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, speciesspecific 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 spermegg 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.

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 $5 \times \text{SET}$ (SET = 0.15 M NaCl/0.03 M Tris·HCl, pH 8/2 mM EDTA), $5 \times$ Denhardt's solution, denatured, sheared calf thymus DNA (50 μ g/ml), and 2 mM sodium phosphate buffer (pH 6.8), and equivalent criterion 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 $poly(A)^+$ RNA was translated in vitro in a commercial rabbit reticulocyte lysate in the presence of 1 μ Ci/ μ l of [³⁵S]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 fractioned over a 12.5% NaDodSO₄/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 3×10^8 cpm/µg. Hybridization was carried out in 50% (vol/vol) formamide, $5 \times$ SSC, $1 \times$ Denhart'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 $5 \times$ SSC, $2 \times$ SSC, and $1 \times$ SSC successively, at 42°C. ($1 \times$ 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 λ gt10 cDNA library (13) was constructed from testis poly(A)⁺ RNA, using random 8- to 12-nt calf thymus DNA fragments as primers for the reverse transcriptase reaction. The poly(A)⁺ RNA was extracted from *Strongylocentrotus purpuratus* males obtained from an intertidal population in January, when spermatogenesis is occurring at a maximal

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.

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.

Developmental Biology: Gao et al.

			-				~			_	-		Rem HT.				Sph I								
	(Ecc	oRI) Bg	I II				Spi			Sst I	I		Barn	<u>יר</u>	al I					SphI	Bgl	I_X	(ho I	(Ec	
		iòo	200	300	40	D 50	0 6	00	700	80	0 9	900	IÒ	00 1	100	1200	0 13	00	1400	1500	160	20	1700	1800	
	•	4	*		*			~		-											•	_			
		•					<			-														100 bp	
							-			_														_	
		1		2	1 [1	1		41 				61 				81				101 					
		TAG CCA	GAA AAT	ACG CTG AM	G TGA AA	A GGA CAU	G ACG AAG	CAT 1	TCT GCT A	CA TTT	CTT C	AT GTG 181	*** **	C TCG /	ATT TCG	AAG ATT 201	AGC AG	GT TCT	TTT TT	Z21	TGC ACA	111	TTA TTA		
Number Corr Corr< Corr<		CAA GAT	CTA CAT	TTA AGC AT	Í CATGUL	T TTC CA	T CAA ATT	THA C	GTC ACT G	TA GTT	GCC C	TC GCT	TTA GO	: 101 (STC AGA	GCC GAG	TTC C	CA TOC	CGT ACC	GAC TCC	CCT ACT	GAC	TGC CCC		
$ \begin{array}{c} 1 & 2 \\ 1 & 2 $					Met Gi	ly Phe Hi	a Gin Ile	Leu	Val Thr N	(al Val	Ale L	eu Ala	Leu Al	s Ser	al Arg	Ala Giu 20	Phe P	ro Ser	Arg Th	r Asp Ser	Pro Thi 34	:Asp J	Cys Pro		
