J Mol Evol (2007) xx:1–13
DOI: 10.1007/s00239-005-0256-3

JOURNAL OF MOLECULAR EVOLUTION

Evolution of Hyperactive, Repetitive Antifreeze Proteins in Beetles

Laurie A. Graham,[1] Wensheng Qin,[2] Stephen C. Lougheed,[2] Peter L. Davies,[1,2,3] Virginia K. Walker[2,3]

[1] Department of Biochemistry, Queen's University, Kingston, Ontario K7L 3N6, Canada
[2] Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada
[3] Protein Function Discovery Group, Queen's University, Kingston, Ontario K7L 3N6, Canada

Received: 1 November 2005 / Accepted: 22 January 2007 [*Reviewing Editor:* Dr. John Oakeshott]

Abstract. Some organisms that experience subzero temperatures, such as insects, fish, bacteria, and plants, synthesize antifreeze proteins (AFPs) that adsorb to surfaces of nascent ice crystals and inhibit their growth. Although some AFPs are globular and nonrepetitive, the majority are repetitive in both sequence and structure. In addition, they are frequently encoded by tandemly arrayed, multigene families. AFP isoforms from the mealworm beetle, Tenebrio molitor, are extremely potent and inhibit ice growth at temperatures below −5°C. They contain a 12-amino acid repeat with the sequence TCTxSxxCxxAx, each of which makes up one coil of the β-helix structure. TxT motifs are arrayed to form the ice-binding surface in all three known insect AFPs: the homologous AFPs from the two beetles, *T. molitor and Dendroides canadensis*, and the nonhomologous AFP from the spruce budworm, *Choristoneura fumiferana*. In this study, we have obtained the cDNA and genomic sequences of additional *T. molitor* isoforms. They show variation in the number of repeats (from 6 to 10) which can largely be explained by recombination at various TCT motifs. In addition, phylogenetic comparison of the AFPs from the two beetles suggests that gene loss and amplification may have occurred after the divergence of these species. In contrast to a previous study suggesting that *T. molitor* genes have undergone positive Darwinian selection (selection for heterogeneity), we propose that the higher than expected ratio of nonsynonymous-to-synonymous substitutions might result from selection for higher AT content in the third codon position.

Key words: Multigene families — Darwinian selection — Codon bias — Nonsynonymous changes — *Tenebrio molitor*

Correspondence to: Virginia K. Walker; *email:* walkervk@post.queensu.ca

Introduction

Antifreeze proteins (AFPs) are a diverse group of proteins that inhibit ice growth by an adsorption-inhibition mechanism (Raymond and DeVries 1977). Adsorption to the ice surface causes the freezing point of the AFP solution to be lowered in a noncolligative manner. To quantify antifreeze activity, a microscopic seed crystal is slowly cooled in the presence of AFPs. The difference in degrees centigrade between the melting point and the nonequilibrium freezing point can be determined and is termed thermal hysteresis (TH).

AFPs and their genes have been isolated from a variety of organisms and reveal a sequence and structural variety that is surprising (Jia and Davies 2002). Some of the most potent AFPs are found in insects like the yellow mealworm beetle, *Tenebrio molitor*, which is a pest of stored grain products. Overwintering larvae can supercool to an average of –12°C (Johnston and Lee 1990), aided by the TH activity of their hemolymph and other body fluids (Ramsey 1964). *T. molitor* AFPs are composed primarily of a variable number of copies of the 12-amino

