Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 4957–4961, May 1995
Evolution

Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa

(invertebrate fertilization/sperm–egg interaction/nonsynonymous nucleotide substitution/protein evolution/acrosome reaction)

WILLIE J. SWANSON AND VICTOR D. VACQUIER

Marine Biology Research Division and Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093-0202

Communicated by Roy J. Britten, California Institute of Technology, Corona del Mar, CA, March 3, 1995 (received for review November 15, 1994)

ABSTRACT **During fertilization in marine invertebrates, fusion between sperm and egg cell membranes occurs at the tip of the sperm acrosomal process. In abalone sperm the acrosomal process is coated with an 18-kDa protein. *In situ*, this protein has no effect on the egg vitelline envelope, but *in vitro* it is a potent fusagen of liposomes. Thus, the 18-kDa protein may mediate membrane fusion between the gametes, a step in gamete recognition known to restrict heterospecific fertilization in other species. The cDNA and deduced amino acid sequences of the 18-kDa protein were determined for five species of California abalone. The deduced amino acid sequences exhibit extraordinary divergence; the percent identity varies from 27% to 87%. Analysis of nucleotide substitution shows extremely high frequencies of amino acid-altering substitution compared to silent substitution, demonstrating that positive Darwinian selection promotes the divergence of this protein. However, amino acid replacement is conservative with respect to size and polarity of residue. The data support the developing idea that in free-spawning marine invertebrates, the proteins mediating fertilization may be subjected to intense, and as yet unknown, selective forces. The extraordinary divergence of fertilization proteins may be related to the establishment of barriers to heterospecific fertilization.**

In most marine invertebrates with external fertilization, sperm–egg interaction exhibits some degree of species selectivity—that is, sperm fertilize conspecific eggs more effectively than heterospecific eggs (1–4). Mechanistically, the species selectivity of fertilization could occur at one or more recognition steps in the fertilization cascade, including attachment of the sperm to the egg envelope, induction of the sperm acrosome reaction, interaction of the acrosomal proteins with the egg envelope, and membrane fusion between the two cells. The sperm proteins mediating gamete recognition are located on the sperm plasma membrane and in the acrosomal granule, which undergoes exocytosis when sperm contact the egg's extracellular matrix.

Spermatozoa of the abalone (*Haliotis*; Archeogastropoda) possess a large acrosomal granule (5) containing roughly equal quantities of proteins of 16 kDa and 18 kDa (6). The 16-kDa protein is located in the proximal portion of the acrosomal granule, whereas the 18-kDa protein is stored in the distal part of the granule at the most anterior tip of the cell, directly anterior to the partially polymerized acrosomal process (7). Exocytosis of the acrosomal granule releases both proteins onto the surface of the egg vitelline envelope (VE). The 16-kDa protein, termed lysin, creates a hole in the VE by a species-selective, nonenzymatic mechanism (6, 8–11). The function of the 18-kDa protein in fertilization remains uncertain. However, this protein coats the acrosomal process of acrosome-reacted sperm and is a potent fusagen of liposomes (unpublished results). In this paper, the cDNA and deduced amino acid sequences of the 18-kDa protein are analyzed from five species of California abalone.*

MATERIALS AND METHODS

Protein Isolation. A supernatant of acrosomal exudate of the red abalone, *Haliotis rufescens*, was obtained as described (6) and dialyzed against 250 mM NaCl/10 mM Mes, pH 6.0. The 20,000 × *g* supernatant of the dialysate was applied to a 60-ml (2.5 × 12 cm) column of CM-cellulose, and the column was washed with 2 liters of 435 mM NaCl/10 mM Mes, pH 6.0, to completely elute the 16-kDa lysin. The NaCl was increased to 1.0 M and 5-ml fractions were collected. Analysis by reducing and denaturing polyacrylamide gel electrophoresis and silver staining showed that the peak fractions contained the purified 18-kDa protein. The protein was dialyzed into distilled water and lyophilized, and 1.5 nmol was subjected to gas-phase amino acid sequencing, which unambiguously yielded the sequence of the first 41 residues.

