J Mol Evol (2002) 54:403–410
DOI: 10.1007/s00239-001-0030-0

JOURNAL OF MOLECULAR EVOLUTION

Positive Darwinian Selection Promotes Heterogeneity Among Members of the Antifreeze Protein Multigene Family

Willie J. Swanson,[1,2] Charles F. Aquadro[2]

[1] Department of Biology, University of California, Riverside, Riverside, CA 92521, USA
[2] Department of Molecular Biology and Genetics, 402 Biotechnology Building, Cornell University, Ithaca, NY 14853-2703, USA

Received: 27 February 2001 / Accepted: 12 September 2001

Abstract. A variety of organisms have independently evolved proteins exhibiting antifreeze activity that allows survival at subfreezing temperatures. The antifreeze proteins (AFPs) bind ice nuclei and depress the freezing point by a noncolligative absorption–inhibition mechanism. Many organisms have a heterogeneous suite of AFPs with variation in primary sequence between paralogous loci. Here, we demonstrate that the diversification of the AFP paralogues is promoted by positive Darwinian selection in two independently evolved AFPs from fish and beetle. First, we demonstrate an elevated rate of nonsynonymous substitutions compared to synonymous substitutions in the mature protein coding region. Second, we perform phylogeny-based tests of selection to demonstrate a subset of codons is subjected to positive selection. When mapped onto the three-dimensional structure of the fish antifreeze type III antifreeze structure, these codons correspond to amino acid positions that surround but do not interrupt the putative ice-binding surface. The selective agent may be related to efficient binding to diverse ice surfaces or some other aspect of AFP function.

Key words: Antifreeze proteins — Darwinian selection — Likelihood ratio test — Adaptive evolution

Correspondence to: Willie J. Swanson; *email:* willies@citrus.ucr.edu

Introduction

Antifreeze proteins (AFPs) inhibit ice crystal growth by binding ice nuclei which depresses the freezing point of water, thereby allowing survival at subzero temperatures (Cheng 1998). Most organisms with AFPs do not have a single component with AFP activity but, rather, a heterogeneous suite of molecules with AFP activity. In fact, some marine fish possess a suite of AFPs with at least 10 identifiable active components (Hew et al. 1984). A few types of AFPs have undergone multiple duplications, with some species having more than 100 copies (Hew et al. 1988). Although it has been suggested that the heterogenous suite of AFPs is functionally important (Hew et al. 1988), currently there is no evidence that the changes are adaptive. Therefore, we performed a rigorous statistical analysis of the AFP genes to determine if the changes observed are neutral or whether there is evidence for positive selection promoting the diversification of the AFP components within an organism.

A classic example of positive selection is the evolution of the major histocompatibility loci (MHC) (Hughes and Nei 1988). In almost all homologous proteins, the frequency of nonsynonymous substitutions per nonsynonymous site (d_N; amino acid replacement) is significantly lower than that of synonymous substitutions per synonymous site (d_S; silent), reflecting selective constraint on amino acid sequence (Li 1997). However, in the antigen recognition site (ARS) of the Class I MHC glycoproteins, d_N is greater than d_S, whereas non-ARS regions show the opposite. This difference has been interpreted as evidence for an adaptive value in diversifying the ARS to allow for the binding of an enormous

variety of processed peptides derived from microbial pathogens (Hughes et al. 1990).

AFPs have multiple independent origins, with approximately eight classes identified (Cheng 1988). Many of the types contain a repetitive structure rich in alanine. The length of the repeats varies within a molecule as well as among homologues making sequence alignments difficult. Here, we focus our analysis on two distinct, independently evolved AFPs for which reliable alignments are possible: fish type III AFPs from *Macrozoarces americanus*, *Anarhichas lupus*, and *Lycodichthys dearborni* (Hew et al. 1984, 1988; Li et al. 1985; Wang et al., 1995) and insect AFPs from the beetle *Tenebrio molitor* (Graham et al. 1997; Liou et al. 1999). The average nucleotide divergence of these proteins is moderate (fish type III, 14.7%; beetle, 7.2%). Therefore, reliable evolutionary inferences can be made without the problem of mutational saturation effects. The AFPs analyzed here from each group represent paralogous comparisons, as each gene is a member of a multigene family. In addition to paralogues within a species, we use sequences from different species. Given the enormous size of this multigene family (100s of copies), it is likely that few if any represent orthologous comparisons. The inclusion of orthologues would not effect our detection of selection. Therefore, we are testing if the heterogeneity among AFPs found within an organism is adaptive or neutral. We are not testing for adaptive changes between species or antifreeze types. We demonstrate that positive Darwinian selection accounts for at least some of the heterogeneity of AFP protein sequence.

