

# Sequence of mRNA coding for bindin, a species-specific sea urchin sperm protein required for fertilization

BONING GAO\*, LAURIE E. KLEIN†, ROY J. BRITTEN, AND ERIC H. DAVIDSON

Division of Biology, California Institute of Technology, Pasadena, CA 91125

Contributed by Eric H. Davidson, July 24, 1986

**ABSTRACT** Bindin, a major protein of the sea urchin acrosome granule, mediates the species-specific adhesion and binding of sperm to egg required to effect fertilization. We report the isolation and sequence of bindin cDNA clones prepared from *Strongylocentrotus purpuratus* testis RNA. The bindin gene appears to be productively expressed only in males and only in testes. The protein is produced from a 51-kDa precursor, which is subsequently processed to yield the mature 24-kDa bindin protein.

Fertilization in echinoderms occurs by a multistep, species-specific process, in which each step may serve as a barrier to gene flow across species lines. The initial physiological event as the sperm approaches the egg is the acrosome reaction, which is induced by the fucose sulfate of the egg jelly coat and is mediated by sperm membrane receptors (1–3). This reaction results in exocytosis of the acrosomal granule and the extrusion of the acrosome process. The sperm binds to the egg vitelline layer by means of the acrosome process, and the plasma membranes of the sperm and egg then fuse. In some interspecific combinations heterologous egg jelly does not induce the sperm acrosome reaction, while in others heterospecific acrosome reactions occur, but even when this is observed, fusion of heterospecific sperm and egg membranes fails to take place (4–6). Vacquier and associates (7–10) demonstrated that the major protein of the acrosome granule, which they named bindin, is the molecular species responsible for the recognition reaction by which the acrosome process is bound to a glycoprotein receptor embedded in the vitelline membrane of the egg. Partial purification and characterization of the sperm receptor glycoprotein have been reported (10–12). On activation of the sperm, bindin is exposed by the eversion of the acrosome granule. Bindin molecules coat the external surface of the acrosomal process and have been detected in the electron microscope by immunocytological methods at the exact site of the sperm-egg bond (7). Bindin has been purified and characterized chemically, as reviewed by Vacquier (10). It contains no detectable carbohydrate, and a partial amino acid sequence has been derived (10).

We report here the isolation of bindin cDNA clones and the complete sequence of the derived protein. This showed unexpectedly that bindin is synthesized from a much larger precursor. Bindin is among the few known proteins of nonimmunogenic origin, the sequence of which specifies a developmentally important interaction between different cell types. Comparison with a homologous fragment of bindin sequence available from another sea urchin species of the same genus shows a remarkable degree of divergence that may be of both current and evolutionary significance in preventing interspecific gene flow.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

## MATERIALS AND METHODS

**Construction of cDNA Library.** Double-stranded cDNA was synthesized by the RNase H procedure of Watson and Jackson (13). Linker ligation, size selection of cDNA, and subsequent procedures are as described by Huynh *et al.* (14). About 40,000 recombinant clones were recovered, the average insert length of which was about 1200 nucleotides (nt).

**Genome Blot Hybridization.** Reactions were carried out at 68°C in a medium containing 5× SET (SET = 0.15 M NaCl/0.03 M Tris·HCl, pH 8/2 mM EDTA), 5× Denhardt's solution, denatured, sheared calf thymus DNA (50 µg/ml), and 2 mM sodium phosphate buffer (pH 6.8), and equivalent conditions were applied for washing (45).

**In Vitro Translation and Immunoprecipitation.** Testes were collected from intertidal male sea urchins at the onset of the spawning season. Following dissection, they were pelleted by centrifugation in a table-top centrifuge for 5 min. The procedure utilized for extracting RNA has been described in detail elsewhere (15). The testis poly(A)<sup>+</sup> RNA was translated *in vitro* in a commercial rabbit reticulocyte lysate in the presence of 1 µCi/µl of [<sup>35</sup>S]methionine (1200 Ci/mmol; 1 Ci = 37 GBq), at 30°C for 1 hr. The lysate was then diluted 1:10 with immunoprecipitation buffer containing 0.16 M NaCl, 20 mM Tris·HCl (pH 7.4), 0.1% deoxycholate, and 1% Triton X-100. Rabbit anti-bindin serum or normal rabbit serum was added, and the mixture was incubated at 4°C for 4 hr. The antibody complex was precipitated with fixed *Staphylococcus aureus* (Cowan I strain) (16) and fractionated over a 12.5% NaDodSO<sub>4</sub>/polyacrylamide gel.

