Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict

Cheng Deng[a,b], C.-H. Christina Cheng[c,1], Hua Ye[a,b], Ximiao He[b], and Liangbiao Chen[a,1]

[a]Key Laboratory of Molecular and Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China; [b]Graduate College of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100101, China; and [c]Department of Animal Biology, University of Illinois, Urbana, IL 61801

The evolutionary model escape from adaptive conflict (EAC) posits that adaptive conflict between the old and an emerging new function within a single gene could drive the fixation of gene duplication, where each duplicate can freely optimize one of the functions. Although EAC has been suggested as a common process in functional evolution, definitive cases of neofunctionalization under EAC are lacking, and the molecular mechanisms leading to functional innovation are not well-understood. We report here clear experimental evidence for EAC-driven evolution of type III antifreeze protein gene from an old sialic acid synthase (*SAS*) gene in an Antarctic zoarcid fish. We found that an *SAS* gene, having both sialic acid synthase and rudimentary ice-binding activities, became duplicated. In one duplicate, the N-terminal *SAS* domain was deleted and replaced with a nascent signal peptide, removing pleiotropic structural conflict between SAS and ice-binding functions and allowing rapid optimization of the C-terminal domain to become a secreted protein capable of noncolligative freezing-point depression. This study reveals how minor functionalities in an old gene can be transformed into a distinct survival protein and provides insights into how gene duplicates facing presumed identical selection and mutation pressures at birth could take divergent evolutionary paths.

Antarctic eelpouts | thermal hysteresis | tandem repeats | positive selection

Gene duplication is well-recognized as an important source of new genes and functions (1), but the underlying evolutionary mechanisms are far from clear (2–5). Most conceptual models propose that mutational changes, whether neutral [mutation during nonfunctionality (MDN) or duplication degeneration complementation (DDC) model] (3, 6, 7) or directional (adaptational models) (3, 4), occur in the daughter duplicate after gene duplication, leading to subfunctionalization (partitioning of ancestral functions and specialization in one of them) and in rare instances, a new function (neofunctionalization). An alternate model, escape from adaptive conflict (EAC), recognizes that an ancestor with an emergent function besides its primary function could be subject to selection and acquire adaptive changes before gene duplication, but inadvertent pleiotropic conflicts between the two functions constrain further improvements (6, 8). Gene duplication resolves the conflict, allowing daughter duplicates to separately optimize one of the functions (8–10). Resolution of adaptive conflicts created by natural selection as an intrinsic driving force of gene duplication during sub- or neofunctionalization is elegantly logical and may occur frequently, because it potentially applies whenever the ancestor gene experiencing positive selection is a generalist capable of more than one function. In fact, widespread observations of gene sharing and promiscuous function of many enzymes have raised considerable interest in the EAC model (3, 11, 12). However, thus far, only two studies provided evidence of gene duplication under EAC (9, 10). The evolutionary partitioning of the two functions of the *Gal1* gene in yeasts (9) is a case of subfunctionalization of preexisting gene functions between the

daughter genes, whereas in the functional evolution of the dihydroflavonol reductase genes in the morning glory (10), the function proposed to have evolved under EAC remained unidentified. Thus, clear experimental evidence for creation of new function or neofunctionalization under EAC is still missing.

Systems with traceable mutational processes after gene duplication, such as the evolutionarily recent novel fish antifreeze proteins, provide promising avenues for furthering our understanding of the various models governing genic and functional evolution. Antifreeze proteins (AFPs) in different polar marine teleost lineages arose under strong selection from late Cenozoic sea-level glaciation, which protects the fish from death from freezing (13). Several fish AFPs evolved from ancestral genes of unrelated function, and thus, they are clear prima facie cases of neofunctionalization and implicitly embody adaptive conflict resolution as the underlying process. Among these, type III AFPs (AFPIII) of various polar zoarcoid fishes (eelpouts, ocean pouts, and wolffishes) are homologous with the small C-terminal domain of sialic acid synthase (SAS) (14, 15). SAS is an old cytoplasmic enzyme present in microbes (16) through vertebrates (17) that catalyzes intracellular synthesis of sialic acids from N-acetylmannosamine or Man-NAc-6-phosphate and phosphoenolpyruvate (16). In contrast, AFPIIIs are secreted plasma proteins that bind to invading ice crystals and arrest ice growth to prevent fish freezing (13). Enzymatic and antifreeze functions within the same ancestral SAS molecule suggest that an adaptive conflict could arise because of disparate substrate specificity and spatial distribution, and thus, *SAS* to *AFPIII* evolution provides a salient system for investigating the role of EAC in gene duplication and neofunctionalization.