cDNA Sequences. mRNA was isolated from abalone testes as described (8, 12) and stored in 2-μg aliquots in 80% ethanol at −20°C. cDNA was synthesized and dG-tailed with terminal transferase (12). A degenerate oligonucleotide primer representing amino acids 34–39, in combination with oligo(dC) (21-mer), was used to amplify a portion of the cDNA by PCR. The sequence of the PCR product confirmed the known amino acid sequence of the 18-kDa protein from *H. rufescens*. An oligonucleotide 5′ to the open reading frame (GATTCGAGGACGGATAGATTC), in combination with oligo(dT), was used to amplify the full-length cDNA from *H. rufescens* and *Haliotis sorenseni*. Additional primers were synthesized based on conserved regions of these two species and were used to obtain PCR products from the three other species. Both strands of cDNA were sequenced twice as single-strand templates (13). A Northern blot was performed by standard methods (14) using 1 μg of poly(A)$^+$ RNA; it yielded a single band of hybridization at 700 bp.

Analysis. The cDNA and deduced amino acid sequences were aligned (15), and the number of nonsynonymous substitutions per 100 nonsynonymous sites (amino acid altering; Dn) and the number of synonymous substitutions per 100 synon-

Abbreviations: VE, egg vitelline envelope; Dn, number of nonsynonymous substitutions per 100 nonsynonymous sites; Ds, number of synonymous substitutions per 100 synonomous sites; PNC, number of conservative nonsynonymous substitutions per 100 nonsynonymous sites; PNR, number of radical nonsynonymous substitutions per 100 nonsynonymous sites.

*The sequences reported in this paper have been deposited in the GenBank data base (accession nos. L36552–L36554, L36589, and L36590).

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.

ymous sites (silent; Ds) were determined by the method of Nei and Gojobori (16). The SCR program (17) was used to calculate the number of nonsynonymous substitutions per 100 nonsynonymous sites that are radical (PNR) versus those that are conservative (PNC) using the amino acid classification of Miyata *et al.* (18). The nucleotide sequences of myoglobins from two species of Japanese abalone, *Haliotis diversicolor* (synonym, *Suculus diversicolor*) and *Haliotis sieboldii* (synonym, *Nordotis madaka*), have been reported (19, 20). Codon usage bias was determined by the scaled χ^2 method (21, 22). Homology searches were performed for each mature 18-kDa protein using the BLAST program (23). Statistical homology of the five 18-kDa protein sequences was demonstrated using the RDF2 program (500 randomizations, Ktup = 2; ref. 24). The Genetics Computer Group program PROFILEMAKE (25) was used to obtain a profile representing the five mature sequences. This profile was used to search the Genpept Library (Release 83) for homologous sequences using PROFILESEARCH (25). These programs were accessed using the suite of programs contained in DNASYSTEM (26). Secondary structure predictions were performed using the five mature protein sequences as input for the program PHDSEC (27). The 16-kDa protein (lysin) secondary structure was calculated using the seven sequences deposited in GenBank (8). A three-dimensional compatibility search (28) was done as described in the Biosym INSIGHT program (29), using the 16-kDa lysin crystal structure for comparison (10).

RESULTS

Sequences of the 18-kDa Proteins. Amino-terminal amino acid sequences were determined by gas-phase sequencing of isolated 18-kDa proteins of *H. rufescens* (residues 1–41) and *Haliotis corrugata* (the first five residues; Fig. 1). The signal sequences (positions −17 to −1) vary from 17 to 19 residues and are typical of eukaryotes (30). The mature proteins vary in length from 132 to 146 residues. Two cysteine residues in the mature proteins (positions 60 and 131) are conserved in all five species. The five proteins are highly basic, with isoelectric points varying from 10.3 (*H. corrugata*) to 10.7 (*H. sorenseni*). There are many dyads of the same amino acid in these sequences (for example, *H. rufescens* positions 42–51 and 68–75). In *H. rufescens* and *H. sorenseni*, there are 16 dyads involving 32 amino acids, which represents 24% of the mature sequences. Amino acids with aromatic side chains (tyrosine, tryptophan, and phenylalanine) are conserved in many positions in four out of five sequences.