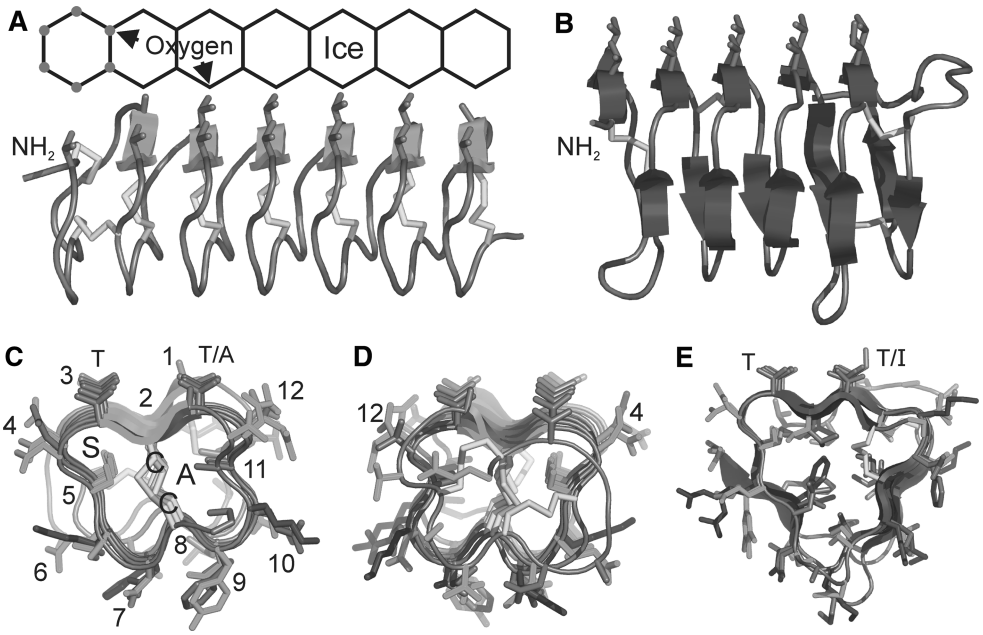

Fig. 1. Models of the crystal structures of nonhomologous β-helical insect AFPs. **A** Side view with N terminus on the left of *T. molitor* AFP (PDB 1EZG) showing the matching of the hydroxyl groups in the TCT ice-binding motif to the ice lattice (Liou et al. 2000). **B, C** End-on views of the right-handed, β-helical *T. molitor* AFP from the C terminus and N terminus, respectively. The six residues conserved within the 12-amino acid repeats are indicated by their one-letter code, and numbers show the position of each residue in the 12-amino acid repeat, TCTxSxxCxxAx. A disulfide bond stabilizes each loop and the only other internal residues are S and A. **D, E** Similar views of *C. fumiferana* AFP (isoform 337; PDB 1L0S), with the N terminus to the right and the front, respectively. It displays a similar array of ice-binding T residues but has a slightly longer repeat (15–17 residues), forms a left-handed β-helix, and has a triangular rather than rectangular cross section. It possesses a hydrophobic core and the disulfide bonds lie between, rather than within, the loops. The side chains are colored as follows: red, acidic; blue, basic; green, polar; yellow, C; gray, nonpolar; and gray with red oxygen, ice-binding T.

acid repeat TCTxSxxCxxAx (Graham et al. 1997; Liou et al. 1999). AFPs from the fire-colored beetle, Dendroides canadensis (Duman et al. 1998; Andorfer and Duman 2000), are clearly homologous. AFP sequences are also known in the spruce budworm moth, Choristoneura fumiferana, and the encoded AFPs contain repeats 15 to 17 amino acids in length (Tyshenko et al. 1997; Doucet et al. 2000, 2002). All of these repetitive insect AFPs show over 5°C of TH activity at millimolar concentrations and are >100-fold more potent than most AFPs from other organisms at comparable concentrations (Graham et al. 1997; Tyshenko et al. 1997; Li et al. 1998). Both the beetle (Liou et al. 2000) and the moth (Leinala et al. 2002a) AFPs are β-helical and contain parallel rows of Thr residues that form the ice-binding surface (Marshall et al. 2002; Graether et al. 2000). Despite these similarities, they are nonhomologous since the primary sequences are distinct, their helices coil in opposite directions, and their disulfide-bonding patterns are different (Fig. 1) (reviewed by Walker et al. 2001; Graether and Sykes 2004). Both AFPs are more potent when the number of repeats (and thus the ice-binding surface) is increased (Leinala et al. 2002b; Marshall et al. 2004).

Most AFPs are encoded by multigene families (Scott et al. 1985, 1988; Hayes et al. 1989), and the insect AFPs are no exception. Together, protein fractionation, Southern analysis, and cDNA cloning indicated that there were 30–50 copies of the AFP gene in *T. molitor* (Liou et al. 1999). Analysis of cDNAs for the multiple isoforms in all three insect species has provided insight into AFP structure and function (Doucet et al. 2000; Liou et al. 1999; Andorfer and Duman 2000). In addition, examination of base changes in cDNAs encoding *T. molitor* AFP suggested that certain residues were subject to positive Darwinian selection, as the rate of nonsynonymous nucleotide substitution was higher than the rate of synonymous substitution (Swanson and Aquadro 2002). We have now obtained the sequences of a large number of additional *T. molitor* AFPs. This new sequence information has allowed us to compare isoforms to each other and to those from *D. Canadensis* to elucidate the evolutionary history of these genes and their families. It has also provided an alternative explanation to Swanson and Aquadro's (2002) hypothesis of positive selection.

Materials and Methods

Cloning of cDNA and genomic sequences. AT molitor fat body cDNA library was screened for additional cDNA sequences (Tf, Th, Tk, and Taa) as previously described (Liou et al.

1999). Genomic clones (Tl, Tv, and Ty) were obtained by screening λZap II (Stratagene, La Jolla, CA, USA) and pBeloBac11 (Bio S&T Inc., Montreal, Quebec, Canada) libraries with labeled cDNA encoding isoform Tb. Additional sequences were obtained from genomic DNA or genomic libraries by PCR amplification using the following primer pairs (upstream primer [US] shown before downstream primer [DS]): US1 5′-CCGCTCGAGATGGCATTC AAIACITGTICTTTTA-3′ with DS1 5′-TGGATAGAATACA GGTTAAGCTTAC-3′ (Tm, Tn, To, Ts, Tt Tu, Tw, Tx, and Tz), US1 with DS2 5′-TGGATATAATACAGGT TAAGCTTAC (Tp), 5′-GAATTCGAATTCCAATGCACTGGGGGTGCTGA-3′ with 5′-TTTCTCGAGGTTTATTATTGTCCGRTAACC-3′ (Tj), 5′-CCGGAATTCGGCAGTCGATCGAAGTAAAC with DS1 (Tq), and 5′-CCGGAATTCCCCAAAATGAAACCGGAAA with DS2 (Tr). PCR products were subcloned into pBluescript SK (Stratagene) using standard techniques and sequenced.

Sequence alignments. Initial alignments were done using ClustalX (Thompson et al. 1997) but extensive manual adjustments were necessary due to the repetitive nature of the sequences. Only sequences showing > 5% divergence were included, as these were considered most likely to represent different genes rather than polymorphic alleles.

Phylogenetic analysis. The DNA sequence alignment was modified prior to phylogenetic analysis to exclude signal-peptide encoding sequence, regions in which the majority of the sequences were gapped (all regions appended with "like"), and sequence extending past the stop codon most commonly used in *T. molitor*. The alignment of the translated sequences and a second DNA alignment in which repeats 3 and 4 were deleted were also analyzed. In addition, an alignment of the internal repeats (excluding bases encoding any 13th residues) from selected divergent isoforms was generated. We used both maximum likelihood and maximum parsimony approaches to infer the evolutionary relationships among AFP isoforms. For the former, we selected the best model of evolution using MODELTEST, version 3.06 (Posada and Crandall 1998), according to the Akaike Information Criterion (AIC). The best-fit model was TVM + I + G (–lnlk = 2850.60) with the proportion of invariant sites = 0.219 and the ? shape parameter = 1.590. Support for the maximum likelihood tree was assessed using nonparametric bootstrap analysis with 100 pseudoreplicates (Felsenstein 1985). Parsimony analysis involved a heuristic search with TBR branch-swapping and 100 random addition replicates. Again, support for the resulting most parsimonious tree(s) was performed using bootstrap analysis, but with 10 random additions for each of 100 bootstrap pseudoreplicates. All phylogenetic analyses were done in PAUP*, version 4.0 (Swofford 2002).