**RNA Blot Hybridizations.** RNAs were displayed on a 1% formaldehyde denaturing agarose gel and transferred to a nitrocellulose filter. The probe was the 1.4-kilobase *Bgl* II fragment of bindin cDNA (Fig. 1), labeled by nick-translation to a specific activity of 3 × 10<sup>8</sup> cpm/µg. Hybridization was carried out in 50% (vol/vol) formamide, 5× SSC, 1× Denhardt's solution, 2 mM phosphate buffer containing denatured and sheared calf thymus DNA (50 µg/ml) at 42°C, for 16 hr. Filters were washed with 5× SSC, 2× SSC, and 1× SSC successively, at 42°C. (1× SSC = 0.15 M NaCl/0.015 M sodium citrate, pH 7.0.) The autoradiographs were exposed with an intensifying screen for periods varying from a few hours to several days.

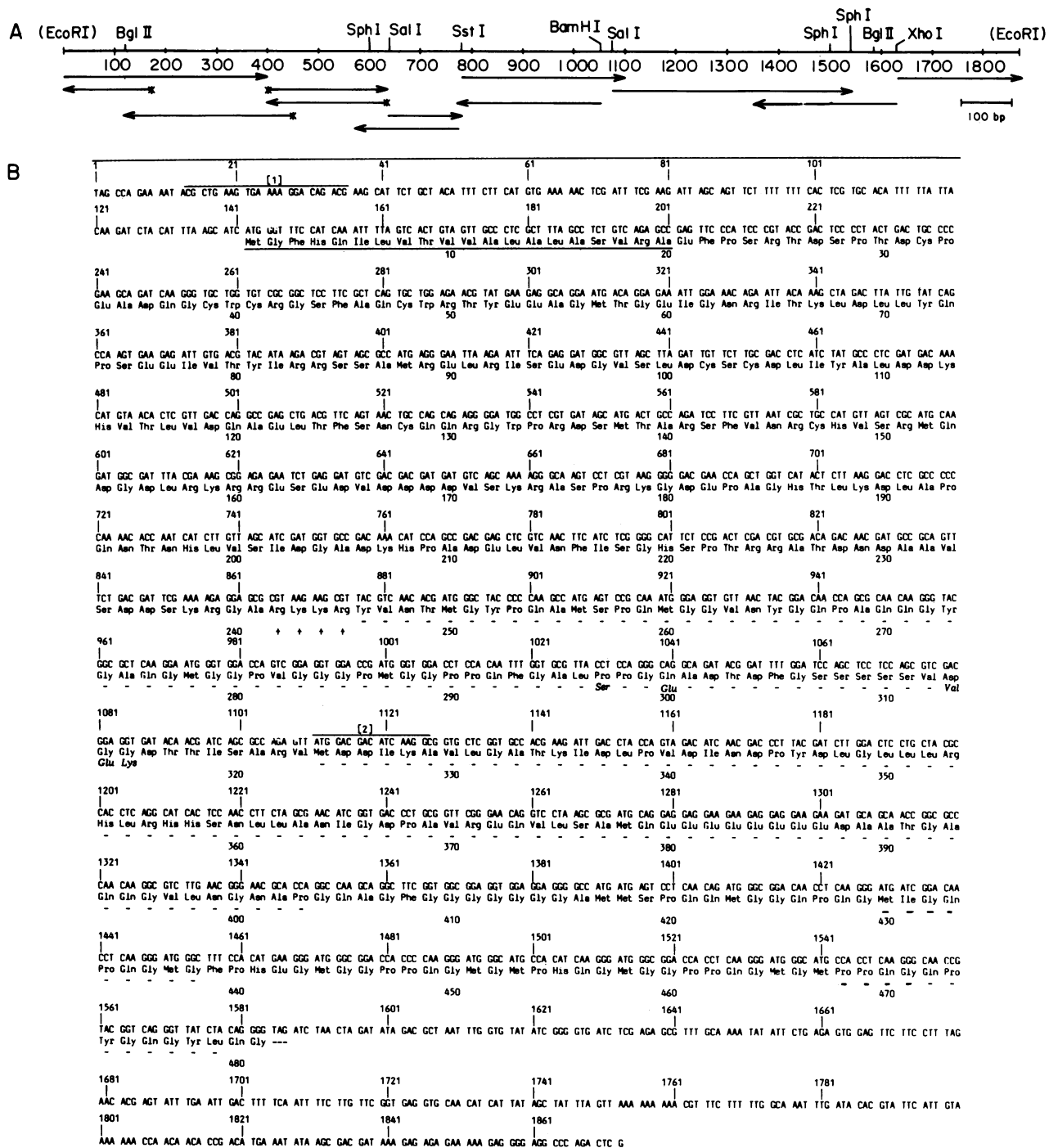
## RESULTS

A λgt10 cDNA library (13) was constructed from testis poly(A)<sup>+</sup> RNA, using random 8- to 12-nt calf thymus DNA fragments as primers for the reverse transcriptase reaction. The poly(A)<sup>+</sup> RNA was extracted from *Strongylocentrotus purpuratus* males obtained from an intertidal population in January, when spermatogenesis is occurring at a maximal

Abbreviation: nt, nucleotides.

\*Present address: Department of Pharmacology, University of Texas Health Science Center, Dallas, TX 75235.

†Present address: University of California Veterinary School, Davis, CA 95616.



**Fig. 1.** Restriction map and structure of bindin cDNA. (A) Cleavage sites and sequencing strategy for bindin cDNA. The lengths of the arrows indicate the direction and number of nucleotides for which sequence was determined by the dideoxynucleotide chain-termination method (17, 18). The fragments marked with asterisks were generated by controlled deletion with exonuclease III (19). The EcoRI sites at the ends of the cDNA derive from the EcoRI linkers used to construct the cDNA library. This clone was selected from the testis cDNA library by screening with a 17-nt probe mixture (see text and B). The probe DNA was labeled by the kinase reaction to a specific activity of  $3 \times 10^8$  cpm/ $\mu$ g and hybridized with filters bearing "plaque lifts" in an aqueous medium at 42°C. (B) Nucleotide sequence of bindin cDNA and predicted amino acid sequence is numbered from the first methionine of the prebindin polypeptide. Some regions of the mature bindin protein