In contrast to the conserved features, the divergence of these orthologous proteins is extraordinarily high. For example, the mature *H. corrugata* and *H. fulgens* amino acid sequences are only 27% identical (Table 1). The *H. corrugata* sequence has an additional six residues at the amino-terminal end and seven residues at the carboxyl-terminal end (Fig. 1). Despite the extreme divergence, the five sequences are homologous; all pairwise alignments have *Z* scores of at least 18 standard deviation units above the randomized alignments (31). There are several positions where a negatively charged residue is replaced by a positively charged residue and vice versa. There are no repetitive elements in any sequence. A GenBank search of each mature sequence yielded no significant similarity to other known sequences.

The extreme divergence of the coding regions contrasts with high conservation in both the 5′ and 3′ untranslated regions of the five cDNAs. For example, the *H. corrugata* and *H. rufescens* nucleotide sequences are 95% identical for 45 bp (one gap) of the 5′ untranslated region and 87% identical for 70 bp (zero gaps) of the 3′ untranslated region. However, in the open reading frames, the 18-kDa proteins of these two species are only 48% identical for 396 bp (five gaps).

Nucleotide Substitution. The values of Dn and Ds were calculated for all 10 pairwise comparisons (Table 1) using the aligned mature coding regions of 132 codons. Dn exceeds Ds in 5 of the 10 comparisons. Three of these 5 comparisons, those among the most closely related sequences, are statistically significant ($P \leq 0.005$). The ratios of Dn to Ds of 4.50 and 4.67 are some of the highest known for a full-length protein. Further analysis showed that the nonsynonymous substitutions are not random. Rather, when residue replacement occurs, there is conservation with respect to size and polarity. In all 10

```
   -17                  -1+1             10        20         31
     ** * *****        * | |          *   |         | **       |
Hr   MRSLVLLCVLL----MAICAA DK------KSTVSKENAAAMKVAMIKFLDSRTDRFKK
Hs   ...........----...... ..------.T...........I.......A.AGK...
Ha   ...........----...... ..------.TS.....E.......M....MKAGV..E
Hc   ..F.L.....MGAVSQ.V.-- R.RPNVWG.IV.KEK.K....IGFMEY..AKLVK..R
Hf   ..........M----AVG.V. FD------DVV..RQEQSYVQRG.VN...EEMHKLV.

             40        50        60        70        80       89
              |         |    *    *      *  |        *|      **|
Hr   R-IEKIGYPITPPQYTTLLYYNRERLMDWCHNYVEVSKKIILLGGNKLNKKNFARMGRI
Hs   .-V.NM........W........Q...E...T...F......M..........T.....
Ha   I-..DM........W...........IEF.RSFLAL..............A........
Hc   HWLVGANWKLQKFETDEMR.LAIK..IKV..G.TIW.QRL.M.KYRP..E.Y.KKV..Y
Hf   .-FRDMRWNLGPGFVFL.KKV....M.RY.MD.ARY....LQ.KHLPV...TLTK...F

    90       100           110       120       130
     |         |             |         |         **
Hr   IGWKNQWILKRRQWHM----VRVMRRYKASAIAKKIVAMKVADLPCN 132
Hs   .......V......E.----........ST................. 132
Ha   .L..S..AVRQ...G.----...S..HTST....R............ 132
Hc   LA.R.-YLIVF.M.IGVL--KKNLK.SEITKPMQ.LLDT.DGE...PVRKIHG 146
Hf   V.YR.-YGVI.ELYADVFRD.QGF.GP.MT.AMR.YSSKDPGTF..KNEKRRG 141
```