Analysis of synonymous and nonsynonymous substitutions, codon usage, and AT bias

Low resolution of phylogenetic trees, especially for *T. molitor*, precluded using likelihood approaches to investigating rates of synonymous versus nonsynonymous evolution. Rather, we examined base changes using the SNAP tool (www.hiv.lanl.gov; Korber 2000) with subsets of data as follows. Both signal-peptide and mature coding sequences were considered separately, as were the sequences from the two different beetle species. Pairwise comparisons were done between each sequence and a consensus sequence derived from the most common nucleotide at each position in the alignment.

A comparison of the ratio between substitution types at every position within the 12-amino acid repeat was done as follows; the numbers of nonsynonymous (n_d) and synonymous (s_d) base changes represented were tallied (irrespective of how many sequences contained each change) and divided by the total changes possible (n and s, respectively) for the consensus codon. These values were averaged for the corresponding position of all repeats, excluding the first repeat and codons after position 8 within the last repeat. The ratio of n_d/n and s_d/s was then calculated.

All coding sequences from the nonredundant nucleotide database of GenBank were retrieved for *T. molitor*, *D. Canadensis*, and *C. fumiferana*. For non-AFPs, only those with < 50% sequence identity to other proteins from the same species were selected for analysis, with only the more divergent AFP isoforms used for comparison. The percentage GC content for each coding region was calculated at each position within the codon, but to reduce bias, those codons for amino acids with an odd number of codons (Met, Trp, and Ile), as well as the stop codon, were excluded from the analysis. For *T. molitor*, AFP isoforms Ta, Ti, Tr, Tt, and Tz were used, as well as the non-AFP sequences with accession numbers AB020738, AB021699, AB028848, AB037697, AB084067, AB108841, AB205184, AF312017, AF395329, AF448479, AJ000043, AJ000044, AJ005685, AJ005765, AJ251542, AJ400904, AJ487081, AJ496728, AJ518833, AJ555539, AY153772, AY153778, AY207373, AY303372, AY325895, AY327800, AY332269, AY332272, AY337517, AY513074, AY714212, DQ097802, D11338, D17670, M95697, M97262, M97917, U21482, U39658, U41298, X72783, X99204, Y09124, and Y11533. *C. fumiferana* AFPs isoforms were AF263009, AY004173, and AY004228, and *C. fumiferana* non-AFPs were AF007767, AF007768, AF016368, AF062383, AF092030, AF128867, AF153367, AF177644, AY098731, AY426538, AY426539, AY426540, AY460342, AY559246, AY563106, DQ005717, U12917, U37528, and U63930. *D. Canadensis* AFPs were Da, De, and Dj (U79777, AF179412, and AF179416), with the single type of non-AFP sequence available (DQ023319).

The GC content was determined for sequences upstream of the start codon and downstream of the stop codon for *T. molitor* genes for which over 1 kb of flanking sequence was available (AJ518833, M95697, U30627, U39658, U57832, X72783, and Y09124) and for the two longest AFP genes with dissimilar flanking regions (DQ224365 and DQ224367). The 3' flanking region of DQ224365 was excluded since it contained a potential exon for an unknown protein.

Results

Sequence alignments and similarities

Gene or cDNA sequences that encode 35 unique *T. molitor* AFPs have been cloned. Isoforms that differed only in the signal peptide encoding region or at one or a few residues within the mature sequence (> 95% amino acid sequence identity) are not considered, leaving 27 isoforms (Fig. 2). Using the same criteria, only 11 of 14 unique AFPs, encoded by a set of 19 cDNA sequences from *D. Canadensis*, are shown. Since the polymorphisms found in the excluded isoforms were usually present in one or more of the other isoforms, only five unique base changes, all of which were silent, are not represented.

These protein and DNA (Supplementary Fig. S1) multiple sequence alignments, generated by ClustalX (Thompson et al. 1997), required manual adjustment since the algorithm often aligned sequences out of register by one or more repeats. Internal repeats showed limited variation at the amino acid level since 6 of the 12 residues are well conserved and only a

