phe-Asp-Leu-Asn-Gly-Gly-Gly-Val-Gly), the first such peptide isolated (1, 2), was shown in early experiments to bind to a surface receptor (9). Based on equilibrium binding studies, it then became clear that the amino-terminal domain of the peptide could be modified without significant decreases in binding or biological activity (10). These studies led to the synthesis of a speract analogue, Gly2Tyr2speract (Gly-Gly-Tyr-Gly-Tyr-Asp-Leu-Asn-Gly-

Gly-Gly-Val-Gly), that could be iodinated and crosslinked to a membrane protein in the presence of disuccinimydyl suberate (11). The apparent receptor appeared to represent a glycoprotein of Mr 77,000. To examine the signaling system of spermatozoa in detail, we have purified the crosslinked protein, obtained limited amino acid sequence information, and isolated a cDNA clone for the protein. The mature protein contains an estimated 502 amino acids and a single transmembrane domain.

MATERIALS AND METHODS

Animals and Reagents. Strongylocentrotus purpuratus were purchased from Marinus (Westchester, CA). Arbacia punctulata were purchased from Gulf Specimen (Panacea, FL). Restriction endonucleases were from New England Biolabs or International Biotechnologies. All other reagents were of the highest purity available.

Purification and Partial Amino Acid Sequence Analysis of the Crosslinked Protein. Radioiodinated Gly2Tyr2speract was covalently bound to intact spermatozoa as described (11); this allowed us to detect the apparent receptor throughout the purification procedure. The radiolabeled cells were solubilized with occasional mixing for 1 hr (0–2°C) in 50 mM Tris-HCl, pH 7.6/250 mM NaCl/0.5% Lubrol PX. The mixture was then centrifuged at 100,000 × g for 1 hr (4°C) and the supernatant was applied to a column of DEAESephrose (Sigma) equilibrated with 50 mM Tris-HCl, pH 7.6/250 mM NaCl/0.1% Lubrol PX. The column was washed extensively with the equilibration buffer and the radiolabeled receptor was subsequently eluted with the same solution containing 500 mM NaCl. The fractions containing radioactivity were pooled, concentrated by ultrafiltration, and subjected to preparative NaDodSO4/polyacrylamide gel electrophoresis using the buffer system of Laemmli (12). The purified receptor was concentrated and analyzed by NaDodSO4/polyacrylamide slab gel electrophoresis.

Peptides, produced by digestion of the purified radiolabeled receptor with Staphylococcus aureus V8 protease according to the method of Cleveland et al. (13), were separated in 15% polyacrylamide slab gels containing NaDodSO4 and visualized by staining briefly with Coomassie blue R-250 (Bio-Rad). The peptides were excised, eluted from the gel, and prepared for microsequencing (14). One of the isolated peptides was submitted to automated Edman degradation on an Applied Biosystems 470A gas-phase sequenator equipped with an on-line Applied Systems 120A PTH amino acid analyzer.

§Present address: The Worcester Foundation for Experimental Biology, 222 Maple Avenue, Shrewsbury, MA 01545.

†To whom reprint requests should be addressed.

The sequence reported in this paper is being deposited in the EMBL/GenBank data base (accession no. J04518).
Site-Directed Antiserum. A synthetic peptide corresponding to the sequence obtained from the crosslinked protein was synthesized (Peninsula Laboratories), conjugated to bovine serum albumin with disuccinimidyl suberate (15), and injected into New Zealand White rabbits. The initial injection was performed in the presence of complete Freund's adjuvant; all subsequent booster injections were with incomplete adjuvant. The antiserum was tested by immunoblot analysis of *S. purpuratus* spermatozoan plasma membranes prepared according to Bentley and Garbers (16). Membrane proteins were separated by NaDodSO4/polyacrylamide slab gel electrophoresis and electroblotted onto nitrocellulose membranes. The nitrocellulose filter was then incubated in phosphate-buffered saline (PBS: 150 M NaCl/10 M phosphate, pH 7.2) containing 5% (wt/vol) non-fat dried milk at 37°C to block nonspecific binding and then with the same solution containing 1% (vol/vol) Tween 20 and a 1:200 dilution of the antiserum for 1 hr at room temperature. The filter was extensively washed with PBS/0.1% NaDodSO4/0.1% sodium deoxycholate/1.0% Tween 20. Immunoreactive proteins were detected with biotinylated anti-rabbit antibodies and avidin-conjugated horseradish peroxidase from a kit purchased from Vector Laboratories.

