Concerted Evolution in an Egg Receptor for a Rapidly Evolving Abalone Sperm Protein

Willie J. Swanson* and Victor D. Vacquier

Gamete interactions during fertilization exhibit species specificity. In abalone, the sperm protein lysin species-specifically creates a hole in the egg envelope. Lysin evolves rapidly by positive Darwinian selection. Evolution of the egg receptor for lysin provides the selective pressure for lysin's divergence. The egg receptor for lysin is a tandemly repeated sequence that evolves by concerted evolution. Concerted evolution in the egg receptor could explain the rapid, adaptive evolution in sperm lysin and may provide an underlying molecular mechanism that gives rise to species-specific fertilization.

In abalone, gamete recognition is mediated by lysin, a sperm protein that species-specifically creates a hole in the egg envelope (*1*). Since the discovery that gamete recognition proteins undergo rapid, adaptive evolution, there have been numerous hypotheses about the selective forces that promote their divergence (*2*). The rapid evolution of gamete recognition proteins could lead to species-specific fertilization. Steps in sperm-egg interaction where heterospecific fertilization is blocked are known in mammals (*3*) and invertebrates (*4*); however, the molecular evolutionary process responsible for the species specificity is unknown. Here we report the cDNA sequences that encode the tandemly repeated motif of the vitelline envelope receptor for lysin (VERL) from seven abalone species and we present a hypothesis to account for the rapid evolution of lysin and species-specific fertilization.

Abalones are marine mollusks that spawn gametes into seawater where fertilization occurs. In California, seven species of abalone overlap in habitat and breeding season, yet natural hybrids are rare (*5*). Abalone eggs are surrounded by a vitelline envelope (VE), which the sperm must penetrate to fuse with the egg. Abalone sperm use 16-kD lysin to nonenzymatically and species-specifically create a hole in the VE (*1*). Analyses of nucleotide substitutions in lysins from different abalone species show that positive Darwinian selection rapidly diversifies lysin (*1*).

VERL is a long, unbranched glycoprotein and it is the only lysin binding component in the VE. Binding curves show species specificity and positive cooperativity with a median effective concentration of approximately 9.5 nM. Each molecule of VERL binds about 60 molecules of lysin, which indicates that VERL contains a repetitive lysin binding motif (*6*). VERL was fragmented and randomly chosen peptides were sequenced (*7*). Portions of the same amino acid sequence motif were obtained from 14 peptides of different sizes, which indicates that VERL is a repeating protein sequence. Polymerase chain reaction (PCR) primers were designed from the peptide sequences and used for amplification (*8*).

Agarose gel analysis of PCR products contained ladders of DNA bands differing by ~500 base pairs (bp). When these bands were sequenced, the VERL repeat was found to be 459 bp (153 amino acids). The translated repeat sequence contained the 14 peptide sequences obtained from VERL fragments and showed the repeats to be in tandem. No homologs of the VERL repeat were found by database searches. A high-density arrayed abalone ovary cDNA library was screened. A 2.6-kb clone was sequenced; it contained 5.5 tandem repeats (*9*). The repeats are >95% identical, which makes it difficult to order repeat units from different cDNA clones. We cloned PCR products to determine the sequences of two to seven VERL repeats from six additional abalone species (*10*) (Fig. 1). PCR products from genomic DNA and cDNA had amplified bands of the same sizes and yielded only VERL repeat sequences, indicating the absence of introns. Because the repeats are in tandem, they are shown from an arbitrary beginning (Fig. 1). The repeats in Fig. 1 represent a subset of repeats that would be found in one VERL molecule from each species; they do not represent polymorphisms within a species and are not known to be orthologous. Some replacements in the repeats are species specific (Fig. 1). Many of these positions are replacements between hydrophobic and hydrophilic or hydrophobic and Ser or Thr residues (which may be O-glycosylated; VERL is 50% carbohydrate) (*6*). The presence of an 11-residue internal repeat in some sequences is the major difference among repeat units. The calculated isoelectric point of the repeat is 4.7, which is the pH at which VEs dissolve; this is consistent with the idea that VERL maintains the structural integrity of the VE.