Methods

Seven fish-type III AFP genes [sequence accession numbers (some contain multiple coding regions): M22125, M11790, J03923, J03924, U20439, and two from Hew et al. (1988) and nine beetle AFP genes (sequence accession numbers AF160494–AF160497 and AF159114–AF159118)] were obtained from GenBank. Other sequences in GenBank contained only a partial coding sequence or were identical to the ones used and were not included in the analysis. The translated protein products were used to create a multiple sequence alignment using ClustalW (http://www2.ebi.ac.uk/clustalw/), and the nucleotide sequences aligned according to the protein alignment using Protal2DNA (http://bioweb.pasteur.fr/seqanal/interfaces/protal2dna-simple.html). Other suboptimal alignments were considered, but they did not change the results. d_N and d_S were calculated using the method of Goldman and Yang (1994). Estimates of d_N and d_S by the method of Nei and Gojobori (1986) produced similar results. The d_N/d_S ratio is abbreviated ω herein. Likelihood models that allow for variation in the selective pressure among lineages or sites were used in phylogenetic analyses, as implemented in the CODEML program of PAML (Yang 1999). For lineage variation, a model with one ω ratio for all sites was compared to a selection model in which a separate ω ratio was calculated for each lineage (Yang and Nielsen 1998; Yang 1998). The negative of twice the log-likelihood difference of the two models was compared to the chi-square distribution, with degrees of freedom $n-1$ (n = number of branches on phylogeny). For site variation, a variety of models of codon evolution were compared (Nielsen and Yang 1998; Yang et al. 2000). First, a neutral model (M1) with two classes of codons (ω = 0 for conserved sites and ω = 1 for neutral sites) was compared to a selection model (M2) with an additional class of codon with a ω ratio estimated from the data. Second, the neutral model M1 was compared to a selection model (M3) with three classes of codons with ω ratios estimated according to a discrete distribution. Finally, we compared a neutral model (M7) with ω ratios estimated according to a beta distribution, β(p,q), with ω limited to the interval (0,1) to a selection model with an additional class of codons with a ω ratio estimated from the data (where ω can be greater than 1). As these models assume a constant d_S across the sequence, and vary the ω ratio, signals of selection can not result from a locally reduced d_S. Our results are therefore dependent upon the assumption of a constant d_S, which also makes the test conservative. Because of the biased amino acid composition of the AFPs, the data were analyzed using the full (F61) model estimating the codon frequencies. The negative of twice the log-likelihood difference between the nested models was compared to the chi-square distribution with degrees of freedom equal to the difference in number of parameters estimated in the models. M1 contains one parameter (one proportion); M2, three parameters (two proportions, one ω ratio); M3, five parameters (two proportions and three ω ratios); M7, two parameters (p and q for the β distribution); and M8, four parameters (β distribution, one proportion, and one ω ratio). Residues identified to be potentially subjected to positive selection were mapped onto the crystal structure of the fish type III AFP structure (PDB file 1MSI) (Zia et al. 1996) using Molscript (Kraulis 1991).

Since the AFPs are short genes, and the likelihood-ratio test statistic is a large sample test (Nei and Kumar 2000), we confirmed the significance by parametric bootstrapping. We used the EVOLVER program from the PAML package (Yang 1999) to simulate 100 data sets under the null model (M1) for both the fish and the beetle antifreeze genes. Under the null model, the simulated data had similar codon frequencies, transition/traversion ratios, divergences, and ω ratios as the real data. The 100 simulated data sets were analyzed with model M1 and M2 and resulting likelihood ratios compared to our experimental value.

To compare radical versus conservative amino acid replacements, we used the method of Hughes et al. (1990). We calculated the number of radical amino acid replacement changes per radical nonsynonymous site and the number of conservative amino acid replacement changes per conservative nonsynonymous site. We used three classifications of amino acids: charge, polarity, and type (which takes into account polarity, charge, hydrophobicity, and size).

Results

Comparison of d_N *and* d_S *Averaged Across All Sites and Lineages*

AFPs have two distinct parts, which we analyzed separately. The first part is the signal sequence, which is cleaved off from the mature AFP and is therefore not expected to be a target of selection associated with the function of the active, mature AFP. The second part is the mature AFP, which is extracellular and interacts with ice and, therefore, may be a target of selection. Protein alignments of the genes used in subsequent analyses are presented in Fig. 1. It is apparent that that the mature protein-coding region is more divergent than the signal sequence.

To test for deviation in the substitution pattern from the neutral expectation, we calculated d_N and d_S for the beetle and fish antifreeze proteins. We performed sepa-

Fish type III AFPs

```
MKSAILTGLLFVLLCVDHMSSA S-QSVVATQLIPINTALTPIMMKGQVVNPAGIPFAE-MSQIVGKQVNRAVAKDETLMPNMVKTYR--AAK-
MKSAILTGLLFVLLCVDHLSSA S-QSVVATQLIPINTALTPIMMKGQVVNPAGIPFAE-MSQIVGKQVNRPVAKDETLMPNMVKTYR--AAK-
MKSVILTGLLFVLLCVDHMT-A S-QSVVATQLIPINTALTPVMMEGKVTNPIGIPFAE-MSQIVGKQVNTPVAKGQTIMPNMVKTYA--AGK-
MKSVILTGLLFVLLCVDHMT-A S-QSVVATQLIPMNSALTPVMMEGKVTNPIGIPFAE-MSQMVGKQVNRPVAKGQTIMPNMVKTYA--AGK-
MKSVILTGLLFVLLCVDHMSSA S-QSVVATQLIPINTALTPAMMVGKVTNPIGIPFAE-MSQIVGKQVNTPVAKGQTLMPNMVKTYV--AGK-
MKSVILTGLLFVLLCVDHMT-A N-QSVVATQLIPINTALTLVMMTTRVIYPTGIP-AEDIPRLVSMQVNQAVPMGTTLMPDMVKFYCLCAPKN
MKSVVLTGLLFVLLCVDHMSSA NKASVVANQLIPINTALTLIMMKAEVVTPMGIP-AEDIPRIIGMQVNRAVPLGTTLMPDMVKNY----EK-
1MSI (PDB Structure)       SVVANQLIPINTALTLVMMRSEVVTPVGIP-AEDIPRLVSMQVNRAVPLGTTLMPDMVKGY---AA--
```