The ability of the antiserum to precipitate the receptor was tested by incubation of the antiserum with solubilized radiolabeled receptor at 4°C for 12 hr. The immunocomplex was then precipitated with formalin-fixed *S. aureus* cells (BRL), solubilized in NaDodSO4 sample buffer, and subjected to NaDodSO4/polyacrylamide slab gel electrophoresis. The gel was dried and the radioactive proteins were detected by autoradiography with Kodak X-Omat film at -70°C.

Identification and Nucleic Acid Sequence Analysis of cDNA Clone for the Speract Receptor. A *S. purpuratus* testis cDNA library in λgt10 was generously provided by E. Davidson (California Institute of Technology, Pasadena, CA). Approximately 3 x 10⁶ colonies were screened on replicate filters with end-labeled, mixed antisense oligonucleotide probes made to the previously determined peptide sequence. The probes, d[IA(T-I/C)/CC-ITC-IAT-IAT-I(A/I)(GC-G(T-C))/C(TC)^2(C-I)/AI-ITC-IAA-I(G/C)/GC^2(C-G)], synthesized with an Applied Biosystems 380A nucleotide synthesizer, contained 2'-deoxyinosine at the first or third positions of several codons and the radioactive probes were detected by autoradiography with Kodak X-Omat film at -70°C.

RESULTS AND DISCUSSION

Purification of the Crosslinked Protein. The apparent speract receptor was previously identified by NaDodSO4/polyacrylamide gel electrophoresis as a membrane glycoprotein of Mr 77,000 (11). Intact sperm cells from *S. purpuratus*, crosslinked with the radioiodinated speract analogue (Gly3-[Tyr2]speract), were treated with Lubrol PX and the solubilized proteins were then subjected to DEAE-Sepharose ion-exchange chromatography. The radiolabeled protein was eluted with 500 mM NaCl and comigrated with a silver-stained protein of Mr 77,000 when analyzed by NaDodSO4/polyacrylamide gel electrophoresis. Subsequent preparative-scale polyacrylamide gel electrophoresis of this material under denaturing conditions resolved the protein, which appeared as a diffuse silver-stained band on analytical-scale NaDodSO4/polyacrylamide gels (Fig. 1). Autoradiography of the gels showed that the radiolabeled receptor comigrated with the stained Mr 77,000 protein.

In the initial studies we performed two experiments to show that the protein-staining band corresponded to the crosslinked protein. We first analyzed the protein by two-dimensional polyacrylamide gel electrophoresis in which proteins were separated by isoelectric point and by molecular size. Under these conditions the radiolabeled receptor comigrated with the silver-stained protein (data not shown). In a second experiment we reasoned that if the stained protein represented the receptor, then radiolabeled fragments of digested, crosslinked receptor would comigrate with peptides from the Mr 77,000 protein. This in fact was the case.

![Fig. 1. NaDodSO4/polyacrylamide gel analysis of the purified Mr 77,000 crosslinked protein under reducing conditions.](image-url)
Crosslinked, radiolabeled receptor digested with *S. aureus* V8 protease yielded a number of well-resolved Coomassie blue-stained bands when subjected to NaDodSO₄/15% polyacrylamide gel electrophoresis, and subsequent autoradiography of the dried gels demonstrated that radioactivity comigrated with some of the stained peptides (data not shown). This observation strongly suggested that these peptides were obtained from the crosslinked protein. Several of the radiolabeled peptides were electrophoresed from the gels and prepared for amino acid sequencing. One of the isolated peptides gave the sequence Val-Ser-Ala-Pro-Phe-Asp-Leu-Glu-Ala-Pro-Phe-Ile-Asp-Gly-Ile.