Southern blots show that VERL is a single-copy gene and that all repeats are contiguous without introns (*11*) (Fig. 2A). Digestion of pink abalone DNA with Eco RI or Hind III, neither of which cuts within the pink repeat, produces a single band of hybridization at about 20 kb (lanes 2 and 3). However, digestion with Pst I, which cuts twice within the pink repeat, yields one hybridizing band at about 400 bp (lane 4), the expected size of the restriction fragment. Two weaker bands at 0.75 and 9 kb are most likely the 5′ and 3′ ends of the gene with one repeat attached (lanes 4 and 7). Similar results were obtained with red abalone DNA (Fig. 2A, lanes 5 to 8). Northern blot analysis showed one band of hybridization at about 13.5 kb, which could code for a protein of 470 kD (*12*) (Fig. 2B). If this is the size of VERL, it could contain ~28 repeats, which suggests that two lysins bind each repeat (*12*).

To determine the selective pressure affecting the evolution of VERL, we computed the number of nonsynonymous (Dn; amino acid altering) nucleotide substitutions per nonsynonymous site, and the number of synonymous (Ds; silent) substitutions per synonymous site (*13*). If there were no selection, Dn would equal Ds. If positive Darwinian selection occurred, Dn would be statistically greater than Ds. If purifying selection occurred, Ds would be greater than Dn. A graph of Dn versus Ds (Fig. 3) shows that most comparisons cluster just below the diagonal line depicting Dn = Ds (the neutral expectation). Of 561 pairwise comparisons, 499 show no significant difference between Ds and Dn. Because most points cluster below Dn = Ds, the analysis suggests that VERL is subjected to weak, purifying selection. However, when Dn approximates Ds it is difficult to determine the relative importance of positive selection versus purifying selection.

A phylogenetic tree of VERL repeats (*14*) (Fig. 4) shows that repeats are more similar within a species than between any two species. This indicates that, like other repeated DNA motifs, VERL repeats evolve by concerted evolution. This is a process by which unequal crossing over and gene conversion homogenize repeats within a species (*15*). The presence or absence of an 11-codon imperfect, internal repeat in some sequences (Fig. 1) is evidence of unequal crossing over. Southern blot analysis (Fig. 2A) is consistent with concerted evolution acting on VERL. For example, the red abalone acquired one Hind III site, which spread through all red VERL repeats. However, this restriction site

Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093–0202, USA.

*To whom correspondence should be addressed. E-mail: jwswanson@ucsd.edu

is not found in pink abalone repeats (Fig. 2A, lanes 3 and 7). Similarly, pink abalone acquired two Pst I sites, whereas the red abalone has only one Pst I site (Fig. 2A, lanes 4 and 8). Further support for concerted evolution comes from the plot of Dn versus Ds (Fig. 3). The within-species comparisons cluster toward the bottom left of the graph, showing that within-species comparisons are more closely related than between-species comparisons.

Concerted evolution has been documented in repeated coding and noncoding sequences (*15*). Genes evolving by concerted evolution have been hypothesized to produce a selective force on the gene of their cognate interacting protein (*16*). Concerted evolution may underlie species-specific lysin-VERL interaction in a similar manner, as is thought to underlie the evolution of species-specific transcription of tandemly repeated ribosomal DNA (rDNA) by RNA polymerase I (*17, 18*). Concerted evolution randomly samples and replaces species-specific differences in the promoters to which RNA polymerase I transcription initiation factors bind [TBP (TATA box binding protein)-associated factors (TAFs)] (*18*). This, in turn, causes the evolution of compensatory changes in the TAFs (*17, 18*). The TAFs are 20 to 34% divergent between mice and humans (*18*), a level of divergence comparable to the rapidly evolving defense proteins (average divergence of 75 defense proteins between mouse and human is 35%) (*19*). The rDNA gene family is large, and concerted evolution homogenizes changes in a population in a cohesive manner (the population mean ratio of mutant to wild-type repeats has low variance as the mutant propagates by concerted evolution); thus there would be time for selection to act while maintaining molecular interaction (*20*).