sequence have been determined (10). Dashes indicate residues that are identical with the predicted amino acid sequence, while the italicized residues differ from the predicted protein sequence. The underlined N-terminal domain is a hydrophobic region that displays the characteristics of a leader sequence. Arrows indicate basic amino acids that are potential cleavage sites for trypsin-like enzymes. The oligonucleotides utilized for the primer extension experiment and for the isolation of this clone were derived from the regions of sequence indicated by overlines [1] and [2], respectively.

rate. A synthetic 17-nt probe mixture including all possible sequences predicted by a known region of the bindin protein sequence was used to screen the testis library (see Fig. 1). A restriction map of one of the selected clones is shown in Fig. 1A, and the 1873-nt sequence of the bindin cDNA insert included in this clone is given in Fig. 1B. To determine the

position of the 5' end of the mRNA, we carried out a primer extension experiment utilizing a second synthetic oligonucleotide complementary to the region between positions 14 and 36 of the sequence shown in Fig. 1B. This oligonucleotide was bound to testis poly(A)<sup>+</sup> RNA and extended with reverse transcriptase, and the reaction products were displayed

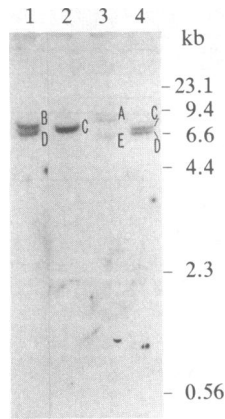


FIG. 2. The bindin gene is a single copy sequence. Genome blots were prepared with sperm DNAs from four different male *S. purpuratus* individuals digested with *Hind*III and reacted with bindin  $^{32}\text{P}$ -labeled cDNA probes. The molecular size markers were  $\lambda$  DNA digested with *Hind*III. As is characteristic for sea urchin genes (15, 20), the flanking sequences are polymorphic, and the genome blot reveals a number of different allelic variants. These are indicated (A–E) to the right of each lane. Thus lane 1 displays alleles B and D; lane 2 appears to be homozygous for allele C; lane 3 displays alleles A and E; lane 4, alleles C and D. This pattern indicates the bindin gene to be single copy since each DNA preparation displays a distinct combination of two of the five alleles observed.

electrophoretically on a sequencing gel (data not shown). These experiments indicated that the 5' terminus of the mRNA is 50 nt beyond the 5' terminus of the cDNA clone.