Results and Discussion

We first determined how *AFPIII* evolved from *SAS*. We constructed a bacterial artificial chromosome (BAC) library for the *AFPIII*-bearing Antarctic eelpout *Lycodichthys dearborni* and isolated and sequenced the clones comprising the genomic *SAS* and *AFPIII* loci. The *SAS* and *AFPIII* loci are spatially distinct (Fig. 1). The *SAS* locus (Fig. 1A) contains two *SAS* genes, *LdSAS-A* and *LdSAS-B*, both with six exons. The two *SAS* genes are separated by a chicken repeat (CR)-type retrotransposon (*LdCR1-3*) and are flanked by the *CLTA* and *NCKX2* genes, respectively. This *CLTA-SAS-NCKX2* microsynteny is conserved
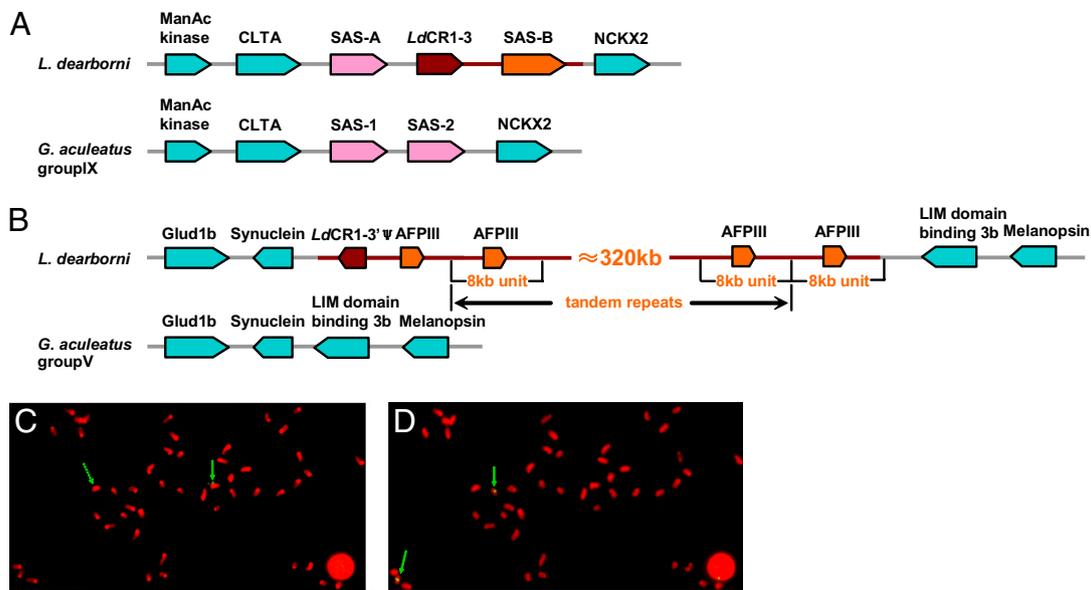
Fig. 1. Genomic organization and chromosomal localization of *SAS* and *AFPIII* loci in the Antarctic eelpout *L. dearborni*. (*A*) *SAS* genomic locus organization is conserved between *L. dearborni* and other teleosts (*G. aculeatus* shown), except for the insertion of a chicken repeat-type retrotransposon (*LdCR1-3*) between the two *SAS* genes in the *L. dearborni* locus. (*B*) *AFPIII* locus of *L. dearborni,* ~400 kbp in size, occurs in the middle of the shared four-gene (gene names as indicated) microsynteny with non-AFPIII teleost (*G. aculeatus* shown). The locus comprises a 5′ pseudogene ψ*AFPIII* followed by >30 *AFPIII* genes predominantly arrayed in 8-kbp tandem repeats. The regions spanned by a brown line (including genes) in the *L. dearborni SAS* locus (*LdCR1-3* and *LdSAS-B*) and the *AFPIII* locus (5′-truncated *LdCR1-3* and *AFPIII* tandem repeats) share strong sequence homology. The relevant BAC sequences were deposited in GenBank under accessions numbers GQ368892, GQ368893, and GQ368894. (*C* and *D*) FISH localized *L. dearborni SAS* (*C*) and *AFPIII* (*D*) loci in separate chromosome pairs, indicative of an interchromosomal translocation of the *AFPIII* progenitor gene.

in annotated fish genomes, as illustrated for *Gasterosteus aculeatus* (Fig. 1*A*). *L. dearborni* and *G. aculeatus* (both percomorphs) have two *SAS* genes, whereas other teleost species that we examined have one copy. The *AFPIII* locus spans ~400 kbp and contains an estimated 30 or more *AFPIII* genes (Fig. 1*B*) that share 97–99% nucleotide sequence identities. *AFPIII* genes have a two-exon structure encoding a signal peptide (exon1) and the mature ice-binding AFPIII (exon2). Except for the 5′ fringe copy Ψ*AFPIII*, they are arrayed in ~8-kbp tandem repeats with one gene per repeat (ref. 18 and this study) (Fig. 1*B*), indicative of gene family expansion through in situ tandem duplications. The 5′ copy Ψ*AFPIII* (Fig. 1*B*), despite having a 214-nt frame-shift insertion, bears greater sequence similarity to the *SAS* homologs (*SAS* exon6) (Fig. S1) and is immediately preceded by a partial CR1-3 retrotransposon (*LdCR1-3′*) similar in sequence to *LdCR1-3* in the *SAS* locus; thus, we reasoned it to be the most deeply diverging member of the *AFPIII* gene family. The *AFPIII* locus is flanked at the two ends by the genes *Glud1b* and *Synuclein* and *LIM domain binding 3b* and *Melanopsin*, respectively. This four-gene microsynteny (without *CR1-3′* and *AFPIII* locus in its middle) is conserved in other teleosts, as illustrated for *G. aculeatus* (Fig. 1*B*). Chromosomal *fluorescence in situ hybridization* (*FISH*) localized *L. dearborni SAS* and *AFPIII* loci to distinct metaphase chromosome pairs (Fig. 1 *C* and *D*). Collectively, these results strongly suggest that a genomic region (estimated at ~12 kbp) containing the *LdSAS-B* gene and its immediate neighbor sequences (including *LdCR1-3*) was duplicated and translocated to a site between *Synuclein* and *LIM domain binding 3b* genes; from this, the primordial *AFPIII* gene evolved, and the large *AFPIII* locus arose from in situ gene family expansion under selection pressure from polar sea-level glaciation.