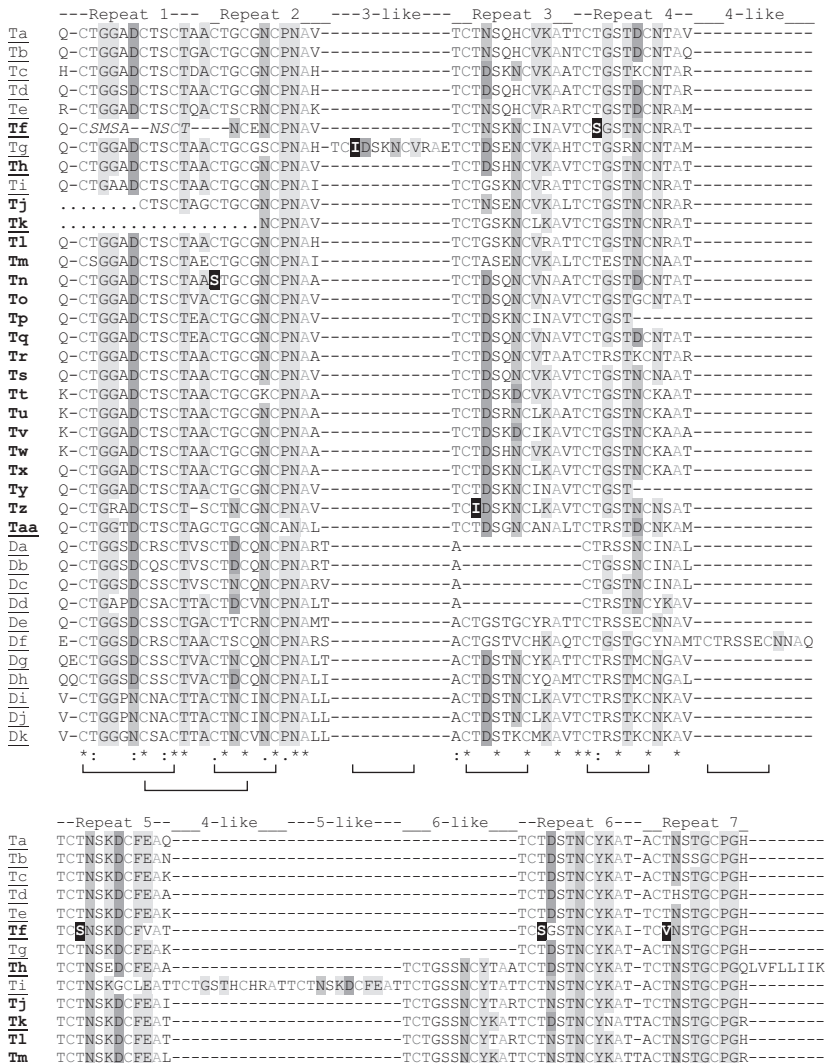

Fig. 2. Alignment of homologous beetle AFPs from *T. molitor* (Ta–Taa) and *D. Canadensis* (Da–Dk). The well-conserved residues within the core repeat motif, TCTxSxxCxxAx$_{1-2}$, are colored as follows: Thr, blue; Cys, red; Ser, brown; and Ala, green. The two Cys within each repeat form a disulfide bond with the exception of repeats 1 and 2, which are linked by an additional disulfide bond (disulfide bonds are shown by brackets below the sequences). The terminal repeat (no. 7) does not contain the conserved Ala. Yellow highlighting indicates residues characteristic of a particular repeat, and Asp and Asn at the fourth and seventh positions in the repeat are highlighted light green and blue, respectively. Exceptional residues that do not correspond to the motif are boxed in black. Residues shown in italics result from frameshifts/deletions. Positions at which all sequences (gaps/missing sequences ignored) contain identical, highly similar, and similar residues are denoted by an asterisk, colon, and period, respectively (Thompson et al. 1997). The names of sequences not published previously are in boldface and are underlined if derived from cDNA rather than genomic DNA. Secretory signal peptides are not shown. The accession numbers of the DNA sequences encoding AFP isoforms are as follows: Ta, AF159114; Tb, AF160494; Tc, AF159116; Td, AF159115; Te, AF159117; Tf, DQ229123; Tg, AF159118; Th, DQ229124; Ti, AF160497; Tj, DQ229112; Tk, DQ229130; Tl, DQ229125; Tm, DQ229113; Tn, DQ229114; To, DQ229115; Tp, DQ229116; Tq, DQ229126; Tr, DQ229127; Ts, DQ229117; Tt, DQ229118; Tu, DQ229119; Tv, DQ229128; Tw, DQ229120; Tx, DQ229121; Ty, DQ229129; Tz, DQ229122; Taa, DQ229131; Da, U79777; Db, AF046862; Dc, AF179410; Dd, AF179408; De, AF179412; Df, AF179415; Dg, AF179414; Dh, AF179413; Di, AF179409; Dj, AF179416; Dk, AF179411.

limited number of substitutions are common at certain positions. However, the DNA sequence alignment (Supplementary Fig. S1) was more informative because amino acid substitutions resulting from only a single base change could be discriminated from those resulting from two or three base changes. In addition, the third codon position was sometimes informative. For example, the infrequently used codon, ACG, encodes the first Thr of the third repeat in all but one of the *T. molitor* sequences.

The two alignments agreed quite well, but four changes were made to the protein alignment (Fig. 2)

based on the DNA alignment: (i) the gap in the first repeat of Tf was adjusted, (ii) the gap in repeats 4–5 for Tp and Ty was inserted in the middle of the repeat, (iii) most of the third repeat of *D. Canadensis* sequences Da–Dd was aligned with repeat 4 of the other sequences rather than repeat 3, and (iv) the central region of the longest *D. Canadensis* isoform (Df) was shifted by one repeat unit. This final alignment was supported by phylogenetic analysis of the repeat units (see below).

The similarities between both the protein (Fig. 2) and the DNA (Supplementary Fig. S1) sequences of the two beetle species indicated that the AFPs of these two beetles are homologous. Pairwise amino acid identities (with pairwise gap exclusion) were as follows: between species, from 55% to 88%, with an average of 67%; within *T. molitor*, a minimum of 60%, with an average of 82%; and within *D. Canadensis*, a minimum of 56%, with an average of 74%.