FIG. 1. Aligned amino acid sequences (single-letter code) of the 18-kDa acrosomal proteins from five species of California abalone. Species names, abbreviated names, and GenBank accession numbers are *H. rufescens*, Hr, L36552; *H. sorenseni*, Hs, L36553; *Haliotis assimilis*, Ha, L36554; *H. corrugata*, Hc, L36590; and *Haliotis fulgens*, Hf, L36589. Sequences are listed in order of similarity to the *H. rufescens* sequence. Dots denote identity to the *H. rufescens* sequence and dashes denote gaps. An extra space separates the signal sequence from the mature protein. Asterisks mark positions that are identical in all sequences. Numbering refers to the *H. rufescens* sequence. Numbers after the carboxyl termini give the lengths of the mature proteins.

Table 1. Pairwise comparison of the percent identity of the mature 18-kDa proteins, Ds, Dn, PNC, and PNR

Species comparison	% identity	Ds	Dn	Dn/Ds	PNC	PNR	PNC/PNR
Hr–Hs	86.6	1.8 ± 1.7	8.1 ± 1.7	4.50*	17.0 ± 3.8	3.3 ± 1.2	5.15*
Hr–Ha	75.2	4.9 ± 2.5	18.5 ± 2.7	3.78*	26.5 ± 4.4	11.5 ± 2.2	2.30*
Hr–Hc	31.1	87.0 ± 17.5	82.6 ± 8.6	0.95	71.1 ± 4.7	41.0 ± 3.4	1.73*
Hr–Hf	33.8	78.2 ± 15.4	81.1 ± 8.5	1.04	62.2 ± 5.0	43.9 ± 3.4	1.42*
Hs–Ha	77.2	3.1 ± 1.9	14.5 ± 2.3	4.67*	22.0 ± 4.1	9.0 ± 2.0	2.44*
Hs–Hf	35.2	83.5 ± 16.8	76.9 ± 8.0	0.92	65.1 ± 4.9	40.6 ± 3.4	1.60*
Hs–Hc	33.1	78.0 ± 15.5	75.6 ± 7.8	0.97	58.1 ± 5.1	42.1 ± 3.4	1.36*
Ha–Hc	31.1	92.2 ± 18.6	82.3 ± 8.8	0.89	63.5 ± 5.0	44.8 ± 3.5	1.42*
Ha–Hf	35.2	84.0 ± 16.6	85.5 ± 9.0	1.02	63.0 ± 4.9	45.3 ± 3.5	1.39*
Hc–Hf	26.9	114.2 ± 23.8	86.9 ± 8.9	0.76	66.4 ± 3.3	45.1 ± 3.3	1.47*

Hr, *H. rufescens*; Hs, *H. sorenseni*; Ha, *H. assimilis*; Hc, *H. corrugata*; Hf, *H. fulgens*.
*Significant at the $P < 0.005$ level.

pairwise comparisons, PNC is statistically greater than PNR (Table 1).

Because the two abalone sperm acrosomal proteins (16-kDa lysin, ref. 8; and the 18-kDa protein, Table 1) show high values of Dn compared to Ds, the possibility existed that comparison of any protein from two abalone species may show high ratios of Dn to Ds. Nucleotide sequences are known for myoglobins from two species of Japanese abalone, *H. diversicolor* (19) and *H. sieboldii* (20). The myoglobins of these two species align perfectly for 377 amino acids (87% identity). Calculations of nucleotide substitutions in these myoglobins yielded a Dn of 7.1 ± 0.9 and a Ds of 52.8 ± 6.0; the ratio of Dn to Ds is 0.13. Thus, as found with most proteins, Ds greatly exceeds Dn in abalone myoglobins.