Through comparative analyses of *LdSAS-A*, *LdSAS-B*, and Ψ*AFPIII* sequences, we deduced that *LdSAS-B* is the closest relative of the *AFPIII* progenitor and the molecular events in the *LdSAS-B* to *AFPIII* transformation (Fig. 2). The six-exon *LdSAS-A* and *LdSAS-B* genes share high nucleotide identities

(~92% between exons and ~67% between introns) but have differentiated in their 5′ flanking region (5′FR; ~594 nt) with no sequence homology (Fig. 2, blue line and Fig. S2*A*). Ψ*AFPIII* shares greater nucleotide identities with *LdSAS-B* (Fig. 2) than with *LdSAS-A* (Fig. S2 *B* and *C*), including 74% identity to the *LdSAS-B* 5′FR, which is nonhomologous between *LdSAS-B* and *LdSAS-A*. Thus, *AFPIII* 5′FR, intron1 (I1), exon2 (E2; ice-binding mature AFPIII), and 3′FR were derived from the 5′FR, I5, E6 (*SAS* C-terminal domain), and 3′FR, respectively, of the ancestral *LdSAS-B* (Fig. 2 and Fig. S2*B*). The emerging *AFPIII* would require a signal peptide for extracellular export of the mature protein. We discovered a precursor signal peptide coding sequence appropriately located in the extant *LdSAS-B*, starting from 54 nt upstream of the translation start site through the first six codons of *LdSAS-B* E1 (Fig. 2 and Fig. S2*B*). An intragenic deletion from the seventh codon of E1 through E5 of *LdSAS-B* and linkage of the new E1, the old I5, and E6 would complete the formation of the nascent two-exon *AFPIII* gene encoding the secretory antifreeze protein.

We then examined for sequence and functional properties that might have compelled the ancestral *LdSAS-B* duplication and neofunctionalization of one duplicate into *AFPIII*, and we found strong evidence that would fulfill the predictions of the EAC model. EAC predicts that (*i*) the ancestor is bifunctional and subject to selection before gene duplication, (*ii*) adaptive conflict between the ancestral and new function constrains improvement of the selected function(s) before duplication, and (*iii*) adaptive changes and functional improvement occur in the daughter genes after duplication. To assess *LdSAS-B* bifunctionality (prediction *i*), we cloned the *SAS* cDNAs from the liver and expressed the recombinant proteins in bacteria. We found that the recombinant SAS-B indeed has incipient antifreeze activity (Fig. 3). A single seed ice crystal in pure water (without antifreeze or other ice-active compounds) grows as a discoid and continues to expand at 0 °C, the equilibrium freezing point (fp) (Fig. 3 *A* and *B*). In contrast, *LdSAS-B* at 2 mg/mL causes strong hexagonal faceting of the test
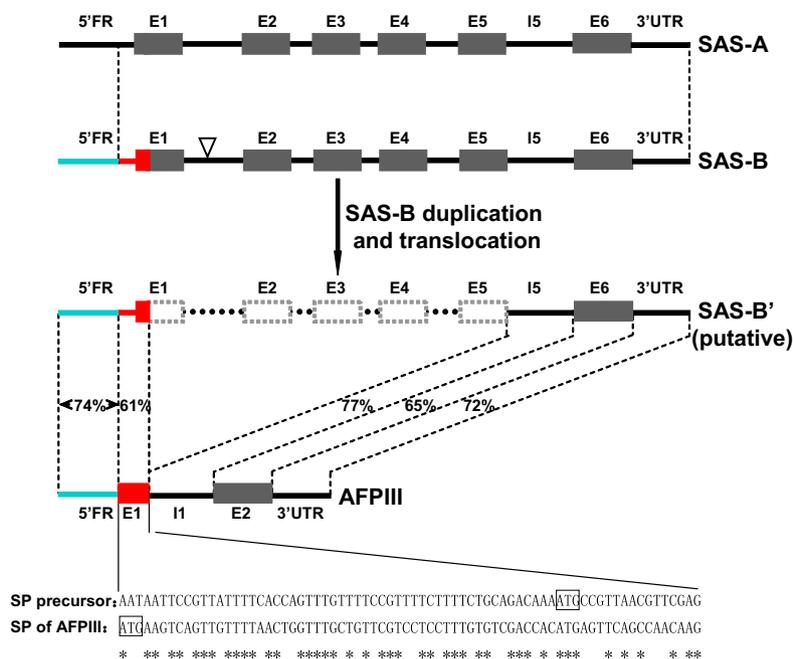
Deng et al.

Fig. 2. Molecular process of the evolution of *AFPIII* from *SAS-B*. One daughter *SAS-B* duplicate (*SAS-B'*) underwent N-terminal domain deletion (seventh codon of E1 through E5) and neofunctionalization into *AFPIII*. Regions in *SAS-B'* corresponding to the regions in the two-exon *AFPIII* gene are indicated with the same colors for the two genes, with nucleotide sequence identities given. The partly nonprotein coding signal peptide (SP) precursor sequence in *SAS-B'* that was modified to become a coding sequence for the AFPIII signal peptide is shown at the bottom. *Ld*SAS-A lacks the 5' flanking sequence homology (blue bar) with *Ld*SAS-B and *AFPIII*; thus, it is not the evolutionary progenitor to *AFPIII*.