Beetle AFPs

```
MAFKTCGFSKKWLVIAVIVMCLCTECYC QCTGGADCTSCTGACTGCGNCPN------------AVTCTNSQHCVKANTCTG
MGFKTCGFSKKWLVTAVIVMCLCTECYC QCTGGADCTSCTAACTGCGNCPN------------AVTCTNSQHCVKATTCTG
MAFKTCGFSKKWLVIAVIVMCLCTECYC RCTGGADCTSCTQACTSCRNCPN------------AKTCTNSQHCVRARTCTG
MAFLTCGFSKKWLVIAVIVMCLCTECYC QCTGGSDCTSCTAACTGCGNCPN------------AHTCTDSQHCVKAATCTG
MAFKTCGFSKKWLVIAVIVMCLCTECYC HCTGGADCTSCTDACTGCGNCPN------------AHTCTDSKNCVKAATCTG
MAFKACGFSKKWLVIAVIVMCLCTECYC HCTGGADCTSCTDACTGCGNCPN------------AHTCTDSKNCVKAATCTG
MAFKTCGFSKKWLIIAVIVMCLCTECYC QCTGGADCTSCTAACTGCGSCPNAHTCIDSKNCVRAETCTDSENCVKAHTCTG
MAFKTCGFSKKWLIIAVIVMCLCTECYC QCTGGADCTSCTAACTGCGSCPNAHTCTDSKNCVRAETCTDSENCVKAHTCTG
MSFKISTFTKIWLIIAVIVMCLCNEYNC QCTGAADCTSCTAACTGCGNCPN------------AITCTGSKNCVRATTCTG

STDCNTAQTCTNSK----------DCFEANTCTD----------STNCYKATACTNSSGCPGH
STDCNTAVTCTNSK----------DCFEAQTCTD----------STNCYKATACTNSTGCPGH
STDCNRAMTCTNSK----------DCFEAKTCTD----------STNCYKATTCTNSTGCPGH
STDCNTARTCTNSK----------DCFEAATCTD----------STNCYKATACTHSTGCPGH
STKCNTARTCTNSK----------DCFEAKTCTD----------STNCYKATACTNSTGCPGH
STKCNTARTCTNSK----------DCFEAKTCTD----------STNCYKATACTNSTGCPGH
SRNCNTAMTCTNSK----------DCFEAKTCTD----------STNCYKATACTNSTGCPGH
SRNCNTAMTCTNSK----------DCFEAKTCTD----------STNCYKATACTNSTGCPGH
STNCNRATTCTNSKGCLEATTCTGSTHCHRATTCTNSKDCFEATTCTGSSNCYTATTCTNSTNCYKATACTNSTGCPGH
```

Fig. 1. Alignments of the fish type III and beetle antifreeze proteins used in statistical tests of neutrality. The *italicized* sequence represents the signal sequence. Sites predicted to be under positive selection with a probability >0.9 are in *underlined boldface;* 0.8–0.9, *underlined;* 0.7–0.8, in *underlined italics;* and 0.5–0.7, in *italics* (mature protein only). The sequence of the structure 1MSI of fish AFP type III used in Fig. 3 is presented. Sequences for fish: M22125, two from Hew et al. (1988), M11790, J03923, U20439. Sequences for beetle: AF160494, AF159114, AF159117, AF159115, AF160495, AF159116, AF159118, AF160496, AF160497.

rate calculations for the signal sequence and mature protein-coding regions. d_N exceeds d_S in 18 of 21 possible pairwise comparisons among the nucleotide sequences of the mature AFP from seven paralogous fish type III AFP genes (Fig. 2a). In contrast, the nucleotide sequence encoding the signal sequence shows the pattern expected in the absence of diversifying selection: d_N is less than d_S in 20 of the 21 comparisons (Fig. 2a). We found similar results for the beetle AFPs as for the fish type III antifreeze: d_N exceeds d_S in 27 of 36 pairwise comparisons for the mature protein, whereas in the signal sequence d_N is less than d_S in 24 of the 36 comparisons (Fig. 2b). However, since these calculations involve pairwise comparisons there are potential problems with the nonindependence of the data. Below, we address this issue by performing phylogeny-based tests of positive selection. Nevertheless, these pairwise comparisons are interpreted as evidence that there is adaptive value in altering the mature AFP mature amino acid sequence.

Variation in the d_N/d_S *Ratio Between Lineages*

We next analyzed for positive selection using a variety of phylogeny-based maximum-likelihood methods. These include analyzing for variation in the d_N/d_S ratio (abbreviated ω herein) between lineages and sites. First, we tested for positive selection by analyzing for variation in the ω ratio between lineages (Yang and Nielsen 1998; Yang 1998). The strictly neutral theory predicts the ω ratio to be constant among all branches of a phylogeny. In some examples of positive selection, it has been demonstrated that a burst of positive selection occurs along one lineage, followed by purifying selection (Messier and Stewart 1997; Yang 1998). To test for variation between lineages we compared the likelihood of a model with one ω ratio for all lineages to a model where a separate ω ratio is estimated for each lineage. For both the fish type III and the beetle AFPs, we detected no difference between the two models (Table 1). This is interpreted as similar selective pressures acting on all AFP types present within an organism.