A similar process could underlie the evolution of species-specific lysin-VERL interaction. With ~28 repeats per VERL molecule, the selective pressure on individual repeats could be relaxed (as shown by a Dn/Ds ratios approaching 1) (Fig. 3). If a suboptimal change for lysin binding occurred in one VERL repeat, lysins could bind the other ~27 unchanged repeats. The change in one

```
White VPITQELGINMMLIQYTRSELPGSPGMCVFWGPYSAPKNDTVVLYTVTARLKWSEGPPTNLSIECYMPKSPVAPKPETGPTSS
White ..................N..LD............V...............................................
White ...................................................................................
White ......F..............LD............V...................................D...........
White ..................N..LD............V...................................D...........
Flat  ......F...........N..LD............V...................................D...........
Flat  ......F...........N..LD............V...................................D...........
Flat  ..................N..LD............V.................R.................D...........
Flat  ......F...........N..LD............V.R.................................D...S.......
Pinto ......F..............LDT...........V.R..S.....................................S...N
Pinto ......F..............LD........G...V...............................................
Red   ......F...........N..LD............V............................................S.N
Red   ......F...........N..LD............V..............................................N
Red   ......F...........N..LD............V..............................................N
Red   ......F...........N..LD............V............................................S.N
Red   ......F...........N..LD............V..............................................N
Red   ......F...........N..LD............V............................................S.N
Green ......Y.F.........N..LD............V.......................D...........D.S......S..
Green ......Y.F.........N..LD.S..........V.........................T.........D.S......S..
Green ......Y.F.........N..LD............V...................................D.S......S..
Green ......Y.F.........N..LD............V...................................D.S......S..
Green ......N.F.........N..LD............V...................................D.S......S..
Black ......SS..........N..LD............V............................................S..
Black ......SS..........N..LD........G...V..........................................A.S..
Black ......S...........ND.LD........G...V............................................S..
Black ......S...........ND.LDG...........V............W.............................A.S..
Black ......S...........N..LD............V............................................S..
Pink  ......S.........SLNDSLD............V............................................I..
Pink  ......S.........SLNDSLD............V........................................K......
Pink  ......S.........SLNDSLD............V...............................................
Pink  ......S.........SLNDSLD............V...................................I...........
Pink  ......S.........SLNDSLD........G...V.................R.............................
Pink  ......S.........SLNDSLD........G...V.........................P..........S.E..------
Pink  ......S.........SLNDSLD............V.................R.............................

White PPEPEASPTSNAPEPETYPTSSAPEKVSSDQPAPSHNQSKLIDWDVYCSQNESIPAKFISRIVT-SKDQALEKTELNCSNGL
White ................................................................-..........I......
White ................................................................-.................
White ................................................................-.................
White ...........................Y......P.............................-..........I......
Flat  .............................................................L..-..........I......
Flat  .............................................................L..-..........I......
Flat  ..........................................N.....................-..........I......
Flat  ........P....................................................L..-..........I......
Pinto A...------------..........................T..................L..-..........I......
Pinto T..............Q..........................T..................L..-..........I......
Red   A...------------.............................................LL.-..........I......
Red   A...Q-----------............................................-L..-..........I......
Red   A...Q-----------.....................................F.......L..-..........I......
Red   A...------------...........Y..............N..................L..-..........I......
Red   A...------------.............................................L..-..........I......
Red   A...------------...........Y..............N..................L..-..........I......
Green TA..G-----------...................D.........................L..-.....V...DI......
Green TA..G-----------...................D.........................L..-.....V...DI......
Green TA..G-----------...................D.........................L..-.....V...DI......
Green TA..G-----------.............................................L..-.....V...DI......
Green TA..G-----------.............................................L..-.....V...DI......
Black T...............S..........F.....L...........................L..-.....V....M......
Black T...........................................H................L..S.....V....I......
Black T........Y......S..........F.....L...........................L..-.....V....M.....P
Black T.....................................HN..N..................L..-.....V....I......
Black T...............S..........F.....L...........................L..-.....V....M..KH..
Pink  T...------------AS...K..G....N.....P........E......D..........P.-.....V....V......
Pink  T..D------------AS...K..G......T...................D..........PI-.....V....V......
Pink  T...------------AS...K..G..........................D..........PI-.....V....V......
Pink  T...------------AS...K..G..........................D..........PI-.....V....V......
Pink  T...------------ASS..K..G....N.....P........E......D..........P.-.....V....V......
Pink  -----..S.............K.............P........E......D............-.....V....V......
Pink  T...------------ASS..K..G....N.....P........E......D..........P.-.....V....V......
```