ice crystal at the equilibrium fp, indicative of ice binding and growth interference by the protein at or near the prism faces (Fig. 3*C*). *Ld*SAS-B could also inhibit bulk ice growth at small cooling, producing a nonequilibrium fp depression ranging from 0.004 °C (Fig. 3*D*) to 0.015 °C (Fig. 3*E*). When temperature was lowered below nonequilibrium fp, ice grew rapidly as a hexagonal disk (Fig. 3*F*). Purified AFPIII protein from *L. dearborni* at the same concentration also caused strong faceting of the seed ice crystal (as a hexagonal bipyramid) at the equilibrium fp (approximately −0.0005 °C) (Fig. 3*G*), but in marked contrast to *Ld*SAS-B, bulk growth of the ice crystal was effectively arrested at large cooling (0.39 °C to 0.67 °

Fig. 3. Ice-binding activity of the recombinant *Ld*SAS-B compared with native AFPIII. Single ice-crystal growth behavior in (*A* and *B*) water, (*C–F*) *Ld*SAS-B (2 mg/mL), and (*G* and *H*) *Ld*AFPIII (2 mg/mL). The crystal morphology of the test ice at the equilibrium fp of water (*A*) is discoid but strongly faceted in *Ld*SAS-B (*C*; hexagonal) and the eelpout antifreeze (*G*; hexagonal bipyramid), indicating growth inhibition at the prism faces. The ice crystal in water expands unrestricted at the equilibrium fp of water (*B*). In contrast, ice expansion is inhibited by *Ld*SASB at its equilibrium fp (*C* and *D*) and at small coolings (~0.004 °C to 0.015 °C) below the equilibrium fp (*E*); at the nonequilibrium fp, (*F*) ice expands quickly as a hexagonal disk. In comparison, the bona fide antifreeze protein *Ld*AFPIII is much more effective in arresting ice-crystal growth to temperatures significantly (*H*, 0.39 °C; *I*, 0.67 °C; data from one of three replicates) below the equilibrium fp until fast spicular growth occurs when the nonequilibrium fp is reached (*J*).

C below the equilibrium fp) (Fig. 3 *H* and *I*). Only when the temperature reached the nonequilibrium fp did the unrestricted burst spicular growth characteristic of AFP solutions occur (Fig. 3*J*). The fp depression activity was measured to be ~0.58 °C ± 0.08 °C for 2 mg/mL *Ld*AFPIII solution and 1.46 °C ± 0.15 °C for 20 mg/mL solution. This large noncolligative fp depression or thermal hysteresis is the unique property of a bona fide antifreeze protein. To determine which structural domain in the *Ld*SAS-B protein contributed to the ice-binding activity, we chemically synthesized its C-terminal domain (69-residue E6 peptide), which is homologous to mature AFPIII, and tested for its ice-binding ability. We found similar ice-binding activity as the full-length *Ld*SAS-B protein, indicating that the ice-binding activity is an intrinsic property of the SAS C-terminal domain (Fig. S3). To verify that the *Ld*SAS-B–encoded protein possesses SAS enzymatic activity, we measured the sialic acid synthetic function using the recombinant *SAS* proteins. Indeed, recombinant SAS-B catalyzed in vitro sialic acid synthesis at twofold greater activity (2.7 ± 0.2 mU/mg) than recombinant *Ld*SAS-A (1.4 ± 0.2 mU/mg) (Fig. S4*B*). Thus, *Ld*SAS-B clearly possesses two activities—the original SAS function and a minor ice-binding activity in its C-terminal domain.

To assess EAC prediction *iii* that daughter duplicates released from adaptive conflict acquire adaptive changes that would improve function, we used the branch model and modified branch-site model A (19) to test for positive selection on the *Ld*SAS-B and *Ld*AFPIII lineages and found its occurrence in both (Table 1 and Fig. S5). The input *SAS* tree includes various teleost *SAS* coding sequences from GenBank and *SAS-A* and *SAS-B* orthologs from the sympatric Antarctic eelpout *Pachycara brachyce-phalum* that we obtained in this study (Fig. S5*A*). The foreground branch (leading to the Antarctic eelpout *SAS-B* clade) has a large nonsynonymous (dN)/synonymous (dS) substitution rate ratio (branch-site dN/dS of $\omega 2 = 9.44$) that is highly significant [likelihood ratio tests (LRT) $P < 0.01$] (Table 1), whereas all *SAS* lineages, including eelpout *SAS-A* genes, have $\omega < 1$, (Fig. S5*A*), suggesting positive Darwinian selection occurring specifically in
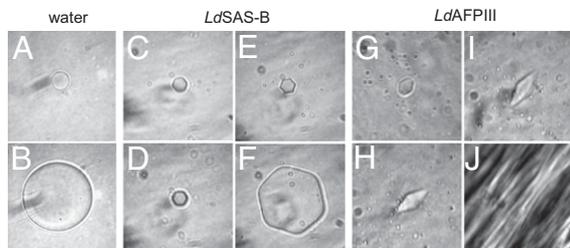
Table 1. The parameters and statistical significances of likelihood ratio tests in the branches of *SAS-B* and *AFPIII*

Foreground branch	Parameters ($\omega1 = 1$; $0 < \omega0 < 1$; $\omega2$ to be estimated)	$2\Delta\ln L$ (LRT; df = 1)	Positive selected sites
SAS-B	p0 = 0.76; $\omega0$ = 0.04; p1 = 0.08; p2 = 0.16; $\omega2$ = 9.44	27.3 ($P = 0.0000$)	135L,* 171I,[†] 228A,[†] 251S,[†] 260K,* 268E,[†] 286A,[†] 327L,*[‡] 328C,[†‡] 351K*[‡]
AFPIII	p0 = 0.37; $\omega0$ = 0.19; p1 = 0.00; p2 = 0.63; $\omega2$ = infinity	16.7 ($P = 0.0000$)	8N,[†] 13I,* 14N,* 16A,[†] 20I,* 22M[†] 28T,* 40I,[†] 44Q,[†] 48A,[†] 50P,[†] 54T,[†] 58D[†]