The initial ATG codon occurring in the sequence shown in Fig. 1B (at position 59) is followed immediately by the stop codon TGA, in the same reading frame. The second ATG signal to occur, at position 142, initiates the 1443-nt open reading frame that codes for the precursor bindin protein. It follows that bindin mRNA has a 188-nt 5'-leader sequence. Following the translation termination signal at position 1585, the cDNA insert continues for an additional 280 nt. This region of the 3'-trailer sequence displays numerous termination codons, and it extends for approximately 500 nt [includ-

ing the poly(A) tract], most of which has not been sequenced, since the size of the mature mRNA is about 2500 nt.

**Derived Sequence of the Bindin Precursor Protein (Probindin).** The mature bindin protein contains 236 amino acids according to the sequence shown. The N-terminal region of the mature bindin protein begins at position 246 of the polypeptide encoded by the bindin mRNA, as determined by the comparison with amino acid sequences obtained from the purified acrosome protein shown in Fig. 1B (10). In all four blocks of mature bindin amino acid sequence, 70, 80, 6, and 12 amino acids in length were available to us (10), and Fig. 1B demonstrates that these can be aligned with the derived amino acid sequence of the cDNA clone with 97% fidelity. The genome blot experiments reproduced in Fig. 2 show that the bindin gene is present in only one copy per haploid genome. Thus the sequenced cDNA clone could not derive from a closely related gene, but must indeed represent the true bindin message. Since the N-terminal 235-amino acid polypeptide is not present in the mature protein, these observations indicate that bindin is initially synthesized as a precursor polypeptide over twice the length of the mature acrosome protein. This polypeptide begins with a characteristic leader sequence (underlined in Fig. 1B) that includes many hydrophobic amino acids (21). Four basic amino acids that occur at the junction between the 245-amino acid N-terminal polypeptide and the mature bindin protein constitute typical cleavage sites for trypsin-like protein processing enzymes (22). These amino acids are marked by arrows in Fig. 1B (amino acids 242–245). Several additional basic amino acid pairs occur in the probindin polypeptide. Although there is no evidence on this point, these sites could also serve as protease cleavage sites. The amino acid compositions of the propolypeptide and the mature bindin protein differ in several respects. As shown in Fig. 3, 34% of the probindin polypeptide residues are hydrophilic, compared to 14% for the mature bindin. Furthermore, all of the cysteine residues are located in the propolypeptide, which suggests the possibility of an internally cross-linked, compacted structure for this region. The mature bindin, on the other hand, is relatively rich in glycine and proline (particularly in the regions between amino acid 274 to 299 and 392 to the C