Fig. 1. Aligned, deduced amino acid sequences of VERL repeats (*23*). Dots denote identity to the first sequence, dashes are for alignment. Regions corresponding to peptide sequences are underlined. Species and scientific names are as follows: white, *Haliotis sorenseni*; flat, *Haliotis walallensis*; pinto, *Haliotis kamtschatkana*; red, *Haliotis rufescens*; green, *Haliotis fulgens*; black, *Haliotis cracherodii*; pink, *Haliotis corrugata*. GenBank accession numbers are AF053652 to AF053681.

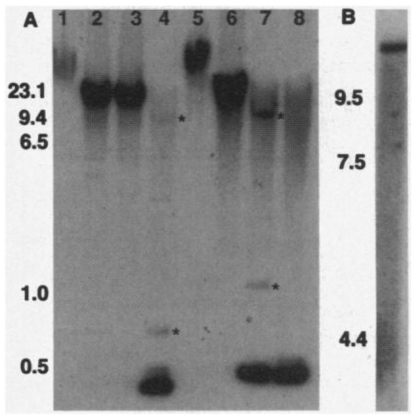

Fig. 2. (**A**) Southern blot analysis. Lanes: 1, undigested pink DNA; 2, Eco RI–digested pink DNA; 3, Hind III–digested pink DNA; 4, Pst I–digested pink DNA; 5, undigested red DNA; 6, Eco RI–digested red DNA; 7, Hind III–digested red DNA; 8, Pst I–digested red DNA. Digestions that cut within the repeat (lanes 4, 7, and 8) produce hybridizing bands of the expected repeat sizes. Asterisks mark locations of weakly hybridizing bands, which may represent the 3′ and 5′ ends. Sizes are indicated in kilobases. (**B**) Northern blot analysis.

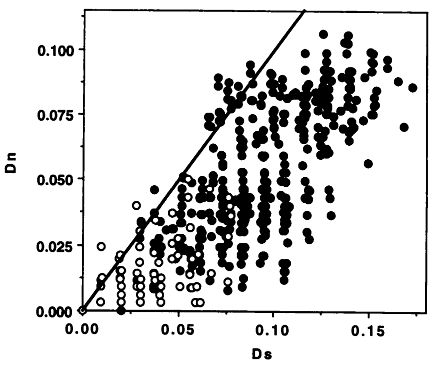

Fig. 3. Plot of nonsynonymous substitutions per nonsynonymous site (Dn) against synonymous substitutions per synonymous site (Ds). Line indicates the neutral expectation where Dn = Ds. Open circles are within-species comparisons, and closed circles are between-species comparisons.

repeat could be randomly sampled by concerted evolution and replacement of all repeats could occur in a cohesive manner. As this variant becomes the major type of repeat in the population, selective pressure for change in lysin could occur to maintain VERL-lysin binding. Such changes, occurring in different lineages, could produce and maintain the species-specific binding of VERL and lysin as species diverge. The process could be continuous with sperm lysin constantly adapting (*1*) to VERL repeats, which are changing by concerted evolution. The process requires no external forces and could occur within populations.