*$P > 0.99$.
[†]$P > 0.95$.
[‡]Residues in the SAS-B gene in the C-terminal (exon6) domain.
   p, Bayes Empirical Bayes (BEB) posterior probability; LTR, likelihood ratio tests; df, degree of freedom.

the Antarctic eelpout *SAS-B* genes. Ten residues in *SAS-B*, mostly (seven) in the N-terminal domain, were identified to be under positive selection (Table 1), suggesting adaptive evolution of the SAS-B enzymatic function after the ancestral SAS-B gene duplication. Likewise, for the tree of homologous *AFPIII* and *SAS* E6 sequences (Fig. S5*B*), we detected a large ω (branch site $\omega2$ = infinite; LRT $P < 0.01$) in the zoarcoid *AFPIII* branch, with a much greater percentage (P_2 = 63%) of residues under positive selection than in the eelpout *SAS-B* branch (P_2 = 16%) (Table 1). This suggests that, after gene duplication and deletion of the N-terminal domain coding sequence of the translocated *SAS-B* daughter duplicate, adaptive amino acid changes in the nascent *AFPIII* gene accelerated, rapidly improving the antifreeze function to the full-fledged thermal hysteresis illustrated in Fig. 3. Nozawa et al. (20) recently questioned the statistical basis of the branch-site model and claimed that it generates too many false positives. However, the false-positive rate found by those authors (20) was lower than the significance level (5%) (21). The branch-site test has been successfully applied in many studies, generating interesting biological hypotheses that have been validated by experimentation (22–24). In this study, 7 of 10 amino acids in SAS-B and 8 of 12 in AFPIII detected under positive selection (Table 1) showed hydrophobicity or charge switches from the corresponding sites of their paralogs or the precursor, implying biochemical property or activity changes. Indeed, mutational analyses of AFPIII have shown these amino acids, such as I13, N14, A16, Q44, and 61K, to be functionally important for antifreeze activity (25, 26).

The foregoing results added to existing zoarcoid *AFPIII* structural information, in turn, shed light on the selection and adaptive conflict that might have transpired in the ancestral *SAS-B* before gene duplication (EAC predictions *i* and *ii*). Of the six conserved residues (Q9, T18, A16, T15, Q44, and N14) in Atlantic ocean pout AFPIII identified through structural studies to constitute a putative flat ice-binding surface (25, 27), two (T15 and T18) are ancestral, preexisting in the C-terminal domain of both SAS-A and SAS-B (T305 and T308, respectively) of *L. dearborni* (Fig. S6) and other teleost SAS. These might have constituted an accidental structural basis for rudimentary ice affinity in SAS. Indeed, *L. dearborni* SAS-A and its E6 peptide also has incipient ice-binding activity similar to *Ld*SAS-B (Fig. S3). However, only in *SAS-B* are adaptive residue changes detected (Table 1). One of the three adaptive changes detected in the *Ld*SAS-B C-terminal domain (Table 1, double dagger), K351 (as opposed to D351 in *Ld*SAS-A and other teleost SAS) corresponds to K61 in AFPIII. K61 in AFPIII forms intramolecular hydrogen bonds with two putative ice-binding residues (Q44 and N14) that are important in stabilizing the flat ice-binding surface (25) or the global protein fold (27), and both are essential for antifreeze activity. Thus, the D351/K351 change in SAS-B might have resulted from positive selection on the ancestral *SAS-B* before gene duplication for improving the accidental ice affinity of its small C-terminal domain. To test whether further improvement of ice-binding activity in the ancestral SAS-B would likely create conflict for its original

SAS function, we substituted four residues (V299, G304, V306, and T334) in the *Ld*SAS-B C-terminal domain with their homologs (Q9, N14, A16, and Q44) in the AFPIII ice-binding surface to mimic adaptive changes to form the AFPIII active site in the C terminal. This mutant (*Ld*SAS-Bm4) showed no detectable SAS activity (Fig. S4*B*), supporting our hypothesis that residue change to ice-binding capability in the C-terminal domain of the bifunctional ancestor would create pleiotropic conflict. The conflict can be explained by the structure of the SAS holoenzyme. *SAS* monomer structure resembles an asymmetric dumbbell (16), and the active enzyme consists of a dimer or tetramer through juxtaposition of the N-terminal domain of one monomer with the C-terminal domain of another, forming the site for substrate (sugar) binding (28) between the interacting surfaces of the swapped domains of each monomer (Fig. S7). It is clear from these studies and the inactive *Ld*SAS-Bm4 in this study that structural coordination between the C- and N-terminal domains of the SAS neighbors is essential for SAS enzyme activity. In addition, for the emerging AFPIII function, the bulky nonice-active N-terminal domain of the ancestral SAS would very likely hinder the small E6 domain in contacting ice crystals because of the propensity of the SAS monomers to form dimers. Furthermore, to test the hypothesis that improvement of the SAS enzyme function in *Ld*SAS-B might conflict with the ice-binding activity of its C-terminal domain, we measured the fp depression capability of recombinantly expressed *Ld*SAS-A (Fig. S3), the outgroup of the *Ld*SAS-B and AFPIII genes (Fig. S5) and the putative ancestral SAS state, to infer the original level of ice-binding activity in the ancestral SAS-B before gene duplication. We found that the noncolligative fp depression by *Ld*SAS-A ranged from 0.004 °C to 0.07 °C and that the noncolligative fp depression by *Ld*SAS-B has a lower maximum, between 0.004 °C and 0.015 °C (both at 2 mg/mL concentration, the maximum solubility of the recombinant SAS proteins in aqueous buffer). The lower fp depression capability observed of *Ld*SAS-B compared with *Ld*SAS-A is consistent with improvement of SAS enzyme function in *Ld*SAS-B (2.7 ± 0.2 mU/mg vs. 1.4 ± 0.2 mU/mg) (Fig. S4*B*) after the gene duplication having an adverse effect on the ice-binding function, supporting the presence of adaptive conflict within the ancestral SAS-B. These conflicts were resolved through a duplication of the ancestral *SAS-B*, whereby each daughter duplicate could freely improve one of the functions. Additionally, conflict resolution was likely quickened and made permanent by deleting the SAS N-terminal domain coding region in the presumptive AFPIII duplicate, eliminating potential dimer formation between the two paralogs. Accelerated adaptive changes subsequently occurred in the nascent *AFPIII* gene, indicated by nearly two-thirds of the residues in *AFPIII* experiencing positive Darwinian selection (Table 1); this resulted in rapid optimization to a full-activity AFP capable of preventing freezing of the fish body fluids.