Solution Solution <td< td=""><td></td><td>241 </td><td></td><td>26</td><td>1 </td><td></td><td></td><td>281 </td><td></td><td></td><td></td><td>301 </td><td></td><td></td><td></td><td>321</td><td></td><td></td><td></td><td>341 </td><td></td><td></td><td></td><td></td></td<>		241 		26	1 			281 				301 				321				341 					
NI NI <th< td=""><td></td><td>ĠAA GCA Glu Als</td><td>GAT CAA Amp Gin</td><td>GGG TGC TC Gly Cys Ti</td><td>ić tGt CC ™p Cys Al</td><td>SC GGC TCI ng Gly Se⊨</td><td>c TTC GC1 c Phe Ali</td><td>CAG Gin I</td><td>tgc tgg / Cys trp /</td><td>NGA ACG</td><td>TAT G Typ G</td><td>AA GAG</td><td>GCA GG Ala Gl</td><td>A ATG . y Met</td><td>ACA GGA The Gly</td><td>GAA ATT Glu Ile</td><td>GGA A</td><td>AC AGA an Arg</td><td>ATT AC</td><td>A AAG CTA r Lys Leu</td><td>GAC TT/ Asp Le</td><td>L TTG</td><td>TAT CAG Tyr Gin</td><td></td></th<>		ĠAA GCA Glu Als	GAT CAA Amp Gin	GGG TGC TC Gly Cys Ti	ić tGt CC ™p Cys Al	SC GGC TCI ng Gly Se⊨	c TTC GC1 c Phe Ali	CAG Gin I	tgc tgg / Cys trp /	NGA ACG	TAT G Typ G	AA GAG	GCA GG Ala Gl	A ATG . y Met	ACA GGA The Gly	GAA ATT Glu Ile	GGA A	AC AGA an Arg	ATT AC	A AAG CTA r Lys Leu	GAC TT/ Asp Le	L TTG	TAT CAG Tyr Gin		
The set of a per at the set of the an a set of at sec of a set of		361		36	10 11			401		50		421				441				461	,	,			
41 S0 121 141 S4 156 156<		CCA AGT	GAA GAG	ATT GTG AG	G TAC AT	A AGA CG	T AGT AG	ccc	ATG AGG (AA TTA	AGA A	TT TCA	GAG GA	TOGC	GTT AGO		TGT T	CT TGC	GAC CT			C GAT	GAC AAA		
and anddddddddddddddddddddddddddddddddddd		10 301	<u> </u>	110 101 1	10 10	le nig ni	y 301 30		not nig t	90	, nig i		010 40	<i>y</i> ur <i>y</i>	J	100	, суа 3	ei cya			11	0	Aap 298		
Nil Vel The Lei Vel Age Chi Alle Cui Lei The Pee Set Age Cya Chi Cui Alle Age Cya Chi Cui Alle Age Cya Chi Cui Alle Age Cya Chi Cui Age Cui Ag		481 		אכ הזד האת: רו		NG CTG AC	G TTC AG		TGC CAG (CAG AGG	GGA T	194 120 20	CGT GA	TAGC	ATG ACT			TC GTT	AAT CG	01 C TGC CAT	GTT AG	TCGC	ATG CAA		
colc		His Val	Thr Leu	Val Asp G. 12	in Ala G. 10	Lu Leu Th	r Phe Se	Aan	Cya Gin (Gin Arg 130	Gly T	rp Pro	Arg As	p Ser	Met Thr	Ala Are 140	Ser P	he Val	Asn Ar	g Cys His	Val Se 15	r Arg O	Met Gin		
A DE CA TITA CA ALC CATA CATA TIT DA ALC CATA CAT TIC LACE ALA ALC CATA CAT TIC LACE ALA CIT CIT CIT ALC ALC CATA CAT CIT CIT CIT ALC ALC CIT CIT CIT ALC ALC CIT CIT CIT CIT ALC ALC CIT CIT CIT CIT ALC ALC CIT CIT ALC ALC CIT CIT ALC ALC CIT CIT ALC ALC CIT CIT ALC ALC CIT CIT CIT ALC ALC CIT CIT ALC ALC CIT CIT CIT CIT CIT CIT CIT CIT CIT CI		601 		6	21			641 				661 				681				701 					