The rate of Ds can be suppressed by either an abnormally high or low percentage of G+C in the third codon position (32–35) and by codon usage bias (21, 22). The percentage of G+C for the open reading frames of the five 18-kDa proteins (Table 2) varies from 40.6 to 45.3. The percentage of C in the third codon position varies from 17.9 to 20.8. Nucleotide usage bias (ref. 36; 0 = no bias and 1.0 = maximum bias) for all three codon positions varies from 0.07 to 0.12, and for the third codon position, it varies from 0.063 to 0.13. These values of nucleotide usage bias are considered low (36). Codon usage bias (0.0 = no bias and 1.0 = maximum bias) was calculated by the scaled χ^2 method (21, 22); the values vary from 0.10 to 0.19 (Table 2), which are also considered low. The values for both indices of bias for the 18-kDa protein are lower than those for abalone tropomyosin, myoglobin, and chymotrypsin (37). These data suggest that nucleotide and codon usage bias cannot account for the high ratios of Dn to Ds and thus do not explain the robust demonstration of positive Darwinian selection in the divergence of the 18-kDa acrosomal protein.

Relatedness of the 16-kDa and 18-kDa Proteins. Proteins homologous to the 18-kDa protein were searched for by using programs utilizing the evolutionary information contained in the five 18-kDa amino acid sequences. An 18-kDa profile, a position-specific scoring table representing a group of homologous sequences, was calculated and used to search a data base (PROFILESEARCH; ref. 25). The results were remarkable; all seven 16-kDa abalone lysin sequences (8) ranked as the top seven highest scoring matches to the 18-kDa profile. The *Z* scores ranged from 5.7 to 9.4 standard deviations above the mean, indicating likely homology between these two acrosomal proteins.

To test if the two acrosomal proteins were related, the predicted secondary structure (PHDSEC; ref. 27) of the 18-kDa protein was compared to the known crystal structure of 16-kDa lysin (10). The utility of the program was demonstrated by predicting the secondary structure of the 16-kDa protein and then comparing it to its known crystal structure (Fig. 2). The prediction of secondary structure for the 16-kDa protein (position and length of α-helices) is in excellent agreement with that of its known crystal structure. The secondary structure prediction of the 18-kDa protein closely matches that of the 16-kDa protein (Fig. 2). It is unlikely that the positions and lengths of the 18-kDa predicted α-helices would match the known 16-kDa lysin α-helices by chance alone.

A three-dimensional compatibility search was performed to determine if the tertiary structure of the 18-kDa protein was similar to that of the 16-kDa protein (28). The best comparison was between the *H. rufescens* 16-kDa lysin crystal structure profile and the *H. assimilis* 18-kDa protein, yielding a *Z* score of 6.92 standard deviations for an alignment of 87 residues. Proteins with a *Z* score ≥7 usually have the same general fold (28). The data suggest it is probable that both acrosomal proteins have a similar tertiary structure, which suggests evolutionary homology of the two proteins.

DISCUSSION

In this paper the sequences of the 18-kDa acrosomal protein from five species of abalone are presented. The conserved features show that they are orthologous proteins; however, they exhibit extraordinary divergence. Analyses of nucleotide substitutions show that their divergence has been promoted by positive Darwinian selection. However, there has been a tendency to conserve the size and polarity of residues. Further analyses indicate that the 18-kDa and 16-kDa lysin proteins

Table 2. Nucleotide and codon bias of the 18-kDa proteins

Species	% G+C	Third position		Nucleotide bias	Third position nucleotide bias	Codon bias
		% G+C	% C			
Hr	41.8	52.7	20.0	0.11	0.12	0.16
Hs	42.1	52.7	19.3	0.12	0.12	0.18
Ha	45.3	54.0	19.3	0.09	0.13	0.10
Hc	40.6	40.6	17.9	0.07	0.13	0.19
Hf	49.7	49.7	20.8	0.11	0.06	0.12

Hr, *H. rufescens*; Hs, *H. sorenseni*; Ha, *H. assimilis*; Hc, *H. corrugata*; Hf, *H. fulgens*.