Core repeats. The different isoforms are variable not only at individual residues, but also in the number of repeats that they contain. As indicated above, the "core" repeat contains the sequence TCTxSxxCxxAx. The first repeat (14 or 15 amino acids) is conserved at only three of the six "core" residues in the repeat motif (i.e., tCTxsxxCxxax, with conserved residues capitalized; Fig. 2). Repeat 2 is also unique in that the first residue is not Thr, but in 29 of 37 sequences, Ala substitutes. The Ser residue, which stabilizes a turn in later repeats (Fig. 1), is often replaced by Gly in the first repeat. In repeat 2, the turn is stabilized by a Cys, which forms a disulfide bond with the second Cys in repeat 1. Two base changes differentiate the Gly codon (GGT) from the Ser codon (TCT) prevalent in repeats 3 and 4, whereas a single change differentiates the Cys codon (TGT). Repeats 3 to 6 are similar and adhere to the core repeat motif with Asp, Asn, and Gly common at positions 4 and 7 (Figs. 1B and 2). In both *T. molitor* and *D. Canadensis*, a 13th residue is sometimes found at the end/beginning of these repeats. Overall, larger and more hydrophobic residues are not abundant since the protein does not possess a hydrophobic core. For example, Trp is not found, and Ile, Leu, Tyr, Phe, and Met make up only 5% of the residues. When present, they are primarily in the ninth position of repeats 3, 5, and 6, well away from the ice-binding surface (residues 1 and 3; Fig. 1). The 12th residue, adjacent to the ice-binding surface, is very variable, likely because of the curvature of the protein backbone at these sites (Liou et al. 2000), which would place side chains away from the adjacent, flat, ice-binding region (Figs. 1B and C).

The final repeat resembles the core repeats except that the first residue is most often Ala, and a Pro residue immediately follows the second Cys residue in most sequences. Different isoforms show length variability at the carboxyl terminus as a result of frameshift or missense mutations that alter the stop codon. Most (18/27) *T. molitor* isoforms terminate with Pro Gly His (with the most common variant being Pro Gly Arg), whereas the terminus of *D. Canadensis* is of more variable length and sequence (Fig. 2).

The ice-binding surface of all isoforms is dominated by a double row of Thr residues (TCT motif; Fig. 1), which are essential for activity (Marshall et al. 2002). This motif is altered to xCT in the first repeat and the Gln commonly found at the N terminus (x) is converted in vivo to pyroglutamate. The position of the backbone at this position (Fig. 1C) suggests that this residue is unlikely to be involved in ice binding, whereas the Thr likely is. In repeat 2, the TCT motif is altered to (A/S)CT. Since the backbone bulges out from the plane of the other repeats at this point (Figs. 1A and C) (Liou et al. 2000), these smaller residues may be required to maintain the flatness of the surface. Similarly, ACT is often found in the last repeat, but TCT is common as well, presumably since the backbone deviation is less than in repeat 2 (Fig. 1). ACT is also found in the first row of repeat 3 of the *D. Canadensis* isoforms. Other, infrequent substitutions of the ice-binding Thr resides also occur, including Ser, which merely removes a methyl group or Val, which substitutes the hydroxyl for a methyl.

Phylogenetic analysis of isoforms. Both maximum likelihood and maximum parsimony analysis of the DNA alignment were performed. The model of evolution used for the maximum likelihood analysis (TVM + I + G) better accommodates rate heterogeneity and complex evolutionary patterns that would be expected of AFP genes. Nevertheless, both trees were essentially identical and thus only the maximum likelihood tree is shown (Fig. 3). Despite their obvious homology, a well-supported split between two clades was apparent, each exclusively containing the AFP isoforms for one of the two investigated species. An additional striking feature of this tree is the markedly greater DNA sequence divergence among *D. Canadensis* versus *T. molitor* isoforms (mean p-distances, 0.22 ± 0.071 and 0.094 ± 0.029, respectively). The two clades and similar groupings were still evident following maximum parsimony analysis of the protein alignment (data not shown), but the divergence within each species was more comparable. This is largely a result of the greater variability in the third codon position. For example, in both species, between 35% and 39% of the first and second codon positions are variable in one or more sequences. For the third position, this increases to 58% for *T. molitor*, whereas *D. Canadensis* shows a greater increase, to 74%. Bootstrap analysis indicates that the topology within the *D. Canadensis*

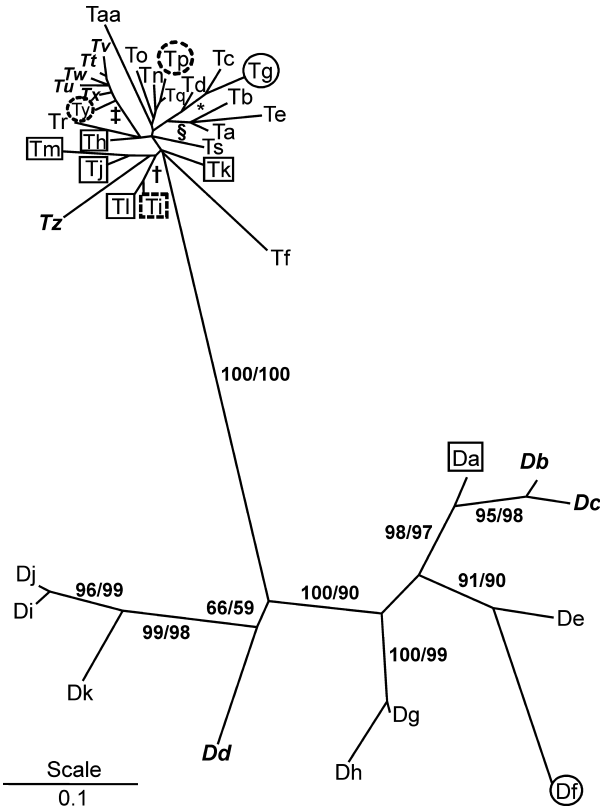

Fig. 3. Maximum likelihood phylogram showing the relationship among isoforms based on analysis of a DNA alignment containing only those repeats common to most isoforms (Supplementary Fig. S1). Bootstrap support from maximum likelihood and maximum parsimony analyses are shown before and after the forward slash, respectively. Only bootstrap values ≥50% are indicated. Space limitations precluded mapping bootstrap values for *T. molitor* directly onto the tree. We have instead used symbols as follows: *, 59/59; †, 75/84; ‡, 92/95; §, 79/64. Boxes and circles, both dashed and solid, as well as boldface italic and regular font, are used to denote isoforms with similar patterns and numbers of repeats as shown in Fig. 4A.

clade is better resolved than within the *T. molitor* clade, with most nodes for the former species supported at ≥90%. Several factors might contribute to this. The divergence is lower and the number of isoforms is higher for *T. molitor*. In addition, the generation of hybrid genes by recombination or gene conversion cannot be ruled out. Although all isoforms are clearly homologous, orthologous or paralogous relationships between the genes of the two species cannot be inferred from these data.