100 170 180 170 721 74		ĠAT GGC Assp Gly	GAT TTA Asp Leu	CGA AAG C Arg Lys A	SĞ AGA GI 19 Arg G.	M TCT GA Lu Ser Gl	G GAT GTI u Asep Va	CGÁC LAsp	GAC GAT (Asp Asp /	GAT GTO Asp Val	CAGCA Ser L	WA ÀGG .ys Arg	GCA AG Ala Se	T CCT r Pro	CGT AAG Arg Lys	GGČ GAU	GAA C Glu P	CA GCI TO Ale	GGT CA	T ACT CIT s Thr Leu	AAG GA	C CTC P Leu	GCC CCC Ala Pro		
An AKE AKE AND CAT CT IT OT AKE AKE AND GOT CAE AKE AND COT CAE CAE ONE CTC DT AKE THE ATT THE THE COLOR OF THE AKE AKE ALS AND AKE ALS THE ATT THE AKE AKE ALS AND AKE AND AKE ALS AND AKE AKE ALS AND AKE AKE AKE ALS AND AKE AKE AKE AL		721		1 74	50 11			761		170		781				180 801				821	19	U			
Lin Am Intr am Inte Cell 200 Set Tis Amp Liy And Liy And Liy And Liy And Tet 2 and Amp Liu Lev Hil Amm The 16 Set U 10 20 12 120 120 120 110 110 120 100 110 11			ACC AAT	CAT CTT G	IT AGE A	TC GAT GG	T GCC GA		CAT CCA	GCC GAG	GAG	TC GTC	AAC TI	C ATC	TCG GGG	CAT TC	T CCG A	CT CG	CGT GC	G ACA GAC	AAC GA	T GCC	GCA GTT		
at1at1at1at1at1at1ct ct c		Gin Asn	Thr Asn	His Leu V	11 Ser 1. 20	Le Alep Gi	y Ala As	p Lys	MIS Pro	A18 A80 210) 610 L	.eu vai	ASN PT	e 11e	Ser GI	220	r Pro I	nr Arç) Arg Al	a Inr Asp	Ash As 23	0	ALS VEL		
See Map Sub See Lys Ang DD, Alls Ang Lys Ang Tye Wal Amen The New DD, Tye Pro DD, Alls Mei Dy DD, Wal Amen Tye DD, DD, Han Amen Tye DD, Han		841 		8			C 007 TA	881 	AAC ACC	ATC 004	. TAC C	901 		C 4CT		921				941	orc c4				
961 200 270 961 1001 1001 1001 1001 961 1001 1001 1001 1001 1001 1001 1001 1001 1001 1001 1001 1001 1001 1001 1001 1001 1001 1001 1001 1001 1001 1001 101 120 1001 101 1001 101 1001 101 1001 101 1001 101		Ser Asp	Asp Ser	Lys Arg G	ly Ala A	rg Lys Ly	s Arg Ty	r Val	Asn Thr	Het Gi	Tyr P	ro Glo	Ala Me	t Ser	Pro Gir	Met Gl	y Gly V	al Asr	Tyr Gl	y Gin Pro	Ala Gi	n Gin	Gly Tyr		
Dec CET CAN GOA ATC COT CAL CAT COL CAL COT CAL CAT CAL CAT THE CET CAL COL CAL THE CET CAL COL CAT THE CLY SET SET SET SET VIL AND 280 290 310 310 1081 1101 1121 1141 1161 1181 310 1084 CET CAL		961		2	40 81	• • •	+	1001	1	250		102	21			260 1041				1061	27	0			
200 200 200 310 1001 1121 1141 1161 1181 1004 1001 1121 1141 1161 1181 1005 1100 1121 1141 1161 1181 1006 1101 1121 1141 1161 1181 1007 1201 1241 1161 1161 1181 1201 1221 1241 1261 1261 1261 1261 1201 1221 1241 1261 1261 1301 300 <			CAA GGA	ATG GGT G	GA CCA G	TC GGA GG	T GGA CC	G ATG	GGT GGA		A CAA 1	TTT GGI Phe Giv	CCG T	A COT	CCA GG		A GAT	NCG GA	T TTT GC	A TCC AGO		C AGC	GTC GAG		
1061 1101 1121 1121 1141 1161 1181 1181 1				2	90			-		290	-		-	Ser		Glu - 300	-					10	- Va	í	