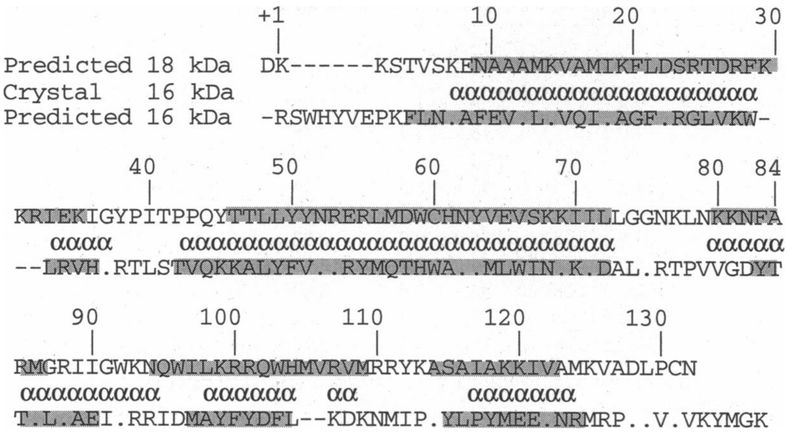
+1 10 20 30
Predicted 18 kDa DK------KSTVSKENAAAMKVAMIKFLDSRTDRFK
Crystal 16 kDa ααααααααααααααααααααα
Predicted 16 kDa -RSWHYVEPKFLN.AFEV.L.VQI.AGF.RGLVKW-
40 50 60 70 80 84
KRIEKIGYPITPPQYTTLLYYNRERLMDWCHNYVEVSKKIILLGGNKLNKKNFA
αααα ααααααααααααααααααααααααααααα ααααα
--LRVH.RTLSTVQKKALYFV..RYMQTHWA..MLWIN.K.DAL.RTPVVGDYT
90 100 110 120 130
RMGRIIGWKNQWILKRRQWHMVRVMRRYKASAIAKKIVAMKVADLPCN
ααααααααα αααααα αα ααααααα
T.L.AEI.RRIDMAYFYDFL--KDKNMIP.YLPYMEE.NRMRP..V.VKYMGK

FIG. 2. Comparisons of the predicted secondary structures for both abalone sperm acrosomal proteins to the known secondary structure of the 16-kDa lysin (10). Predicted α-helices are shaded. No β-sheets were predicted. Only those residues where the reliability of the prediction was >78% are shaded (27). "Crystal 16 kDa" shows the known α-helical domains of the *H. rufescens* lysin represented by ααα. Numbering refers to the *H. rufescens* 18-kDa sequence (Fig. 1), dots in the 16-kDa sequence denote identity to the 18-kDa sequence, and dashes are inserted for alignment.

have similar tertiary structures and therefore probably arose by gene duplication.

Previous work (38) with a Japanese abalone, *Haliotis discus*, suggested that the 18-kDa protein dissolves an outermost zone of the egg VE and that the actions of both the 18-kDa and 16-kDa acrosomal proteins are required to dissolve the VE. Eggs of *H. rufescens* with intact jelly layers and VEs were exposed to purified 16-kDa and 18-kDa proteins for time periods of hours and microscopically assessed for any effect on the VE. When present alone, the 18-kDa protein (1 mg/ml) had no effect on the integrity of the egg jelly layer or VE. Mixtures of both purified proteins were no more effective than the 16-kDa protein alone in solubilizing egg VEs. Thus, we were unable to demonstrate that the 18-kDa protein altered the integrity of the *H. rufescens* egg VE. It could be argued that the 18-kDa protein was denatured during purification, but the purified protein crystallizes, suggesting that it is native (W.J.S., unpublished results).

The interspecific divergence of both the 16-kDa and 18-kDa abalone sperm acrosomal proteins has been promoted by positive Darwinian selection, suggesting there is strong adaptive value to alter the sequences of these proteins between species. Gene genealogies of both proteins show identical branching patterns (8), indicating it is likely that both proteins share a similar evolutionary history. However, interspecific divergence of the 18-kDa protein is 2–3 times greater than that of the 16-kDa lysin (8, 9). The Dn to Ds ratios of the 18-kDa protein are some of the highest known for a full-length protein. The decrease in the Dn to Ds ratios with increasing divergence has been previously observed and is probably caused by functional constraints on the protein (39). The sequences presented in Fig. 1 represent single individuals from each species; therefore, there is no information on sequence variation within a species. The 16-kDa lysin was previously shown to exhibit species selectivity in the dissolution of isolated egg VEs, suggesting that it is a component in the block to heterospecific fertilization in abalones (9). Demonstrating the species selectivity of 16-kDa lysin is easy due to an unambiguous, rapid, and quantitative assay for VE dissolution (9). Unfortunately, there is no such assay to demonstrate directly that the 18-kDa protein mediates gamete fusion.