By molecular evolutionary analysis of the proteins involved in sperm-egg interaction, we have uncovered a possible hypothesis by which species-specific fertilization may evolve in abalones. Whether concerted evolution affects other gamete recognition proteins is unknown. However, as is true in the abalone, many extracellular proteins (including those involved in gamete recognition) exhibit a modular design composed of repeating units that could evolve by concerted evolution (*21*). Such a process for the evolution of gamete recognition systems could be involved in establishment of prezygotic reproductive isolation in the speciation process (*22*).

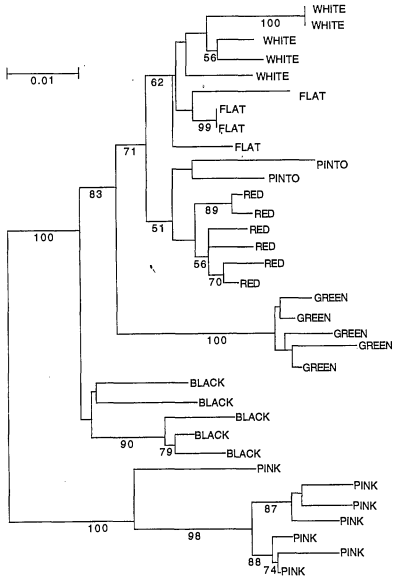

Fig. 4. Neighbor-joining phylogeny of the VERL repeats. Scale bar, Jukes-Cantor corrected nucleotide distance. Bootstrap values > 50% (1000 replicates) are shown below the branches. A tree of amino acid replacements yields the same topology. The topology of this tree is in general agreement with that of published lysin trees (*1*).