GCA GCT CAT ACA ACC ACC ACA LUT ATC CAC CAC ATC ALC CAC CAC ATC ALC CAC CAC ACC ATT CAC CT ACC ATT CAC CT ACC ATT CAC CT TAC CAT CTT GCA CTC TG CT ACC CT CT CAC CAT CTT GCA CTC TG CT ACC CT C		1081 		1	101 		(1	1121 1	I			114	11			1161 				1181 I					
Grin Liye 320 330 340 350 1201 1221 1221 1241 1261 1281 1301 CAC CTC AGE CAA CAC TC CTA ACC ACT TA CTA CTA CTA C		GGA GGI Gly Gly	GAT ACA	ACG ATC A Thr Ile S	GC GCC A er Alm A	GA UTT AT	G GAC GA	C ATC	AAG GCG	GTG CT Val Le	C GGT (u Gly /	GCC ACI	GAAGA	T GAC	CTA CC.	A GTA GA D Val As	CATC/	AAC GA		NC GAT CTI	GGA CT		CTA CGO		
1201 1221 1241 1261 1281 1301 CAC CTC AGE CAT CAC TCC AAC CTT CTA COG AAC ATC GET GGE GAT CGE GAA CAG GTC CTA ACC GCE ATC CGA GAG GAG GAG GAA CAA CAT GCA ACC GCE CCC 1301 1301 CAC CTC AGE CAT CAC TCC AAC CTT CTA COG AAC ATC GET GGE GAT CGE GAA CAG GT GGA ACA GCE GAA CAG GGA GAA GAA GAG GAA CAA CAT GCA ACC GCE CCC 300 300 300 320 320 320 300 300 300 300 300 1321 1341 1361 1301 1421 1401 1421 CAA CAA GGE GT TTG GAA CGE AC CCA GGE CAA GEA TTC GET GGE GGA GGT GGA GGT GGA GGE GGE CAT GA GGE GGE CAT GA AC CT CAA AGE ATC GGA AC CT CAA GGE ATE GGE GGA CCA CCT CAA GGE ATE GGE GAA CCA CCT CAA GGE ATE CGE GAA CCA CCT CAA GGE ATE CGE GAA CCA CCT CAA GGE ATE GGE GAA CCA CCT CAA GGE ATE GGE GAA CCA CCT CAA GGE ATE GGE GGA CCA CCT CAA GGE ATE CGE GGA GGE CAA CCA CCT CAA GGE AT		Glu Ly		3	20			-		330	-		-		• •	340	-				3			•	
CAC CYC AGE CAT CAC TCC AAC CYT CTA GOS AAC AYC GI' GAC GAC GCA GCA CAC GCA GCA GAA GAA GAG GAG		1201 		1	221 			1241 	1			120	51			1281 				1301 1					
360 370 380 390 1321 1341 1361 1381 1401 1421 CAA CAA CAC GCG CT TIG AAC GCG CAA CCA CCA CCA CCA CCA CCA CC		CAC CTO His Leu	C AGG CAT I Arg His	CAC TCC A His Ser A	AC CTT C an Leu L	TA GCG AM eu Ala Aa	NC ATC GG In Ile Gl	TGAC yAsp	CCT GCG Pro Ala	GTT CG Val Ar	G GAA (g Glu (CAG GTI Gin Val	CIAN Lleu S	GC GCG Pr Ala	ATG CA Met Gl	G GAG GA n Glu Gl	G GAA (u Glu (GAA GA Glu Gl	G GAG GA U Glu Gi	A GÁA GAI Lu Glu Asp	GCA GC	CA ACC La Thr	GGC GCC Gly Ala	: •	
121 1241 1241 1241 1421 CAA CAA GGC GTC TIG AAC GGA CGA CGA CGA CGC CAA GGC TIC GGT GGC GGA GGG GGC AGG GGC AGC AT CT CAA CGG ATG GGC GGA CAA CCT CAA GGG ATG ACT CGA GGG GGC GGA CGC GGA CGA CCT CAA GGG ATG ACT CGA GGG ATG ACT CGA GGG GGC GGA CAA CCT CAA GGG ATG ACT CGA GGG ATG ACT CGA GGG ATG ACT CGA GGG ATG ACT CAA GGG ATG GGC GGA CGA CCT CAA GGC ATG CGA CGG GGC GGA CCA CCT CAA GGC ATG CGA CGC CT CAA GGG ATG GGC GGA GGC GGA CCA CCT CAA GGC ATG CGA CGC CT CAA GGG ATG GGC GGA GGC CGA GGC CGA GGC CAA GGC ATG CGA CAT CAA GGG ATG GGC GGA CGA CCT CAA GGG ATG GGC GGA GGC GGA CGA CCT CAA GGG ATG GGC GGA G			• •	•••	60			-		370	-				• •	380	-				- 39	ю.			