The genesis of the extracellular secretory signal in the primordial *LdAFPIII* gene by incorporating a piece of 5′ UTR into a functional protein coding sequence with some modifications (Fig. 2) is an added evolutionary innovation. The capacity to

code for a functional signal peptide (SP), in fact, existed in the precursor sequence in *LdSAS-B*, because we found comparable levels of AFPIII-exporting activity by the SP precursor-mature *AFPIII* construct and the native pre-*AFPIII* cDNA (Fig. S8). Thus, the evolution of Antarctic eelpout *AFPIII* has entailed tapping into two inconspicuous functionalities in the same cytoplasmic ancestor and remarkably, transforming them into a quintessential lifesaving antifreeze function. The partly nonprotein coding origin of *AFPIII* signal peptide represents an example of recruiting a hidden function by incorporating a translation start site, and sheds light on how products of duplicated genes could be targeted to different cellular localizations.

A remaining question is how the weakly ice-active ancestral SAS protein could be initially beneficial for natural selection to act on. Observations from our laboratory testing of SAS ice inhibition offers an explanation. Ice growth in SAS solution could be inhibited at temperatures slightly below its equilibrium fp when the seed ice crystals were very small (∼10 μm) and the cooling rate was very slow (∼0.0009 °C/min). These two conditions quite certainly apply in the wild over evolutionary time. Sea surface ice formed before water temperatures of the Antarctic Ocean reached freezing in its entirety, and thus, ice crystals that could appear in the water column were likely thermally unstable and therefore, small. The cooling rate of the polar waters was very slow, an overall cooling of ∼12 °C in the past 50 Myr in the Antarctic deep water (29), compared with the laboratory rate of ∼0.0009 °C/min (473 °C/y). Thus, the incipient ice activity of the ancestral SAS molecule or its detached C-terminal peptide would be quite sufficient in inhibiting ice-crystal expansion in the fish body fluids. By the time water temperatures chilled to the equilibrium fp of the fish and fp depression became essential for survival, one can assume that the ongoing process of natural selection would have led to the refinement of the incipient ice-binding activity to a full-fledged antifreeze protein.

To summarize, through tracing the processes in the SAS-B to AFPIII evolution in the Antarctic eelpout, we provided strong and comprehensive molecular and functional evidence for a clear example of EAC-compelled duplication of a bifunctional ancestral gene and additionally, acceleration of conflict resolution through intragenic domain deletion in one duplicate and its neofunctionalization into a protein of distinctive function. It also reveals that an ancestral molecule (*SAS*) can be subject to a separate modality of natural selection (advent of freezing marine conditions) for an accidental functional property (ice binding) concomitant with preexisting cellular selection on improving the ancestral function (enzymatic). Thus, this study provides a fresh evolutionary perspective that gene duplicates, although bearing identical sequence at birth, could be exposed to divergent mutational and selection pressures that propel them on different evolutionary trajectories (2). The evolutionary process of zoarcid *AFPIII*, in total or part, likely typifies the evolution of other antifreeze proteins from functionally unrelated ancestors. For example, the antifreeze glycoprotein (*AFGP*) gene in Antarctic notothenioid fishes evolved from a duplicated trypsinogen-like protease (*TLP*) gene and entailed intragenic deletion of most of the coding region for the protease in the chimeric *AFGP-TLP* evolutionary intermediate (30, 31). Type II AFPs in sea raven, herring, and smelt are homologous with the carbohydrate-binding domain of C-type (Ca²⁺-dependent) lectins (13, 32), and thus, EAC-driven gene duplication may apply to their evolution. In principal, any protein gene with more than one function that may be at odds with each other because of incompatible structural requirements or regulation of expression is candidate for EAC-based gene duplication and subsequent optimization. We envisage further interesting empirical investigations and evidence for EAC as an important and common underlying mechanism in generating genetic and functional novelty in time to come.