Variation in the d_N/d_S *Ratio Among Sites*

Next we tested for positive Darwinian selection using maximum-likelihood methods to analyze the sequences with variation of the ω ratio among sites (Nielsen and Yang 1998; Yang et al. 2000). The ω ratio averaged across all sites and lineages is greater than 1 for both the beetle and the fish type III antifreeze proteins but not significantly different from 1 (Table 2). However, averaging across all sites is not a sensitive measure for se-

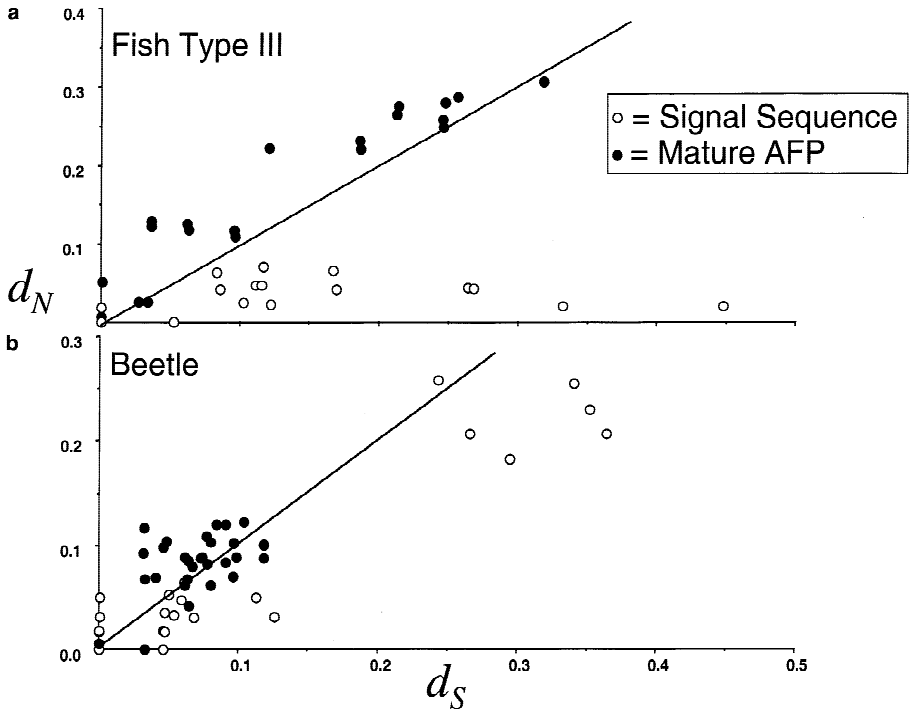

Fig. 2. The number of nonsynonymous substitutions per nonsynonymous sites (d_N) plotted against the number of synonymous substitutions per synonymous sites (d_S). The line is where $d_N = d_S$. (●) Mature protein-coding region; (○) signal sequence. **a.** Comparison for the fish type III AFPs. **b.** Comparison for the beetle AFPs.

Table 1. Likelihood-ratio tests for variation of the d_N/d_S ratio between lineages[a]

Gene	ln*l* 1 ratio	ln*l* free ratio	$-2\Delta\, l$	df	*p*
Fish type III	−587.00	−583.11	7.8	11	0.73
Beetle	−683.3	−673.4	19.8	14	0.14

[a] $-2\Delta\, l$ = the negative of twice the log-likelihood difference between the two models; df = degrees of freedom.

lection, since it is likely that only a few sites may be subjected to positive selection while others are conserved. To analyze for the variation in ω among sites, we utilized maximum-likelihood ratio tests. Other statistical tests for variation in the ω ratio among sites are also possible (e.g., Suzuki and Gojobori 1999) but require more data than are currently available for the AFP genes. The maximum-likelihood method involves comparing the likelihood of two models, a neutral model to a selection model. We used a variety of models of codon evolution ranging from two ω ratios to a β distribution of ω ratios (see Methods). The neutral models all have ω ratios limited between 0 and 1, while the selection models allow for a class of codons with ω estimated from the data and possibly greater than 1. In all cases, positive selection is indicated if the selection model fits the data significantly better and has a class of codons with a ω ratio > 1. For the mature fish type III and the beetle AFPs, the selection models are significantly more likely than their corresponding neutral models with an ω class > 1 ($p < 0.005$), indicating positive selection (Table 2). For the fish type III AFPs, the selection model M8 predicts 27% of the codons to be under positive selection with an average ω ratio of 7.6. Similar results are obtained with the mature beetle AFPs, the selection model M8 predicts 6.7% of the codons to be under positive selection with an average ω ratio of 32.1.

Parametric Bootstrapping

To confirm the significance of the likelihood-ratio test based on comparison with the χ^2 distribution, we performed parametric bootstrapping. One hundred data sets were simulated using the estimated parameter values (tree topology, tree length, number of codons, number of taxon, transition/transversion ratio, codon frequencies, proportion with $\omega = 0$ or 1) obtained for the null model from both the beetle and the fish AFPs. The resulting simulated data sets were analyzed using models M1 and M2. In the simulated data sets, we found 5 and 7 significant comparisons of 100 ($p < 0.05$) comparing the likelihood-ratio test statistic to the χ^2 distribution for the beetle and fish, respectively. This is close to the expectation of five significant comparisons. This suggests that in this case just 100 codons may be sufficient when applying the likelihood-ratio test based on the χ^2 distribution. The observed likelihood-test statistic from the real data were 3.5-fold (for fish) or 7-fold (for beetle) higher than the simulated data (Fig. 3). This confirms the significance of our results.