CIN GIN GIY VAL LAW ANN GUY AAN ALB PPO GIY GIN ALL GUY PHO GIY				TTC AAC C		ra (ra (r				acr ac	A (CCT)		91 A CCC C		ATC AC		A . CAC	ATC . 00							
400 410 420 430 1441 1461 1481 1501 1521 1541 CCT CAA GGC ATG GGC THT CCA CAT GAA GGC ATG GGC GGA CCA CCC CAA GGC ATG GCC ATG CCA CAT CAA GGC ATG GCC ATG CCA CCT CAA GGC ATG CCA ATT TTG ATG TTA GTT ATG CTA ATG ATT ATT CTG AGA GTG GAG TTC CTT TAG 1561 1561 1601 1621 1641 1661 1681 1701 1721 1741 1761 1781 1681 1701 1721 1741 1761 1781 1681 1701 1721 1741 1761 1781 1681 1701 1721 1741 1761 1781 1681 1701 1721 1741 1761 1781 1681 1701 1721 1741 1761 1781 1681 1701 1721 1741 1761 1781 1681 1701 1721 1741 1761 1		Gin Gir	Gly Val	Leu Aan G	ly Asn A	la Pro Gi	ly Gln Al	a Gly	Phe Gly	Giy Gi	y Gly (Gly Gl	Gly A	la Met	Met Se	r Pro Gl	n Gin I	Met Gl	y Gly Gl	In Pro Gir	Gly M	st Ile	Gly Gl	1	
CT CAA GGG ATG GGC THT CCA TGAA GGG ATG GGC GGA CCA CCC CAA GGG ATG GGC ATG GGC ATG GGC GGA CCA CCA CCA CCA CCA CCA CGA GGG ATG GGC GGA CCA CCA CCA CCA CCA CCA CCA CC		1441		4	00 461			1481		410		15	n 1			420 1521				1541	43	50			
Pro GLIn GLY Met GLY Phe Pro His GLu GLY Met GLY GLY Pro Pro GLin GLY Met GLY GLY Pro Pro GLin GLY Met GLY GLY Pro Pro GLin GLY Met GLY Met Pro Pro GLin GLY Met Pro Pro GLin GLY Met GLY Met Pro Pro GLin GLY Met Pro Pro GLin GLY GLI Pro 440 450 460 460 470 1561 1561 1661 1661 1661 1661 1661 166			GGG ATG	GGC TTT C	CA CAT G	AA GGG AT	rg gac ga	A CCA	CCC CAA	GGG AT	G GGC /	ATG CC	A CAT C	MA GGG	ATG GG	C GGA CC	ACCT	CAA GG	G ATG GI	C ATG CC	A CCT C	AA GGC	G CAA CO	6	
470 460 470 1561 1581 1601 1621 1661 1661 TAC GGT CAG GGT TAT CTA CAG GGG TAG ATT GAT ATA GAT GAT GGT GGT AAT TTG GTG AAT TTG GTG ATT TGG GGG G		Pro Gir	Gly Met	Gly Phe P	ro His G	ilu Gly Me	et Gly Gl	y Pro	Pro Gin	Giy Me	t Gly I	Het Pr	o His G	in Giy	Met Gi	y Giy Pr	o Pro	Gin Gi	y Met G.	ly Met Pro	Pro G.	in Giy	Gin Pro	0	
TAC GCT CAC GCT TAT CTA CAG GGC TAG ATC TAA CTA GAT GAT GAT GAC GCT AAT TTG GTG TAT ATC GGG GTG ATC TCG AGA GCG TTT GCA AAA TAT ATT CTG AGA GTG GAG TTC TTC CTT TAG Tyr GJy GIA GIY Tyr Lew GIA GIY 980 1681 1701 1721 1741 1751 1781 AAC ACG AGT ATT TGA ATT TGA ATT TTC TTG TTC GGT GAG GTG CAA CAT CAT TAT AGC TAT TTA GTT AMA AMA AGG TGT TTC TTT TTG GCA AAT TTG ATA CAC GTA TTC ATT GTA 1801 1841 1841 1861 AAA AMA CCA ACA ACA ACA CCG ACA TGA AAT ATA AGC GAC GAT GAG AGA GAA AGA GGG AGG CCC AGA CTC G		1561		1	40 581			1601		450		16	21			460 1641				1661	4	70			
ABD 1681 1701 1721 1741 1761 1781 AAC ACG AGT ATT TGA ATT GAC TIT ICA ATT TIC TIG TIC GET GAG GIG CAA CAT CAT TAT AGC TAT TIA GIT AMA AMA AMA CGT TIC TIT TIG GCA AMT TIG ATA CAC GIA TIC ATT GTA 1801 1821 1841 1861 1801 1821 1841 1861 AMA AMA CCA ACA ACA CCG ACA TGA AMT ATA AGC GAC GAT AGA GAG AGA AMA GAG GGG AGG CCC AGA CTC G		TAC GGI	CAG GGT		AG GGG T	AG ATC T	M CTA GA	ATA TA	GAC GCT	AAT TT	G GTG	TAT AT	C GGG G	TG ATC	TCG AG		T GCA	ааа та	T ATT C	TG AGA GTI	GAG T	10 110	CTT TA	G	
1681 1701 1721 1741 1761 1781 AAC ACG AGT ATT TGA ATT GAC TTT TCA ATT TTC TTG TTC GGT GAG GTG CAA CAT CAT TAT AGC TAT TTA GTT ANA ANA ANA CGT TTC TTT TTG GCA AAT TTG ATA CAC GTA TTC ATT GTA 1801 1821 1841 1861 ANA ANA CCA ACA ACA CCG ACA TGA AAT ATA AGC GAC GAT AGA GAG AGA GAA AGA GGG AGG CCC AGA CTC G				4	80	-																			