Phylogenetic analysis of repeats. The alignments suggest that the AFPs from both beetles are composed of seven basic and distinct repeat units. Different isoforms contain from 6 to 10 copies of these repeats in different combinations (Figs. 2 and 4A), with each repeat forming one turn of the β-helix (Figs. 1A–C). As mentioned previously, each of the repeats 1, 2, and 7 (first two and last one) have unique features and are therefore more divergent relative to other repeats. The internal repeats, 3–6, are more similar and show high conservation of the Thr within the TCT motif. These internal repeats seem to be largely interchangeable, with each one either duplicated or deleted in one or more isoforms. Such flexibility is consistent with their structural and sequence similarity (Figs. 1 and 2). Thus, there is variability in the length of the AFPs, and also, it was relatively easy to add or delete internal repeats in a protein engineering study (Marshall et al. 2004). Although both species possess shorter isoforms with only 6 repeats, the most common repeat length is 7 and 8, for *T. molitor* and *D. Canadensis*, respectively. All gains or losses of repeats appear to have occurred within the TCT (or ACT) motif, except for isoforms Tp and Ty (Fig. 2 and Supplementary Fig. S1), in which the DNA alignment supports a rearrangement in the middle of a repeat. To further examine the relationship between the repeats, the DNA sequences encoding the first 12 amino acids of each repeat from selected divergent isoforms, including the two longest isoforms (Ti and Df), were aligned and subjected to maximum likelihood phylogenetic analysis (Fig. 4B) as before, using the GTR + I + G model of evolution (I = 0.249, shape parameter = 6.59 [Posada and Crandall 1998]). Repeats 1 and 7 were excluded to simplify the interpretation, as they are obviously unique relative to the others. The remaining five repeats of *T. molitor* isoforms Ta and Taa are distributed across the phylogram in five groups. Although isoform Ti has eight internal repeats, they all fall into one of these groups. This suggests that this long isoform arose by insertion of additional repeats from a seven-repeat precursor. Repeats 2, 5, and 6 of the *D. Canadensis* isoforms cluster with like-named repeats of the *T. molitor* isoforms. However, clustering was less evident with repeats 3 and 4. Since this section of the alignment was considered less reliable, it was subsequently excluded and the phylogenetic analysis performed again, but the tree remained identical in all important details (data not shown). Again, all of the repeats found in the longer isoform (Df) cluster with repeats from the shorter isoforms.

Isoforms having similar patterns of repeats (Fig. 4a) were mapped onto the phylogenetic tree (Fig. 3). For *T. molitor*, the poorer resolution of the tree complicated the analysis, but there is some evidence of clustering of types, as five of the six isoforms lacking repeat 5 (Tt, Tu, Tv, Tw, and Tx) form one clade. For *D. Canadensis*, three of the four isoforms missing repeat 3 (Da, Db, and Dc) also cluster.

The longest isoform from each species (10 repeats) contains duplicates of repeats 4, 5, and 6, but in a slightly different pattern: Ti = 1, 2, 3, 4, 5, 4, 5, 6, 6, 7; and Df = 1, 2, 3, 4, 4, 5, 5, 6, 6, 7. Neither pattern can be generated by a single event from a seven-repeat progenitor. Pairwise comparisons between the