Radical Amino Acid Replacements

To investigate the type of amino acid changes which occurred, we calculated the number of radical amino acid

Table 2. Likelihood-ratio test of variation in the d_N/d_S ratio (ω) between sites[a]

Gene	*n*	Lc	S	ω_{total}	$-2\Delta l$			Parameter estimates under M8 (β & ω)	Positively selected sites
					M2 vs M1	M3 vs M1	M8 vs M7		
Fish type III AFPs	7	62	1.40	1.3	21.4	24.6	21.4	**p_1 = 0.277, ω = 7.6** p_0 = 0.723 β(0.0256, 0.0202)	2,*14*, **19**, 22, 23, **24**, 26, *27*, **29**, **38**, **40**, **45**, *46*, *49*, *50*, **51**, **60**
Beetle AFPs	9	81	0.89	1.3	70.4	71.2	70.3	**p_1 = 0.067, ω = 32.1** p_0 = 0.933 *B*(0.005, 0.005)	**13**, **24**, **36**, **48**, **59**

[a] *n* = number of sequences; Lc = number of codons after gaps removed; S = tree length as substitutions per codon; $-2\Delta l$ = the negative of twice the log-likelihood difference between the two models; $\omega_{total} = d_N/d_S$ ratio averaged across all sites and lineages; $\omega = d_N/d_S$ ratio; M1 = model 1, etc.; p_1 = proportion of codons under positive selection; p_0 = proportion of codons not under positive selection; $\beta(p,q)$ = parameters for the β distribution. Parameters indicating positive selection are in boldface. In all cases, the selection model is significantly better than the neutral model ($p < 0.005$). Sites predicted to be under positive selection with a probability >0.9 are in underlined boldface; 0.8–0.9, underlined; 0.7–0.8, in underlined italics; and 0.5–0.7, in italics.

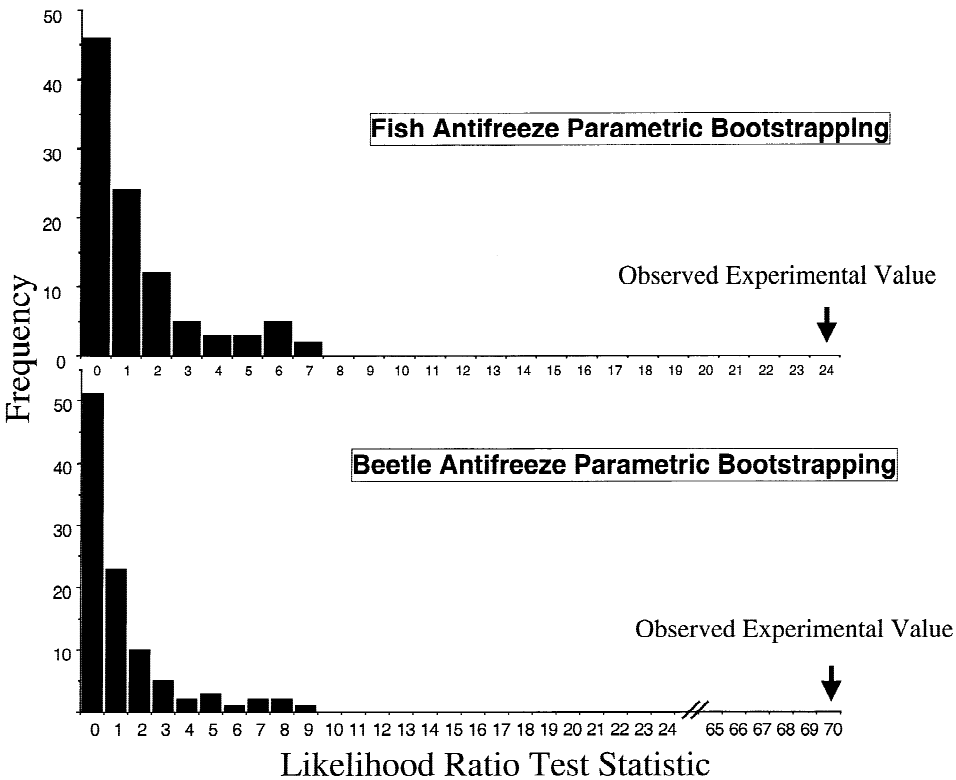

Fig. 3. Parametric bootstrapping demonstrates the significance of the likelihood test statistic. One hundred data sets were simulated under the null model M1 for each fish and beetle AFP gene. The test statistic from the real data falls far outside the distribution of the simulated data, demonstrating the significance of the comparison.

replacement changes per radical nonsynonymous site (p_{NR}) and the number of conservative amino acid replacement changes per conservative nonsynonymous site (p_{NC}) by the methods of Hughes et al. (1990). If there is no selection, then $p_{NR} = p_{NC}$. While under purifying selection $p_{NR} < p_{NC}$ and under positive selection $p_{NR} > p_{NC}$. We found that, for charge changes, p_{NR} was generally greater than p_{NC} (Fig. 4). However, the comparison was only significant for the beetle AFPs, where all but two of the comparisons showed that $p_{NR} > p_{NC}$ and many were significant by a *t* test (data not shown). For polarity and type, p_{NR} was typically less than p_{NC} (Fig. 4), and in many cases significantly different by a *t* test. These results indicate that the positive selection favors charge differences, while maintaining overall type of amino acid most likely to conserve the structure.