AÀC ACG AGT ATT TGA ATT GAČ TIT TCA ATT TIC TTG TIC GĠT GAG GTG CAA CAT CAT TAT ÁGC TAT TTA GTT ANA ANA ANÀ CGT TIC TIT TIG GCA ANT TÌG ATA CAC GTA TIC ATT GTA 1801 - 1821 - 1841 - 1861 ANA ANA CCA ACA ACA CCG ACA TGA AAT ATA AGC GAC GAT ANA GAG AGA ANA GAG GGG AGG CCC AGA CIC G		1681 		1	701 I			1721 I	1			17	41			1761				1781					
ANA ANA CCA ACA ACA CCG ACA TGA AAT ATA AGC GAC GAT ANA GAG AGA GAA AAA GAG GGG AGG CCC AGA CTC G		AÁC ACC 1801	G AGT ATT	TGA ATT O	AC TTT 1 821	CA ATT T	IC TTG TI	C GÖT	GAG GTG	CAA CA	T CAT	TAT AG	C TAT T	TA GTT	MA M	a ana co	TTC TTC	111 11	G GCA A	AT TTG AT	CAC G	TA TTO	C ATT GT	A	
			CCA ACA	ACA CCG A	CA TGA A	IAT ATA A	SC GAC GA		GAG AGA	GAA AA	A GAG	0 000 AG	G CCC A	GA CTC	G										
	_	h	CCA ACA	ACA CCG A	CA TGA A	AT ATA A	SC GAC G/		GAG AGA	GAA AA	A GAG	GGG AG	G CCC A	GA CTC	G										

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 *Eco*RI sites at the ends of the cDNA derive from the *Eco*RI 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 3×10^8 cpm/µg 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 $poly(A)^+$ 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 ³²P-labeled cDNA probes. The molecular size markers were λ 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 $poly(A)^+$ mRNA, and displayed by NaDodSO₄ 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 $poly(A)^+$ 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)⁺ 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)+ 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)⁺ RNA, immunoprecipitated with bindin antibody in the presence of $5 \times 10^{-3} \mu g$ of unlabeled ovalbumin; lane E, total testes poly(A)⁺ RNA in vitro translation product prior to immunoprecipitation; lane F, egg poly(A)⁺ RNA in vitro translation product prior to immunoprecipitation; lane G, in vitro translation with egg poly(A)⁺ RNA, immunoprecipitated with normal serum; lane H, in vitro translation with egg poly(A)⁺ RNA, immunoprecipitated with anti-bindin antibody. The apparent molecular size of the polypeptide precipitated in the sample displayed in lane D is ≈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 NaDodSO4/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 probindin 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 is 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 malespecific 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

- Sp 26 GInGlyTyrGlyAlaCInGlyMetGlyGlyProValGlyGlyProMetGlyGlyProProGInPheGlyAla Sf - - [- - AlaPhe - -][GInGlyMet ClyAlaVal * Cly
- LeuProProGlyGlnAlaAspThrAspPheGlySerSerSerSerSerValAspClyGlyAspThrThr [ClnGlyNet Cly ValClyGlyCly] * Phe * AlaPhe ProGlyGluAlaGluAlaAsp Sp 51 Sf

FIG. 5. N-terminal sequences of S. franciscanus (Sf) and S. purpuratus (Sp) bindins. The Sp sequence is from Fig. 1B of this paper. The Sf sequence is from Vacquier and Moy (46). The bracket indicates the 10-amino acid motif that is repeated in the Sf sequence. Amino acids that were not determined are marked with an asterisk.