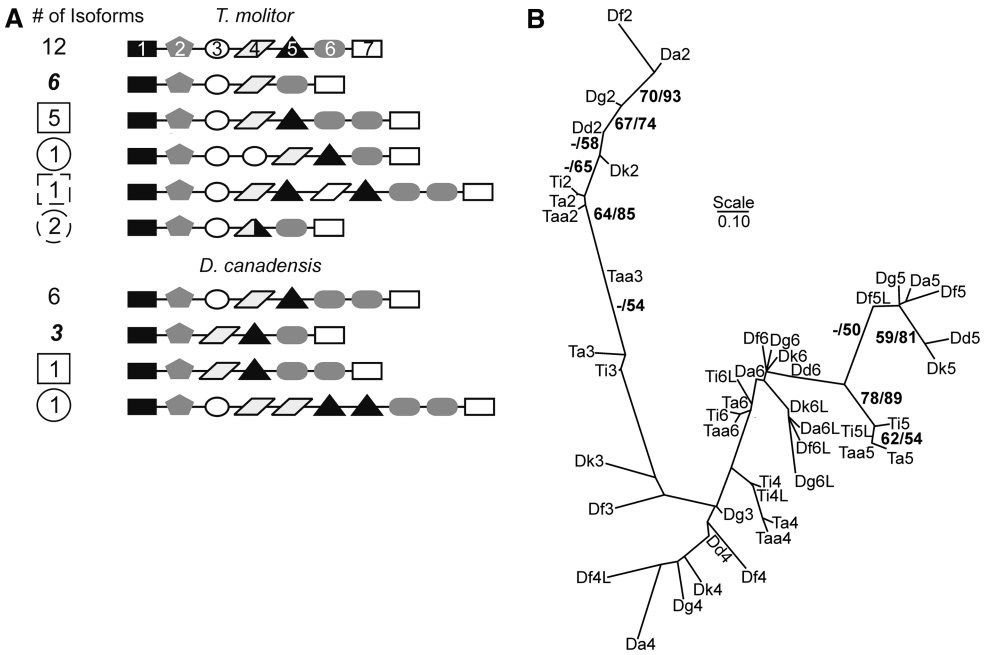

Fig. 4. Analysis of repeats found in beetle AFPs. **A** Schematic showing the pattern of repeats found in all AFP isoforms. Each repeat is given a unique shape and shading pattern. All duplications and deletions occurred within the TCT motif with one exception, indicated by the hybrid symbol (see Tp and Ty; Fig. 2). The sequences were grouped by category as follows: for *T. molitor*, 12 isoforms (Ta–Tf, Tn, To, Tq–Ts, Taa), 6 isoforms (Tt–Tx, Tz), 5 isoforms (Th, Tj–Tm), 1 isoform (Tg), 1 isoform (Ti), and 2 isoforms (Tp, Ty); and for *D. Canadensis*, 6 isoforms (De, Dg–Dk), 3 isoforms (Db–Dd), 1 isoform (Da), and 1 isoform (Df). The numbers beside each category are modified in the same manner as the isoform names in Fig. 3. **B** Maximum likelihood radial phylogram based on the DNA sequence of individual internal repeats from selected isoforms. They are named according to isoform, followed by a number indicating repeat number as in Fig. 2, with L representing "like" and bootstrap values indicated as in Fig. 3.

internal repeats of the *T. molitor* isoform Ti (Supplementary Fig. S2) show that there are two base pair differences between repeat 4 and repeat 4-like and between repeat 5 and repeat 5-like and five between repeat 6 and repeat 6-like. However, 11–14 differences are seen when different numbered repeats are compared. Given the high similarity of isoform Ti to Tl, in which only repeat 6 is duplicated (Figs. 2 and 4A) and the greater divergence of repeat 6-like, repeat 6 was probably duplicated first. Repeats 4 and 5 may have been duplicated later. Whether this was an intra- or an intergenic duplication is difficult to determine due to the high similarity between these repeats in all isoforms. A similar scenario in which repeat 6 was duplicated first is likely for *D. Canadensis* isoform Df. However, in this case, pairwise alignments (Supplementary Fig. S3) suggest that repeats 4- and 5-like may have arisen by an intergenic mechanism. This is because repeats 4- and 5-like are more similar to repeats 4 and 5 of isoform De (four and one differences, respectively) than to the corresponding repeats in the same gene (nine and six differences, respectively).

Codon bias, GC content, and synonymous and nonsynonymous substitutions. Examination of several of the more divergent TmAFP isoform sequences revealed that although the GC content in the first two codon positions averaged 47%–51%, it was only 26%–32% at the third position (Fig. 5), resulting in the low overall GC content of 42%–44% for these genes. In contrast, the GC content in the first two codon positions for 44 non-AFP sequences ranged from 41% to 75% (mean, 52%), while the third position was even more variable, ranging from 28% to 85% (mean, 56%). TmAFPs are synthesized primarily in larval fat body (Graham et al. 2000) and the GC content at the third position for other non-AFPs synthesized by the same tissue and during the same developmental stage (AB205184, AY153772, AY153778, and U41298) averaged 59% (Fig. 5). A comparison of the codon usage for the seven residues most abundant in the AFPs revealed a consistent bias toward a low GC content at the third position relative to the non-AFPs (Supplementary Table S1). *D. Canadensis* AFP sequences appear to show the same bias as those from *T. molitor*. AFP sequences from *C. fumiferana* showed the slightly lower third position GC content of 47%–52% than the 19 *C. fumiferana* non-AFPs (52%–79%). Examination of the rate of synonymous and nonsynonymous substitutions in the coding sequences of the two species is complicated by a lack of independence of the sequences. Therefore, instead of pairwise comparisons between all sequences, we compared each sequence to a consensus sequence derived from the alignment. Bases that could not be

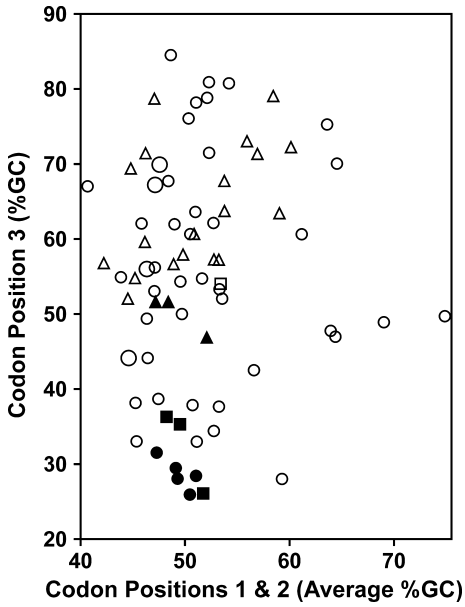

Fig. 5. Codon bias in AFP coding sequences. The GC content at the third codon position is compared to the average GC content at the other two positions. The sequences of *T. molitor* are represented by circles; *C. fumiferana*, by triangles; and *D. Canadensis*, by squares. AFPs are represented by filled symbols, and non-AFPs by open symbols. Non-AFPs known to be expressed in the larval fat body are represented by larger symbols.

unequivocally aligned and C-terminal sequences with frameshifts were excluded from the analysis (Supplementary Fig. S4). In *T. molitor*, the ratio of nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site (d_N/d_S) in the mature sequence averages 1.09 (range, 0.25–2.86). The small number of base differences, from 5 to 22 in pairwise comparisons, is a likely source of some of this variation. In contrast, the signal-peptide encoding sequence averaged 0.45 (range, 0.07–1.43). The results for *D. Canadensis* were more consistent with d_N/d_S averaging 0.34 (range, 0.20–0.54) for the mature sequence, with 22–48 base differences. The signal-peptide encoding region averaged 0.45 (range, 0–2.56), similar to that observed for *T. molitor*. These calculations suggest that the *D. Canadensis* genes are primarily under negative selection pressure, but it is less clear for *T. molitor*.