Identification of Sites Subjected to Positive Selection

The maximum-likelihood method can identify which residues have been subjected to positive Darwinian se-

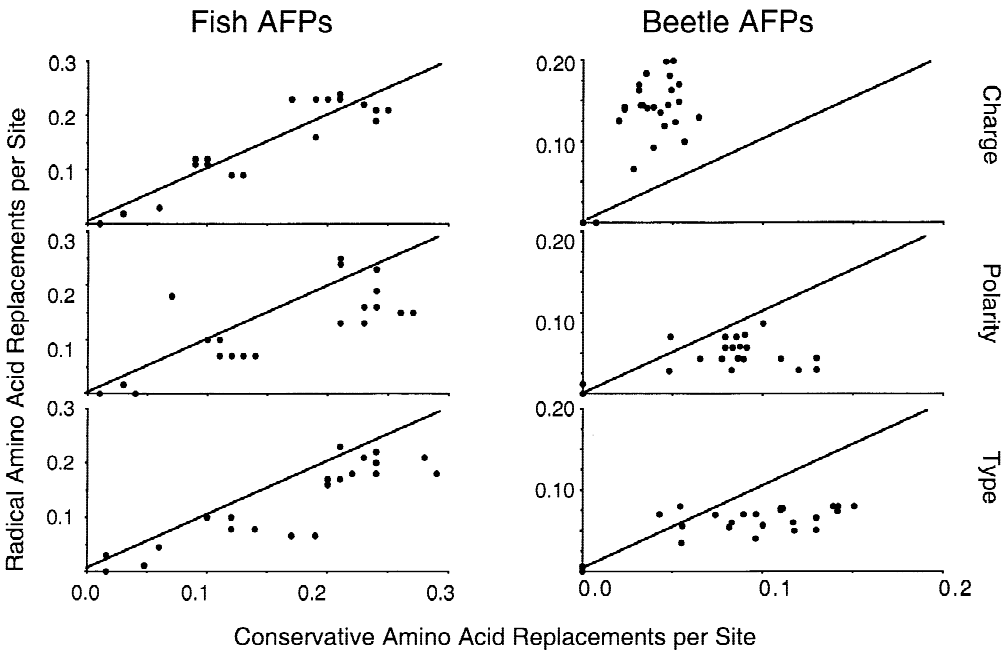

Fig. 4. The number of radical amino acid replacement changes per radical nonsynonymous site (p_{NR}) plotted against the number of conservative amino acid replacement changes per conservative nonsynonymous site (p_{NC}). Three amino acid classifications are used: charge, polarity, and type. Charge typically shows $p_{NR} > p_{NC}$, indicating that selection favors charge changes.

lection using an empirical Bayes approach (Nielsen and Yang 1998). This analysis appears to be very robust and correctly identifies the antigen binding cleft of the MHC Class I glycoprotein as being under strong diversifying selection (Swanson et al. 2001). The sites predicted to be under positive selection under model M8 are given in Table 2 and indicated in Fig. 1. Many of the sites predicted to be under positive selection contain charge charges between paralogues (Fig. 1). Since the crystal structure for the fish type III antifreeze protein has been determined, it is possible to map these positions in three-dimensional space. The 17 amino acid positions predicted to be under selection by maximum likelihood are located on the faces of the fish type III AFP surrounding the ice binding plane, rather than on the putative ice-binding face itself (Fig. 5) (Jia et al. 1996; Graether et al. 1999; DeLuca et al. 1998). Thus, the functional ice-binding surface remains intact. The residues identified as being under positive selection have been shown not be involved directly in ice binding. It is possible that the changes could slightly modulate ice-binding specificity. However, the selective pressure producing the signal of positive selection demonstrated here remains unknown.

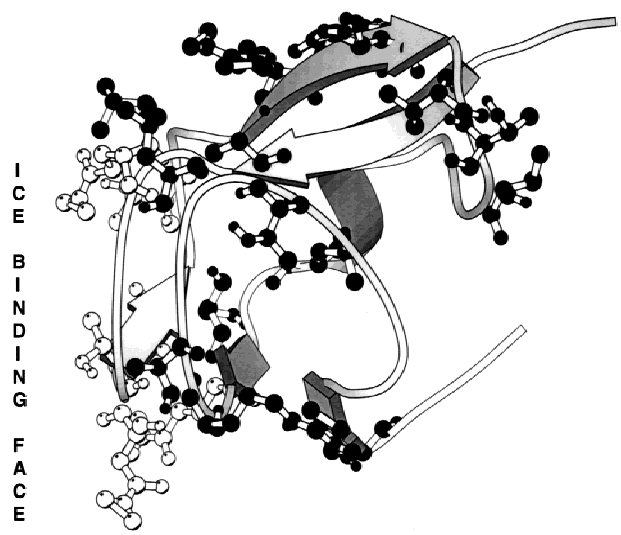

Fig. 5 Fish type III antifreeze protein residues identified as targets of selection with a posterior probability of $p > 0.95$ mapped on the three-dimensional structure 1MSI (Zia et al. 1998). *Black residues showing side chains* were predicated to be subjected to positive selection using a Bayesian approach (Table 1). *White residues showing side chains on the left face* of the structure represent the putative ice-binding residues (Jia et al. 1996; DeLuca et al. 1998; Graether et al. 1999). No other side chains are shown.