these species (40-42), and rare naturally occurring adult hybrids of these species have been reported (43). Thus the interspecific fertilization barrier rather than developmental incompatibility is probably a limiting factor in preserving the genetic separateness of these species. S. purpuratus sperm bindin fails to react with the receptor on S. franciscanus eggs and vice versa (6, 8, 10). The bindin proteins of these species are of the same molecular weight, and the sequence of the initial 73 amino acids of the mature S. franciscanus bindin has been obtained (46). A comparison between this sequence and that of the corresponding region of the S. purpuratus bindin is shown in Fig. 5. For the first 40 residues of the mature protein, the amino acid sequences are 80% homologous. However, between residues 31 and 59, the S. franciscanus sequence consists of three imperfect tandem repeats of a 10-amino acid sequence element, only the first of which is present in the S. purpuratus bindin at this location. Elsewhere in the S. purpuratus bindin molecule, sequences homologous to the first 5 amino acids of this repeat occur in eight additional locations. The repetition of this sequence motif might suggest that it is functionally important, and the differing organization of the same or similar repeats in the bindin molecules of the two species could contribute to the molecular basis for their different functional recognition specificities. The extent and nature of the distinctions in the primary sequences of the two proteins will be clarified when the S. franciscanus bindin gene is cloned and fully sequenced. The development of methods for inserting genes into the sea urchin genome (reviewed in ref. 44) provides the means to obtain more exact knowledge of the functional molecular changes that during evolution resulted in the genetic isolation of the populations ancestral to these two species.

We thank Dr. V. Vacquier for the generous gift of bindin protein and bindin antibody. We also thank Dr. S. J. Horvath for providing synthetic probes and primer, and Mr. J. A. Ferrier for technical assistance. B.G. was supported in part by the California Foundation for Biomedical Research. This research was supported by Grant DCB8502191 from the National Science Foundation.

- 1. Dan, J. C. (1967) in Fertilization, eds. Mertz, C. B. & Monroy, A. (Academic, New York), Vol. 1, pp. 237-293.
- 2. Summers, R. G., Hylander, B. L., Colwin, L. H. & Colwin, A. L. (1975) Am. Zool. 15, 523-551.
- 3. Podell, S. B. & Vacquier, V. D. (1985) J. Biol. Chem. 260, 2715-2718.
- 4. Summers, R. G. & Hylander, B. L. (1975) Exp. Cell. Res. 96, 63-68.
- 5 Segall, G. K. & Lennarz, W. J. (1979) Dev. Biol. 71, 33-48.
- Glabe, C., Buchalter, M. & Lennarz, W. J. (1981) Dev. Biol. 6. 84, 397-406.
- 7. Moy, G. W. & Vacquier, V. D. (1979) Curr. Top. Dev. Biol. 13, 31-44.

- 8. Glabe, C. G. & Vacquier, V. D. (1977) Nature (London) 267, 836-838.
- Vacquier, V. D. & Moy, G. W. (1977) Proc. Natl. Acad. Sci. USA 74, 2456-2460.
- 10. Vacquier, V. D. (1980) in The Cell Surface, Mediator of Developmental Processes, ed. Wessells, N. K. (Academic, New York), pp. 151-168.
- 11. Rossignol, D. P., Earles, B. J., Decker, G. L. & Lennarz, W. J. (1984) Dev. Biol. 104, 308-321.
- 12. Acevado-Duncan, M. & Carroll, E. (1986) Gamete Res., in press.
- 13. Watson, C. J. & Jackson, J. F. (1984) in DNA Cloning: A Practical Approach, ed. Glover, D. (IRL, Oxford), pp. 79-88.