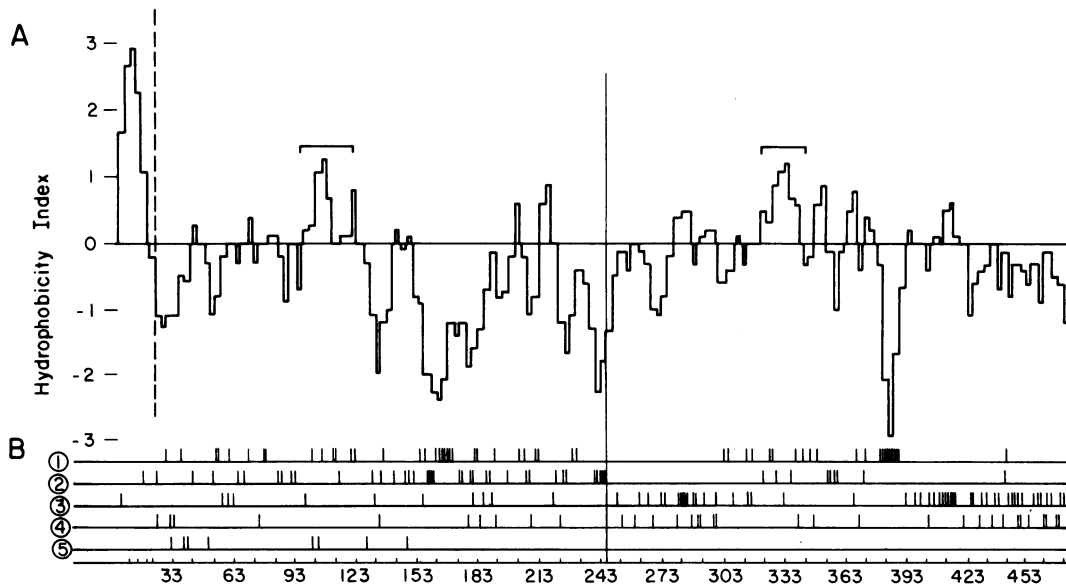


FIG. 3. (A) Hydrophobicity plot for the bindin protein sequence, obtained by the program of Kyte and Doolittle (23), using a search length of nine amino acids. Positive values indicate hydrophobicity and negative values, hydrophilicity. The dotted vertical line denotes the putative leader sequence cleavage site. The N terminus of the mature bindin is indicated by the solid vertical line. The brackets demarcate two regions that are relatively hydrophobic. (B) Distribution of selected amino acids in the bindin protein. Each vertical line indicates one amino acid. Line 1, acidic amino acids including aspartic acid and glutamic acid. Line 2, basic amino acids, including histidine, arginine, and lysine. Line 3, glycine. Line 4, proline. Line 5, cysteine. The amino acids are numbered from the N terminus of the probindin.

terminus), a characteristic of rigid structural protein domains. Two hydrophobic regions (indicated under the brackets in Fig. 3A), which are shorter than the typical transmembrane domain, might serve as sites for interaction with other hydrophobic sequences. The molecular sizes calculated from the deduced protein sequences are 51 kDa for the total prebindin molecule and 24 kDa for the mature bindin protein.

The conclusion that bindin is derived by processing from a much larger precursor is supported by the immunoprecipitation experiment shown in Fig. 4. An anti-bindin antibody generously provided by V. Vacquier was reacted with the *in vitro* translation products of testis poly(A)<sup>+</sup> mRNA, and displayed by NaDodSO<sub>4</sub> gel electrophoresis. As expected from the sequences shown in Fig. 1, the polypeptide that reacts with the anti-bindin antibody in lane D has an apparent molecular size over twice that of the mature bindin protein (see legend to Fig. 4).

**Developmental Expression of the Bindin Gene.** As an initial enquiry into the specificity of expression of the bindin gene, we carried out a search for bindin mRNA in various male and female sea urchin tissues. Total RNA was extracted from testis, 40-hr gastrula stage embryos, and adult sea urchin tube feet; and poly(A)<sup>+</sup> RNAs were prepared from testis, eggs, ovaries, adult sea urchin coelomocytes, lantern tissue, and intestine. The RNAs were hybridized on RNA gel blot with bindin cDNA probes (data not shown). To control the condition of the extracted RNAs, intact actin mRNAs were demonstrated in all of the preparations, as reported (26).