Discussion

We have demonstrated that at least some of the striking amino acid diversity among antifreeze proteins of two distinct evolutionary origins has been driven by positive Darwinian selection. Observing positive selection in two independent molecules that have converged upon the same function bolsters the suggestion that the adaptive changes involve a functional advantage in having a heterogeneous suite of AFPs. Of the examples of positive selection documented at the molecular level (e.g., Long and Langley 1993; Endo et al. 1996; Sharp 1997; Nurminsky et al. 1998; Yokoyama et al. 1999; Wyckoff et al.

2000), most involve recognition of other protein molecules in "red queen" scenarios (Endo et al. 1996). For example, host-pathogen recognition involves change by the pathogen to avoid recognition, countered by host change to maintain recognition. Sperm proteins involved in sperm–egg interaction have been shown to be evolving rapidly (Metz et al. 1998), which is hypothesized to be driven by the need to maintain functional interaction with a constantly changing egg surface (Swanson and Vacquier 1998).

The diversification demonstrated here occurs between paralogous members of the antifreeze multigene family. Diversification of members of a multigene family may allow for the binding of a wide variety of molecules. An analogous situation is the diversification of olfactory receptors, which allows for recognition of a wide variety of odorant molecules (Hughes and Hughes 1993). The functional significance of the adaptive changes demonstrated here for the AFPs is unknown but could involve solubility of the AFPs (which are at concentrations of up to 10 mg/ml in the blood), AFP–AFP interaction on the ice surface, other factors associated with inhibition of ice crystal growth, or some other unknown function. Another hypothesis is that the function of AFPs may not yet be optimized, and we could be observing the signal of positive Darwinian selection for increased antifreeze activity. Finally, the selective pressure may involve binding to a heterogeneous surface. Structural studies of the ice surface morphology suggest that, unlike the underlying ice lattice, the ice crystal surface is disordered and morphologically heterogenous (Delzeit et al. 1997; Rowland et al. 1995). However, there is some controversy regarding the structure of the ice surface.

The binding specificity of proteins can in some cases evolve by modulating a conserved binding surface with substitutions at neighboring sites (Wallis et al. 1998; Clackson and Wells 1995; Foote and Winter 1992). For example, antibody specificity can be mediated by both variable loops and substitutions on the protein framework that do not have direct contact with the antigen (Foote and Winter 1992). In particular, the substitutions at neighboring residues often involve charge changes (Wallis et al. 1998; Clackson and Wells 1995), which is the pattern we observe in the beetle AFPs (Fig. 4). This model would allow for specificity to evolve while maintaining a functional ice-binding surface, which is crucial for AFP function and survival of the animal. A fundamental challenge to testing this and other ice-binding hypotheses is the difficulty in quantitative measurements (i.e., determination of binding kinetics) of ice–protein interactions (Greather et al. 1999; Chao et al. 1995). Previous results with mixtures of nonhomologous AFP mixtures did not show increased AFP activity (Chakrabartty and Hew 1991), however, the adaptive value we suggest arises from mixtures of AFPs from paralogous loci within an organism. Since AFPs interact with specific ice planes (Graether et al. 1999) and the binding surface of the AFPs remains intact, the AFP heterogeneity observed may have been selected to interact with morphologically heterogeneous ice crystal growth on specific ice planes. Our hypothesis for the selective pressure diversifying the AFPs is thus consistent with the current models of AFP inhibition of ice crystal growth (Graether et al. 1999). AFPs may present a situation where a biological change is selectively favored to counteract the heterogeneous surface of an ice crystal that is acting as a chemical "pathogen."

Our analyses have demonstrated that at least some of the heterogeneity in AFPs has been driven by positive Darwinian selection. However, the functional significance of these adaptive changes remains unknown. Additionally, the type of selection remains unknown. Future analyses of polymorphism and divergence of orthologous AFP loci could provide insights into whether selection acts to promote diversity within each locus in a manner analogous to MHC or between loci. More importantly, functional studies will be necessary to determine the significance of the adaptive changes identified here.

Acknowledgments. We thank J.P. Devlin, V. Vacquier, R. Nielsen, Z. Yang, M. Wolfner, J. Swanson, G. Feigenson, J. Calkins, and V. Bauer Dumont for discussions. We thank three anonymous reviews for comments. This work was supported by an NSF/Alfred P. Sloan postdoctoral fellowship in molecular evolution and NIH/NRSA postdoctoral fellowship to W.J.S. and N.I.H. Grant GM36431 to C.F.A. Some of these ideas were explored by W.J.S. during a NSF-sponsored course on biological adaptations of Antarctic marine organisms.