**Synthetic Peptide Antisera.** Antibodies were raised against the above peptide sequence in order to verify the identity of the protein from which the peptide originated. The antiserum recognized a single band at *M*ₗ 77,000 when tested on immunoblots of spermatozoan plasma membrane proteins from *S. purpuratus* (Fig. 2A). No immunoreactive bands were detected, however, in immunoblots of plasma membrane proteins prepared from spermatozoa of *A. punctulata*, a sea urchin that is not activated by sperm. Further, the site-directed antibodies were able to immunoprecipitate the crosslinked, radiolabeled protein from solubilized *S. purpuratus* spermatozoa, demonstrating that the peptide sequence was from the apparent speract receptor (Fig. 2B).

Sequence of the cDNA Clone. The sequencing strategy for the full-length cDNA clone is presented in Fig. 3. The nucleotide sequence and deduced amino acid sequence are presented in Fig. 4. The nucleotide sequence of the two isolated clones span 2461 base pairs (bp). We have assigned the initiation codon to the ATG at position 118 because (i) there are two upstream in-frame stop codons; (ii) this ATG is flanked by sequences that fit Kozak's criteria for a translation initiation codon (22), and (iii) the 26-amino acid sequence following this ATG possesses the features characteristic of signal sequences (23). The initiation codon is followed by an open reading frame of 1593 bp. An in-frame stop codon occurs at position 1714 and is followed by a 3' untranslated region composed of 747 bp that includes a polyadenylylation consensus signal sequence (AATAAA) 16 bp upstream from the poly(A)⁺ tail.

The apparent receptor can be divided into various tentative domains based on the deduced amino acid sequence. Following the hydrophobic signal peptide domain is an amino-terminal domain composed of 461 amino acids. This is followed by a second hydrophobic region of 29 amino acids that is flanked on the carboxyl-terminal side by Lys-Phe-Val-Lys and on the amino-terminal end by a single acidic residue; these features are typical of membrane-spanning domains of membrane proteins (24). The distance from the first basic residue following the putative transmembrane region to the carboxyl terminus is 12 amino acids. It is predicted that the amino-terminal domain of the protein represents the extracellular portion and the carboxyl-terminal domain the intracellular region of the receptor.

The deduced amino acid sequence suggests that cleavage of the signal peptide would yield a protein of 502 amino acids with a calculated molecular weight of 54,641. The discrepancy between the apparent molecular weight of the protein in NaDodSO₄/polyacrylamide gels (*M*ₗ 77,000) and the deduced molecular weight may be due to the presence of carbohydrate on the mature protein. There are three potential N-linked glycosylation sites (Asn-Xaa-Thr) at positions 78-80, 115-117, and 459-461; additional O-linked glycosylation sites may exist. Previous work demonstrated that the crosslinked speract receptor is a glycoprotein (11).

A striking feature of the deduced amino-terminal domain is the relatively high cysteine content (5%). Computer analysis of the speract receptor sequence with a dot matrix homology program shows that the amino-terminal, cysteine-rich domain is composed of a sequence of about 115 residues that is repeated four times with minor variations (Fig. 5). By the introduction of a few small gaps, each of the sequences can be aligned so that the cysteine residues are in register. In addition, several residues other than the cysteine appear to

---

**Fig. 3.** Restriction endonuclease map and sequencing strategy for speract receptor cDNA clone. Endonuclease restriction sites are shown above the line: B, Bgl II; C, Cla I; E, EcoRI; H, HindIII; Hc, HincII; K, Kpn I; P, Pst I; R, Rsa I; X, Xho I. Arrows denote the direction and extent of nucleotide chain sequenced by dideoxynucleotide chain termination. Arrows beginning with solid circles represent sequences determined by primer extension using synthetic oligonucleotide primers.
be conserved (Fig. 5). Twelve amino acids (10%) are conserved and 18 others (16%) are in the same position in three of the four repetitive sequences. Twenty-five percent of the invariant residues are acidic.

![Fig. 4](image-url) Complete nucleotide sequence and deduced amino acid sequence of the cDNA corresponding to the crosslinked protein. Nucleotides are numbered on the right. Amino acids are numbered above the sequence. The signal sequence and the putative membrane-spanning sequence are boxed. The sites to which N-linked carbohydrate could be attached (Asn-Xaa-Thr) are indicated by dotted underlines. Cysteine residues are circled, and the predicted amino acid sequence that corresponds to the sequence of the *S. aureus* V8 peptide is underlined. A polyadenylation signal sequence in the 3' untranslated region is indicated by the underlines and overlines.