References and Notes

1. Y.-H. Lee and V. D. Vacquier, *Biol. Bull.* **182**, 97 (1992); Y.-H. Lee, T. Ota, V. D. Vacquier, *Mol. Biol. Evol.* **12**, 231 (1995); V. D. Vacquier and Y.-H. Lee, *Zygote* **1**, 181 (1993); A. Shaw, D. E. McRee, V. D. Vacquier, C. D. Stout, *Science* **262**, 1864 (1993).
2. E. C. Metz and S. R. Palumbi, *Mol. Biol. Evol.* **13**, 397 (1996); S. C. Tsaur and C.-I. Wu, *ibid.* **14**, 544 (1997); P. J. Ferris, C. Pavlovic, S. Fabry, U. W. Goodenough, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 8634 (1997); V. D. Vacquier, W. J. Swanson, Y.-H. Lee, *J. Mol. Evol.* **44**, S15 (1997); W. R. Rice, *Nature* **381**, 232 (1996).
3. R. Yanagimachi, in *The Physiology of Reproduction*, E. Knobil and J. Neil, Eds. (Raven Press, New York, 1988), p. 135; E. R. S. Roldan and R. Yanagimachi, *J. Exp. Zool.* **250**, 321 (1989).
4. R. G. Summers and B. L. Hylander, *Exp. Cell. Res.* **96**, 63 (1975); G. Giudice, *Developmental Biology of the Sea Urchin Embryo* (Academic Press, San Diego, 1973), p. 162; K. Osanai and K. Kyozuka, *Gamete Res.* **5**, 49 (1982); E. C. Metz, R. E. Kane, H. Yanagimachi, S. R. Palumbi, *Biol. Bull.* **187**, 23 (1994).
5. B. Owen, J. H. McLean, R. J. Meyer, *Bull. Los Angel. Cty. Mus. Nat. Hist. Sci.* **9**, 1 (1971).
6. W. J. Swanson and V. D. Vacquier, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6724 (1997).
7. ^{125}I-labeled VERL was excised from an SDS gel and cleaved with cyanogen bromide [V. Nikodem and J. R. Fresco, *Anal. Biochem.* **97**, 382 (1979)]. Endo-LysC cleavage of VERL excised from SDS gels was performed by the PAN Facility (Stanford University).
8. PCR primers were TAYTGDATNARCATCATRTT and GTNCCNATHACNCARGA (Y = C or T; D = A, G, or T; N = A, G, C, or T; R = A or G; H = A, C, or T), template was ovary cDNA primed with oligo(dT). PCR products were cloned and sequenced by using ABI prism FS chemistry on an Applied Biosystems model ABI 370.
9. A pink abalone ovary, random-primed cDNA library was made in pSport (BRL), and 92,160 independent clones were arrayed (Q-Bot) on five high-density filters. Filters were hybridized with a random-primed probe of the VERL repeat. Nested deletions (Promega) were generated and sequenced.
10. PCR was performed with genomic DNA or with cDNA from the other abalone species. PCR primers are as in (*8*). The products were cloned and individual colonies were sequenced. To avoid biased amplification, two or three PCRs were done for each species.
11. Ten micrograms of genomic DNA were digested, separated on a 0.8% agarose gel, denatured, neutralized, transferred to nylon, and probed as in (*9*).
12. One hundred micrograms of total RNA were separated on a 0.8% agarose formaldehyde gel, transferred to nylon, and probed as in (*9*). A pink VERL repeat is 459 bp. Assuming 500 bp is untranslated, 13 kb codes for 28.3 repeats. The average molecular mass of the pink repeats is 16,560 daltons, which gives a total molecular mass of ≈470 kD. Recalculating the binding stoichiometry (*6*) with this VERL molecular mass yields 58 to 66 lysins per VERL.
13. Maximum-likelihood estimates of Dn and Ds were calculated [S. V. Muse, *Mol. Biol. Evol.* **13**, 105 (1996)]. Identical values were obtained by the Nei and Gojobori method [M. Nei and T. Gojobori, *Mol. Biol. Evol.* **3**, 418 (1986)].
14. A neighbor-joining tree was constructed with MEGA (S. Kumar, K. Tamura, M. Nei, Pennsylvania State University) using Jukes-Cantor corrected nucleotide distances. Positions with alignment gaps were omitted only in pairwise comparisons.
15. G. P. Smith, *Cold Spring Harbor Symp. Quant. Biol.* **38**, 507 (1973); F. J. Elder and B. J. Turner, *Q. Rev. Biol.* **70**, 297 (1995).
16. G. A. Dover, *Curr. Opin. Genet. Dev.* **3**, 902. (1993); _____ and R. B. Flavell, *Cell* **38**, 622 (1984).
17. I. Grummt, E. Roth, M. R. Paule, *Nature* **296**, 173 (1982); R. Miesfeld and N. Arnheim, *Mol. Cell. Biol.* **4**, 221 (1984); U. Rudloff, D. Eberhard, L. Tora, H. Stunnenberg, I. Grummt, *EMBO J.* **13**, 2611 (1994).
18. J. Heix, J. C. B. M. Zomerdijk, A. Ravanpay, I. Grummt, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1733 (1997).
19. P. M. Murphy, *Cell* **72**, 823 (1993).
20. T. Ohta and G. A. Dover, *Genetics* **108**, 501 (1984).
21. M. Baron, D. G. Norman, I. D. Campbell, *Trends Biochem. Sci.* **16**, 13 (1991); R. F. Doolittle and P. Bork, *Sci. Am.* **269**, 50 (October 1993).
22. S. R. Palumbi, *Annu. Rev. Ecol. Syst.* **25**, 547 (1994).
23. Abbreviations for amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

24. We thank E. H. Davidson, Y.-H. Lee, T. Walkup, E. C. Metz, M. E. Hellberg, J. R. Schulz, R. S. Burton, R. R. Hudson, J. D. Calkins, J. M. Swanson, C. F. Aquadro, M. Schug, K. Schmid, J. B. Walsh, and R. R. McConnaughey for their research support and editorial help. Supported by NIH grant HD12986 to V.D.V.

16 March 1998; accepted 11 June 1998