References

Chakrabartty A, Hew CL (1991) The effect of enhanced alpha-helicity on the activity of a winter flounder antifreeze polypeptide. Eur J Biochem 202:1057–1063

Chao H, DeLuca CI, Davies PL (1995) Mixing antifreeze protein types changes ice crystal morphology without affecting antifreeze activity. FEBS Lett 357:183–186

Cheng C-HC (1998) Evolution of the diverse antifreeze proteins. Curr Opin Gen Dev 8:715–720

Clackson T, Wells JA (1995) A hot spot of binding energy in a hormone-receptor interface. Science 267:383–386

DeLuca CI, Davies PL, Ye Q, Jia Z (1998) The effects of steric mutations on the structure of type III antifreeze protein and its interaction with ice. J Mol Biol 275:515–525

Delzeit L, Devlin JP, Buch V (1997) Structural relaxation rates near the ice surface: Basis for separation of the surface and subsurface spectra. J Chem Phys 107:3726–3729

Endo T, Ikeo K, Gojobori T (1996) Large scale search for genes on which positive selection may operate. Mol Biol Evol 13:685–690

Foote J, Winter G (1992) Antibody framework residues affecting the conformation of the hypervariable loops. J Mol Biol 224:487–499

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11: 725–736

Graether SP, et al. (1999) Quantitative and qualitative analysis of type III antifreeze protein structure and function. J Biol Chem 274: 11842–11847

Graham LA, Liou Y-C, Walker VK, Davies PL (1997) Hyperactive antifreeze protein from beetles. Nature 388:727–728

Hew C, et al. (1984) Antifreeze polypeptides from the Newfoundland ocean pout, *Macrozoarces americanus:* Presence of multiple compositionally diverse components. J Comp Physiol B 155:81–88

Hew C, et al. (1988) Multiple genes provide the basis for antifreeze protein diversity and dosage in the ocean pout. *Macrozoarces americanus.* J Biol Chem 263:12049–12055

Hughes AL, Hughes MK (1993) Adaptive evolution in the rat olfactory receptor gene family. J Mol Evol 36:249–254

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex loci reveals overdominat selection. Nature 335:167–170

Hughes AL, Ota T, Nei M (1990) Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility complex molecules. Mol Biol Evol 7: 515–524

Jia Z, DeLuca CI, Chao H, Davies PL (1996) Structural basis for the binding of globular antifreeze protein to ice. Nature 384:285–288

Kraulis, PJ (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J Applogr Crystallogr 24:946–950

Li, W-H (1997) *Molecular evolution,* Sinauer, Sunderland, MA

Li X-M, et al. (1985) Structure of an antifreeze polypeptide and its precursor from the ocean pout, *Macrozoarces americanus.* J Biol Chem 260:12904–12909

Liou Y-C, Thibault P, Walker VK, Davies PL, Grahm LA (1999) A complex family of highly heterogeneous and internally repetitive hyperactive antifreeze proteins from the beetle *Tenebrio molitor.* Bioch 38:11415–11424

Long M, Langley CH (1993) Natural selection and the origin of jingwei, a chimeric processed functional gene in *Drosophila.* Science 260:91–95

Messier W, Stewart CB (1997) Episodic adaptive evolution of primate lysozymes. Nature 385:151–154

Metz EC, Robles-Sikisaka R, Vacquier VD (1998) Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. Proc Natl Acad Sci USA 95:10676–10681

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous substitutions. Mol Biol Evol 3:418–426

Nei M, Kumar S (2000) *Molecular evolution and phylogenetics.* Oxford University Press, New York

Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936

Nurminsky DI, Nurminskaya MV, De-Aguiar D, Hartl DL (1998) Selective sweep of a newly evolved sperm-specific gene in *Drosophila.* Nature 396:572–575

Rowland B, et al. (1995) Infrared spectra of ice surfaces and assignment of surface-localized modes from simulated spectra of cubic ice. J Chem Phys 102:8328–8341

Scott GK, Hayes PH, Fletcher GL, Davies PL (1988) Wolfish antifreeze protein genes are primarily organized as tandem repeats that each contain two genes in inverted orientation. Mol Cell Biol 8: 3670–3675

Sharp PM (1997) In search of molecular Darwinism. Nature 385:111–112

Suzuki Y, Gojobori T (1999) A method for detecting positive selection at single amino acid sites. Mol Biol Evol 16:1315–1328

Swanson WJ, Vacquier VD (1998) Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. Science 281: 710–712

Swanson WJ, Yang Z, Wolfner MF, Aquadro CF (2001) Positive Darwinian selection drives the Evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci USA 98:2509–2514

Wallis R, et al. (1998) Specificity in protein-protein recognition: Conserved Im9 residues are the major determinants of stability in the colicin E9 DNase-Im9 complex. Biochemistry 37:476–485

Wang X, DeVries AL, Cheng CH (1995) Antifreeze peptide heterogeneity in an antarctic eel pout includes an unusually large major variant comprised of two 7 kDa type III AFPs linked in tandem. Biochim Biophys Acta 1247:163–172

Wyckoff GJ, Wang W, Wu C-I (2000) Rapid evolution of male reproductive genes in the descent of man Nature 403:304–309

Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15: 568–573

Yang Z (1999) Phylogenetic analysis by maximum likelihood (PAML), Version 2.0j. University College London. http://abacus.gene.ucl.ac.uk/ziheng/paml.html

Yang Z, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol Evol 46:409–418

Yang Z, Nielsen R, Goldman N, Krabbe Pedersen A-M (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Yokoyama S, Zhang H, Radlwimmer FB, Blow NS (1999) Adaptive evolution of color vision of the Comoran coelacanth (*Latimeria chalumnae*). Proc Natl Acad Sci USA 96:6279–6284