![Fig. 5](image-url) Alignment of four repetitive, cysteine-rich sequences in the amino-terminal domain of the crosslinked protein. Gaps have been introduced in the sequences to optimize alignments. Identical residues are boxed. Residues that occur three or more times are highlighted with bold letters, and those that occur in all four sequences are indicated by a single asterisk. Cysteine residues are denoted by double asterisks. Numbers at left indicate the position of the amino acids according to the numbering system of Fig. 4.
A search of the GenBank and National Biomedical Research Foundation data bases (July 1, 1988) revealed no significant identity with the crosslinked protein sequence. However, the repetitive cysteine-rich substructure represents a common feature of the extracellular domains of various proteins, including those for the low density lipoprotein, epidermal growth factor, and insulin receptors (25-27) as well as members of the gene superfamily of cellular adhesion proteins (28). It has been proposed that the cysteine-rich repetitive sequences in the low density lipoprotein receptor represent ligand-binding domains (25).

**Northern Blot Analysis.** Analysis of total RNA prepared from testis of *S. purpuratus* demonstrated a 2.5-kb mRNA (Fig. 6). Hybridization of total RNA purified from testis of a closely related species, *Lytechinus pictus*, gave identical results (data not shown). *L. pictus* spermatozoa are activated by speract and, in fact, the eggs of this species have been shown to produce a peptide, Gly-Ph-e-Asp-Leu-Thr-Gly-Gly-Gly-Gly-Val-Gln (4), whose structure is similar to that of speract.

We also have identified a related RNA species in *A. punctulata* testis (Fig. 6). Hybridization of poly(A)+ RNA from the testes of *A. punctulata* revealed an RNA of ~2.5 kb (Fig. 6). These results are of interest in that speract does not activate *A. punctulata* spermatozoa, analysis of membranes with site-directed antiserum by immunoblotting does not detect an antigen, and the radioiodinated speract analogue does not appear to crosslink to any proteins in *A. punctulata* spermatozoa (11). The relatively weak signal from the hybridized species suggests that the RNA shares only limited sequence similarity with that in *S. purpuratus* and *L. pictus* or that the *A. punctulata* RNA is present in relatively low levels.

Resact (Cys-Val-Thr-Gly-Ala-Pro-Gly-Cys-Val-Gly-Gly-Arg-Leu-NH2), an egg peptide from the sea urchin *A. punctulata*, appears to cause the same biochemical responses in *A. punctulata* spermatozoa as described above for speract (3). However, an analogue of resact covalently binds to a M, 160,000 protein identified as guanylate cyclase (29), instead of a M, 77,000 protein. Since RNA for the M, 77,000 protein seems to exist in *A. punctulata*, a species with which speract does not interact, the existence of a family of receptors with variable ligand-binding sites, similar to bacterial chemoreceptors (30), must still be considered possible. The relationship, if any, between guanylate cyclase and the protein that covalently binds to speract remains an important area of study. It is possible that these two proteins lie in close apposition in the membrane and only one represents the active peptide receptor or that they are both subunits of the intact receptor. Alternatively, both proteins may represent independent receptor molecules. The membrane form of guanylate cyclase from *A. punctulata* spermatozoa has been cloned and sequenced (31), but expression studies have so far failed in that neither guanylate cyclase activity nor receptor activity has been detected when the protein has been expressed in mammalian cells. This failure appears to be due to inadequate posttranslational modification, since the enzyme is not phosphorylated and only partially glycosylated. The M, 77,000 protein described here shows only limited sequence similarity to guanylate cyclase. With the isolated cDNA clone for the protein that covalently binds speract in hand, we are now in a position to pursue expression studies that should resolve whether this protein is a receptor or a protein closely associated with a receptor.

We express appreciation to Dr. D. Martin Watson and Paul Matrisian for their assistance in amino acid sequencing. We thank Phil Zentner for help with Northern blot analyses and Cindy Reeder for typing the manuscript. This research was supported by National Institutes of Health Grant HD10254.