- Huynh, T. V., Young, R. A. & Davis, R. W. (1984) in DNA 14. Cloning: A Practical Approach, ed. Glover, D. (IRL, Oxford), pp. 49-78.
- Posakony, J. W., Scheller, R. H., Anderson, D. M., Britten, 15. R. J. & Davidson, E. H. (1981) J. Mol. Biol. 149, 41-67.
- Kessler, S. W. (1975) J. Immunol. 115, 1617-1624. 16.
- 17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- Messing, J. (1983) Methods Enzymol. 101, 20-79. 18
- Henikoff, S. (1984) Gene 28, 351-359. 19.
- 20. Britten, R. J., Cetta, A. & Davidson, E. H. (1978) Cell 15, 1175-1186.
- Watson, M. E. E. (1984) Nucleic Acids Res. 12, 5145-5164.
- Bergmann, M. & Fruton, J. S. (1941) Adv. Enzymol. 1, 63-98. 22.
- 23. Kyte, J. & Doolittle, R. F. (1982) J. Mol. Biol. 157, 105-132.
- Greenwood, F. L., Hunter, W. M. & Glover, J. S. (1963) Biochem. J. 89, 114-123. 24
- Vacquier, V. D. (1983) Anal. Biochem. 129, 497-501. 25.
- Shott, R. J., Lee, J. J., Britten, R. J. & Davidson, E. H. 26. (1984) Dev. Biol. 101, 295-306.
- 27. Burgos, M. H. & Fawcett, D. W. (1955) J. Biophys. Biochem. Cytol. 1, 287.
- Burgos, M. H. & Fawcett, D. W. (1956) J. Biophys. Biochem. 28. Cytol. 2, 223-240.
- 29. Viebrock, A., Perz, A. & Sebald, W. (1982) EMBO J. 1, 565-571.
- 30. Chance, R. E., Ellis, R. M. & Bromer, W. W. (1968) Science 161, 165-166.
- Bornstein, P. (1974) Annu. Rev. Biochem. 43, 567-603. 31
- Kassell, B. & Kay, J. (1973) Science 180, 1022-1027. 32.
- 33. Levine, A. E., Walsh, K. A. & Fodor, E. J. B. (1978) Dev. Biol. 63, 299-306.
- 34. Csernansky, J. G., Zimmerman, M. & Troll, W. (1979) Dev. Biol. 70, 283-286.
- Conway, A. F. & Metz, C. B. (1976) J. Exp. Zool. 198, 39-48. 35
- Shyu, A., Raff, R. A. & Blumenthal, T. (1986) Proc. Natl. 36. Acad. Sci. USA 83, 3865-3869.
- 37. Strickland, M., Strickland, W. N., Brandt, W. F. & Von Holt, C. (1977) Eur. J. Biochem. 77, 263-275.
- Strickland, W. N., Strickland, M., de Groot, P. C., Von Holt, 38. C. & Wittmann-Liebold, B. (1980) Eur. J. Biochem. 104, 559-566.
- Busslinger, M. C. & Barberis, A. (1985) Proc. Natl. Acad. Sci. 39. USA 82, 5676-5680.
- 40. Loeb, J., King, R. & Moore, A. R. (1910) Wilhelm Roux' Arch. Entwicklungsmech. Org. 29, 354-362. Chaffee, R. R. & Mazia, D. (1963) Dev. Biol. 7, 502-512.
- 41.
- Barrett, D. & Angelo, G. M. (1969) Exp. Cell Res. 57, 159-166. 42.
- Swan, E. (1953) Evolution 7, 269-273. 43.
- 44. Davidson, E. H., Flytzanis, C. N., Lee, J. J., Robinson, S. J., Rose, S. J., III, & Sucov, H. M. (1985) Cold Spring Harbor Symp. Quant. Biol. 50, 321-328.
- 45. Lee, J. J., Shott, R. J., Rose, S. J., III, Thomas, T. L., Britten, R. J. & Davidson, E. H. (1984) J. Mol. Biol. 172, 149-176
- Vacquier, V. D. & Moy, G. W. (1978) in Cell Reproduction, 46. eds. Dirksen, E. R., Prescott, D. M. & Fox, C. F. (Academic, New York), pp. 379-389.