The adaptive value underlying the interspecific positive selection on the 18-kDa protein might be in the maintenance of species-selective sperm–egg fusion at fertilization. Several proteins that mediate sperm–egg fusion are known. Bindin, the acrosomal protein of sea urchin sperm, coats the acrosomal process (40) and is a fusagen of liposomes (41). Recordings of sea urchin egg membrane potentials suggest that membrane fusion with sperm may exhibit species selectivity (2). After removal of the sea urchin egg coat, fusion with sperm can still exhibit species selectivity (42). PH-30, a protein from guinea pig sperm, is involved in gamete fusion (43), and the species selectivity of gamete fusion has been documented in several mammalian species (43–45).

Significant similarity between the two abalone acrosomal proteins was demonstrated at the primary (PROFILESEARCH), secondary (Fig. 2), and tertiary structural levels, suggesting that they are homologous proteins that arose by gene duplication. Subsequent divergence of the duplicated genes may have been promoted by intraspecific selective pressure to specialize the functions of both proteins (46). Hypothetically, the ancestral acrosomal protein had two functions, to dissolve a hole in the VE and to mediate membrane fusion. At present, the 16-kDa lysin dissolves a hole in the VE, yet still retains the vestigial ability to fuse membranes (47). The 18-kDa protein has no effect on the VE and fuses liposomes at a faster rate and to a greater extent than does 16-kDa lysin (unpublished results). The intraspecific divergence of 16-kDa and 18-kDa proteins is too great to test for positive selection.

Few examples are known of interspecific positive Darwinian selection at the molecular level. Most involve immune defense and host–pathogen recognition by cell surface proteins (9). The nature of the selective pressure acting on the interspecific divergence of abalone acrosomal proteins remains unknown. However, the adaptive value in utilizing species-selective, cell surface recognition proteins to restrict heterospecific fertilization could in theory provide an explanation.

We thank Drs. M. E. Hellberg, R. S. Burton, J. D. Calkins, and J. M. Swanson for reviewing the manuscript and D. Geiger for identification of species. This work was supported by National Institutes of Health Grant HD 12986 to V.D.V.

1. Minor, J. E., Fromson, D. R., Britten, R. J. & Davidson, E. H. (1991) *Mol. Biol. Evol.* **8**, 781–795.
2. Metz, E. C., Kane, R. E., Yanagimachi, H. & Palumbi, S. R. (1994) *Biol. Bull.* **187**, 23–34.
3. Lopez, A., Miraglia, S. J. & Glabe, C. G. (1993) *Dev. Biol.* **156**, 24–33.
4. Summers, R. G. & Hylander, B. L. (1976); *Exp. Cell Res.* **100**, 190–194.
5. Lewis, C. A., Leighton, D. L. & Vacquier, V. D. (1980) *J. Ultrastruct. Res.* **72**, 39–47.
6. Lewis, C. A., Talbot, C. F. & Vacquier, V. D. (1982) *Dev. Biol.* **92**, 227–240.
7. Haino-Fukushima, K. & Usui, N. (1986) *Dev. Biol.* **115**, 27–34.
8. Lee, Y.-H. & Vacquier, V. D. (1992) *Biol. Bull.* **182**, 97–104.
9. Vacquier, V. D. & Lee, Y.-H. (1993) *Zygote* **1**, 181–196.