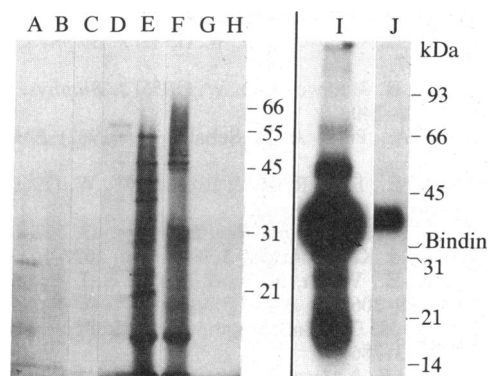


FIG. 4. Immunoprecipitation with anti-bindin antibody of translation products of testis poly(A)<sup>+</sup> RNA. Lane A, *in vitro* translation with testis RNA, immunoprecipitated with normal serum; lane B, *in vitro* translation with testis RNA, precipitated without antibody; lane C, *in vitro* translation with testis poly(A)<sup>+</sup> RNA, immunoprecipitated with bindin antibody, but in the presence of  $5 \times 10^{-3}$   $\mu$ g of unlabeled mature bindin protein; lane D, *in vitro* translation with testis poly(A)<sup>+</sup> RNA, immunoprecipitated with bindin antibody in the presence of  $5 \times 10^{-3}$   $\mu$ g of unlabeled ovalbumin; lane E, total testes poly(A)<sup>+</sup> RNA *in vitro* translation product prior to immunoprecipitation; lane F, egg poly(A)<sup>+</sup> RNA *in vitro* translation product prior to immunoprecipitation; lane G, *in vitro* translation with egg poly(A)<sup>+</sup> RNA, immunoprecipitated with normal serum; lane H, *in vitro* translation with egg poly(A)<sup>+</sup> RNA, immunoprecipitated with anti-bindin antibody. The apparent molecular size of the polypeptide precipitated in the sample displayed in lane D is  $\approx 60$  kDa (indicated by an arrow at left), rather than the 51 kDa (calculated from the sequence shown in Fig. 1B). This is probably due to anomalous migration of this protein in NaDodSO<sub>4</sub>/polyacrylamide gels, relative to standards. Thus the mature bindin protein was reported to migrate in gels as a 30.5-kDa protein (10), while HPLC measurements yielded a mass of 25 kDa (10), close to the value deduced from the sequence of the mature protein. Lane I, partially purified, iodinated (24) preparation of acrosomal protein (25), the major component of which is bindin, prior to immunoprecipitation. After iodination this protein migrates more slowly than does native mature bindin (see lane A). Lane J, immunoprecipitated component of preparation shown in lane I, demonstrating the specificity of the antibody utilized by lanes A–H.

While the testis RNAs contained copious amounts of bindin mRNA, no reactions could be detected with any of the other RNA preparations even after long exposures.

## DISCUSSION

The bindin gene is interesting from physiological, regulatory, and evolutionary vantage points. An unexpected aspect of these observations that pertains to the physiological function of the proteins coded by the bindin message is the presence of the 245-amino acid prebindin N-terminal polypeptide. During spermatogenesis this polypeptide is evidently cleaved from the mature 236-amino acid bindin moiety. The acrosome granule is initially derived from the Golgi complex (1). Thus prebindin may be synthesized in the rough endoplasmic reticulum of the spermatocyte, transported to the Golgi complex, and later incorporated in the membrane-bound acrosome granule (27, 28). Precedents in which mature proteins are initially synthesized as significantly larger precursors and are subsequently processed include certain cellular, viral, hormonal, and neuronal peptides (e.g., refs. 29–32). Among the functions proposed for such precursor regions are that they might facilitate transport of the protein to its ultimate intracellular destination (29), that they might be needed to ensure correct structural conformation (30, 31), and that they are required for storage of the protein as an inactive form (32). An interesting analogy to the bindin case might be provided by the glycine- and proline-rich structural protein collagen. This protein is initially synthesized as a soluble procollagen precursor, in which the cysteines of the propeptide form interchain disulfide bonds, the function of which is to stabilize the triple helical body of the protein (30). On the other hand, it is also possible that the N-terminal prebindin polypeptide has an independent function of its own. Though bindin is indeed the major protein of the isolated acrosome granule (7), other components may have been leached out during the isolation procedure. Several studies have indicated that on exocytosis the acrosome granule releases various enzymes (33–35), and at present it cannot be excluded that the N-terminal bindin polypeptide has this or some other functional acrosomal activity that is also required for fertilization.

We found the bindin gene to be productively expressed only in testis among the tissues studied, and hence it would appear that it is utilized only in males. These qualitative experiments show that bindin mRNA is not present in female gonads or in eggs. This is in contrast to the pattern of expression of the gene for the sea urchin egg yolk protein vitellogenin. Shyu *et al.* (36) showed that in sea urchins vitellogenin is produced both in testes and in ovaries, as well as in intestine cells of both sexes. The observations reported here confirm that sexual determination in sea urchins does indeed involve differential expression of genes in male and female gonads, and they provide a specific molecular marker useful for further studies of this aspect of developmental gene regulation. Other evidence (37–39) demonstrates that there are also H1 and H2B histone genes that are expressed specifically in testis, the products of which are found only in sperm nuclei. Thus there is evidently a battery of male-specific genes that are activated during the differentiation of the sperm and are otherwise quiescent.

Change in the primary sequence of essential coding elements of the bindin gene could have played a crucial role in the evolutionary processes leading to speciation. Thus preference in the bindin–bindin-receptor interaction within a population isolate would lead directly to isolation of its gene pool. In contemporary sea urchins this interaction indeed apparently functions as a barrier to gene flow between species such as *S. purpuratus* and *Strongylocentrotus franciscanus*. Viable hybrid embryos can be formed between