Materials and Methods

BAC Library Construction, Screening, and Sequencing. *L. dearborni* was caught with deep-water traps from McMurdo Sound, Antarctica. A BAC library containing >120,000 clones with insert sizes ranging from 100 to 200 kbp was constructed using the Copy Control pCC1BAC kit (Epicentre) following published protocols with optimization (33). The library was screened for SAS and AFPIII clones using respective cDNA probes; 8 *SAS*-positive and 48 *AFPIII*-positive BAC clones were obtained. The positive clones were fingerprinted using the SNaPshot Multiplex kit (Applied Biosystems). The labeled restriction fragments were resolved on an ABI 3730 sequencer, and the elution profiles were collected using GeneMapper version3.5. The data were cleaned of vector sequence using Genoprofiler (34) and assembled with FPC version 8 software (35). A minimal tiling path of one and seven BAC clones covered the *SAS* and *AFPIII* locus, respectively. A shotgun library of 1.5- to 2-kbp inserts was constructed for each BAC clone using a pUC18 vector for sequencing. The relevant clones were sequenced at 7× coverage, with estimated sequencing accuracies greater than 99.98% for all clones. Sequence contigs of each BAC clone insert were assembled using the Phred/Phrap/Consed package (36). Gene annotation was obtained by BLAST, and the contigs were compared with the National Center for Biotechnology Information (NCBI) database.

Chromosomal FISH of *AFPIII* and *SAS* Genes. The full-length digoxigenin-labeled *AFPIII* gene probe was hybridized to metaphase chromosomal preparations from *L. dearborni* head kidney and spleen cells following previously published protocol (37). The same slide was stripped of the *AFPIII* probe after visualization and image capture, and an *SAS* gene probe containing exon2–5 was applied for another round of hybridization under the same conditions (37).

Detection of Ice-Binding Activity of Recombinant SAS Proteins and Their C-Terminal Domains. The full-length cDNAs for *LdSAS-A* and *LdSAS-B* were directionally cloned into the EcoRI and SalI sites of pET28a+ (Novagen). The two recombinant plasmids were transformed into *Escherichia coli* Rosetta (DE3) for protein expression. The expressed proteins were purified by chelating a Sepharose Fast Flow gel column charged with Ni²⁺ and dialyzed. Sixty-nine residues of the C-terminal domain encoded by exon6 of *LdSAS-A* and *LdSAS-B* genes (homologous to mature AFPIII) were chemically synthesized and purified by HPLC. The ice-binding activity of the SAS proteins and synthetic peptides was assessed using the Clifton nanoliter osmometer cryoscope equipped with image capture (38). The activities of purified native *L. dearborni* AFPIII were also determined for comparison.

Selection Analysis of *SAS* and *AFPIII* Genes. SAS- and AFPIII-translated amino acid sequences from Antarctic eelpouts *L. dearborni* and *P. brachycephalum* and other teleost species were aligned with ClustalW version 1.83 (39) with default settings, and the results were converted to nucleotide sequence alignments. Phylogenetic trees were constructed using three distinct algorithms: neighbor joining (NJ) with 1,000 bootstrap replicates in MEGA version 4 (40), maximum likelihood (ML) in PAUP version 4.0b10 (41), and Bayesian inference (BI) in MrBayes version 3.1.2 (42). The best substitution model for ML and BI was evaluated using Modeltest version 3.7 (43) and MrModelTest version 2.2 (44), respectively. The best ML tree was determined using heuristic search. PAML version 4 was used to test for positive selections on interested branches and identify the codons under selection (45).

Test of Sialic Acid Synthetic Activity of Recombinant *LdSAS-A*, *LdSAS-B*, and *LdSAS-B* Mutant. The recombinant *LdSAS-A* and *LdSAS-B* were prepared as above. An *LdSAS-B* mutant (*SAS-Bm4*) with four amino acids substitutions (V299Q, G304N, V306A, and T334Q) in the C-terminal domain was created by chemical synthesis of the gene (Sangon). Together with the two preexisting threonine residues (T305 and T308), *SAS-Bm4* contains all six amino acids comprising the ice-binding surface of a functional AFPIII, and thus, this mutant mimics the evolutionary conversion of SAS-B C terminal to AFPIII in the active sites. *SAS-Bm4* was directionally cloned into the same sites in pET28a+ as the *LdSAS* genes, and the authenticity of all three expression clones was verified by sequencing. The expression and purification followed the same protocol as *LdSAS-A* and *LdSAS-B*. Sialic acid synthetic activity of each protein was assayed using a classic method previously described (46).

Test of the Secretion Functionality of the Precursor Signal Peptide. The 5′sequence in *LdSAS-B* gene (signal peptide precursor) homologous to AFPIII signal peptide coding sequence was spliced with exon2 of AFPIII and cloned into the expression vector pCS2-flag4 vector with a built-in flag tag (termed

EVOLUTION

SP precursor construct); the native pre-AFPIII cDNA (pre-AFPIII construct) and the second exon without any leading sequence (non-SP construct) were similarly constructed. These constructs and the vector were separately transfected into HEK293T cells and cultured for 3 d; 12 μL culture medium from each transfection were resolved on SDS/PAGE gel followed with immunostaining of the blotted gel for the flag tag to detect recombinant protein secretion (details in Fig. S8).