10. Shaw, A., McRee, D. E., Vacquier, V. D. & Stout, C. D. (1993) *Science* **262,** 1864–1867.
11. Shaw, A., Lee, Y.-H., Stout, C. D. & Vacquier, V. D. (1994) *Semin. Dev. Biol.* **5,** 209–215.
12. Lee, Y.-H. & Vacquier, V. D. (1992) *Anal. Biochem.* **206,** 206–207.
13. Lee, Y.-H. & Vacquier, V. D. (1993) *BioTechniques* **14,** 191–192.
14. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
15. Feng, D.-F. & Doolittle, R. F. (1990) *Methods Enzymol.* **183,** 375–387.
16. Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3,** 418–426.
17. Hughes, A. L., Ota, T. & Nei, M. (1990) *Mol. Biol. Evol.* **7,** 515–524.
18. Miyata, T., Miyazawa, S. & Yasanaga, T. (1979) *J. Mol. Evol.* **12,** 219–236.
19. Suzuki, T. & Takagi, T. (1992) *J. Mol. Biol.* **228,** 698–700.
20. Suzuki, T. (1994) *J. Protein Chem.* **14,** 9–13.
21. Shields, D. C., Sharp, P. M., Higgins, D. G. & Wright, F. (1988) *Mol. Biol. Evol.* **5,** 704–716.
22. Moriyama, E. N. & Hartl, D. L. (1993) *Genetics* **134,** 847–858.
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
24. Pearson, W. R. (1990) *Methods Enzymol.* **183,** 63–98.
25. Genetics Computer Group (1991) *Program Manual for the GCG Package* (Univ. of Wisconsin, Madison), Version 7.
26. Smith, D. W. (1988) *Comput. Appl. Biosci.* **4,** 212.
27. Rost, B. & Sander, C. (1994) *Proteins* **19,** 55–72.
28. Bowie, J. U., Luthy, R. & Eisenberg, D. (1991) *Science* **253,** 163–170.
29. *Biosym Profiles 3D User Guide* (Biosym Technologies, San Diego), Version 2.3.0.
30. von Heijne, G. (1985) *J. Mol. Biol.* **184,** 99–105.
31. Doolittle, R. F. (1987) *Of URFS and ORFS: A Primer on How to Analyse Derived Amino Acid Sequences* (University Science Books, Mill Valley, CA).
32. Sharp, P. M. & Li, W.-H. (1987) *Mol. Biol. Evol* **4,** 222–230.
33. Ticher, A. & Graur, D. (1989) *J. Mol. Evol.* **28,** 286–298.
34. Wolfe, K. H., Sharp, P. M. & Li, W.-H. (1989) *Nature* (*London*) **337,** 283–285.
35. Moriyama, E. N. & Gojobori, T. (1992) *Genetics* **130,** 855–864.
36. Irwin, D. M., Kocher, T. D. & Wilson, A. C. (1991) *J. Mol. Evol.* **32,** 128–144.
37. Lee, Y.-H. (1994) Ph.D. thesis (Univ. of California, San Diego).
38. Usui, N. & Haino-Fukushima, K. (1991) *Mol. Reprod. Dev.* **28,** 189–198.
39. Hughes, A. L. & Nei, N. (1988) *Nature (London)* **335,** 167–170.
40. Moy, G. W. & Vacquier, V. D. (1979) *Curr. Top. Dev. Biol.* **13,** 31–43.
41. Glabe, C. G. (1985) *J. Cell Biol.* **100,** 800–806.
42. Aketa, K. (1982) *Cell Differ.* **11,** 277–278.
43. Myles, D. G. (1993) *Dev. Biol.* **158,** 35–45.
44. Yanagimachi, R. (1988) *Curr. Top. Membr. Transp.* **32,** 3–43.
45. Roldan, E. R. S. & Yanagimachi, R. (1989) *J. Exp. Zool.* **250,** 321–328.
46. Hughes, A. L. (1994) *Proc. R. Soc. London B* **256,** 119–124.
47. Hong, K. & Vacquier, V. D. (1986) *Biochemistry* **25,** 543–550.