1. Ohno S (1970) *Evolution by Gene Duplication* (Springer, New York).
2. Lynch M, Katju V (2004) The altered evolutionary trajectories of gene duplicates. *Trends Genet* 20:544–549.
3. Conant GC, Wolfe KH (2008) Turning a hobby into a job: How duplicated genes find new functions. *Nat Rev Genet* 9:938–950.
4. Hahn MW (2009) Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered* 100:605–617.
5. Zhang J (2006) Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. *Nat Genet* 38:819–823.
6. Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. *Proc Biol Sci* 256:119–124.
7. Force A, et al. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.
8. Piatigorsky J, Wistow G (1991) The recruitment of crystallins: New functions precede gene duplication. *Science* 252:1078–1079.
9. Hittinger CT, Carroll SB (2007) Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449:677–681.
10. Des Marais DL, Rausher MD (2008) Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 454:762–765.
11. Barkman T, Zhang J (2009) Evidence for escape from adaptive conflict? *Nature* 462: E1.
12. Des Marais DL, Rausher MD (2009) Des Marais and Rausher reply. *Nature* 462:E2–E3.
13. DeVries AL, Cheng CHC (2005) Antifreeze proteins and organismal freezing avoidance in polar fishes. *Fish Physiology*, eds Anthony PF, John FS (Academic Press), Vol 22, pp 155–201.
14. Baardsnes J, Davies PL (2001) Sialic acid synthase: The origin of fish type III antifreeze protein? *Trends Biochem Sci* 26:468–469.
15. Lander ES, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
16. Gunawan J, et al. (2005) Structural and mechanistic analysis of sialic acid synthase NeuB from Neisseria meningitidis in complex with Mn2+, phosphoenolpyruvate, and N-acetylmannosaminitol. *J Biol Chem* 280:3555–3563.
17. Reaves ML, Lopez LC, Daskalova SM (2008) Replacement of the antifreeze-like domain of human N-acetylneuraminic acid phosphate synthase with the mouse antifreeze-like domain impacts both N-acetylneuraminic acid 9-phosphate synthase and 2-keto-3-deoxy-D-glycero-D-galacto-nonulosonic acid 9-phosphate synthase activities. *BMB Rep* 41:72–78.
18. Wang X, DeVries AL, Cheng CH (1995) Genomic basis for antifreeze peptide heterogeneity and abundance in an Antarctic eel pout: Gene structures and organization. *Mol Mar Biol Biotechnol* 4:135–147.
19. Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22: 2472–2479.
20. Nozawa M, Suzuki Y, Nei M (2009) Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proc Natl Acad Sci USA* 106:6700–6705.
21. Yang Z, Nielsen R, Goldman N (2009) In defense of statistical methods for detecting positive selection. *Proc Natl Acad Sci USA* 106:E95.
22. Shen YY, et al. (2010) Adaptive evolution of energy metabolism genes and the origin of flight in bats. *Proc Natl Acad Sci USA* 107:8666–8671.
23. Briscoe AD, et al. (2010) Positive selection of a duplicated UV-sensitive visual pigment coincides with wing pigment evolution in Heliconius butterflies. *Proc Natl Acad Sci USA* 107:3628–3633.
24. Aagaard JE, Yi X, MacCoss MJ, Swanson WJ (2006) Rapidly evolving zona pellucida domain proteins are a major component of the vitelline envelope of abalone eggs. *Proc Natl Acad Sci USA* 103:17302–17307.
25. Jia Z, DeLuca CI, Chao H, Davies PL (1996) Structural basis for the binding of a globular antifreeze protein to ice. *Nature* 384:285–288.
26. Chao H, Sönnichsen FD, DeLuca CI, Sykes BD, Davies PL (1994) Structure-function relationship in the globular type III antifreeze protein: Identification of a cluster of surface residues required for binding to ice. *Protein Sci* 3:1760–1769.
27. Yang DS, et al. (1998) Identification of the ice-binding surface on a type III antifreeze protein with a "flatness function" algorithm. *Biophys J* 74:2142–2151.
28. Huang HH, et al. (2005) Structural characterization of sialic acid synthase by electrospray mass spectrometry—a tetrameric enzyme composed of dimeric dimers. *J Am Soc Mass Spectrom* 16:324–332.
29. Lear CH, Elderfield H, Wilson PA (2001) Cenozoic deep-sea temperatures and global ice volumes from Mg/Ca in Benthic foraminiferal calcite. *Science* 287:269–272.
30. Cheng CHC, Chen L (1999) Evolution of an antifreeze glycoprotein. *Nature* 401: 443–444.
31. Chen L, DeVries AL, Cheng CHC (1997) Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc Natl Acad Sci USA* 94: 3811–3816.
32. Ewart KV, Fletcher GL (1993) Herring antifreeze protein: Primary structure and evidence for a C-type lectin evolutionary origin. *Mol Mar Biol Biotechnol* 2:20–27.
33. Amemiya CTOT, Litman GW (1996) *Construction of P1 Artificial Chromosome (PAC) Libraries from Lower Vertebrates* (Academic Press, San Diego), pp 223–256.
34. You FM, et al. (2007) GenoProfiler: Batch processing of high-throughput capillary fingerprinting data. *Bioinformatics* 23:240–242.
35. Luo MC, et al. (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. *Genomics* 82:378–389.
36. Gordon D (2002) Viewing and editing assembled sequences using Consed (John Wiley & Sons, Inc.), pp unit 11.12.11–11.12.43.
37. Jiang J, Gill BS, Wang GL, Ronald PC, Ward DC (1995) Metaphase and interphase fluorescence in situ hybridization mapping of the rice genome with bacterial artificial chromosomes. *Proc Natl Acad Sci USA* 92:4487–4491.
38. Jin Y, DeVries AL (2006) Antifreeze glycoprotein levels in Antarctic notothenioid fishes inhabiting different thermal environments and the effect of warm acclimation. *Comp Biochem Physiol B Biochem Mol Biol* 144:290–300.
39. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
40. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
41. DL: S (2000) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4 (Sinauer Associates, Sunderland, MA).
42. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
43. Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.
44. JAA: N (2004) MrModeltest v2.2. Program distributed by the author.
45. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
46. Hao J, Balagurumoorthy P, Sarilla S, Sundaramoorthy M (2005) Cloning, expression, and characterization of sialic acid synthases. *Biochem Biophys Res Commun* 338: 1507–1514.