To determine whether nonsynonymous mutations were overrepresented at any location within the 12-amino acid repeat, we performed a comparison largely based on the method of Nei and Gojobori (1986) but with each base change scored once regardless of the number of sequences in which it was found. This simple approach was chosen for several reasons. First, as noted above, the *T. molitor* clade was poorly resolved and it was difficult to deduce whether a particular mutation occurred once or on several occasions. Second, the methods used to screen for new clones may have resulted in an overrepresentation of groups of similar clones. Third, 20 of the 84 potentially degenerate codons within the mature sequence are invariant across all 27 *T. molitor* sequences and only 4 of 252 nucleotide positions contain all four bases. This suggests that most changes are a result of a single mutation rather than multiple mutations, at least in *T. molitor*.

Table 1. Differences in the ratios of nonsynonymous-to-synonymous nucleotide changes ($(n_d/n)/(s_d/s)$) at each position in the 12-amino acid repeat

Residue	*T. molitor*	*D. Canadensis*
1 T	0.27	0.36
2 C	0.04	0.00
3 T	1.50 (0.25)[a]	0.21
4 X	0.88	0.82
5 S	0.00	0.00
6 X	0.77	0.55
7 X	0.33	0.34
8 C	0.00	0.00
9 X	0.66	0.43
10 X	1.06	0.28
11 A	0.00	0.00
12 X	2.23	1.58

[a]If a single isoform (Tf) is omitted from the analysis.

By this method, the three underlined positions within the *T. molitor* repeat consensus (TC*T*xSxxCx*x*A*x*) showed an excess of nonsynonymous changes ($(n_d/n)/(s_d/s) > 1$; Table 1). The unusual isoform Tf contains four substitutions at the second Thr residue, and if ignored, the ratio drops to 0.25. The ratio at position 10 (X) is close to unity. Only position 12 shows a large excess of nonsynonymous changes with a ratio of 2.2, similar to a previous observation using different methods and a smaller dataset (Swanson and Aquadro 2002). This suggests that this site, adjacent to the ice-binding surface (Figs. 1B and C), could be under positive selection. Certainly it is more variable than any of the other positions. It should be noted that A or T is almost exclusively used in at the third codon position of residue 12 in four of the five internal repeats, similar to that found for the entire gene.

When a similar analysis was carried out on the repeat consensus of *D. Canadensis*, a similar pattern was observed. Generally, the ratios were lower due to the greater variability found in the third codon position. The 12th position again showed an excess of nonsynonymous changes, with a ratio of 1.58.

The GC content of the sequences flanking seven non-AFP and two AFP genes in *T. molitor* was consistently low, averaging 35% (range, 31%–37%; Supplementary Table S2). However, for these genes, the GC content of the third codon position ranged from 38% to 70% for the non-AFPs, compared to 26% for the AFPs.

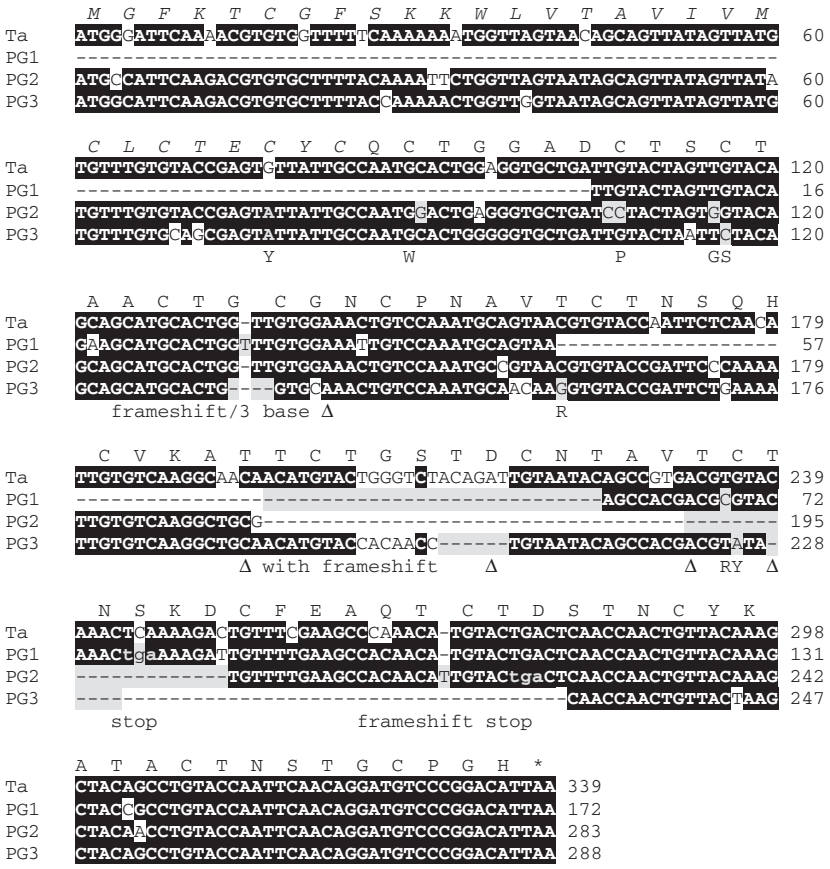

Fig. 6. Alignment of AFP pseudogenes (PG1, -2, and -3) with the *T. molitor* isoform Ta coding sequence. Bases shared by two or more sequences are shaded black. The protein sequence of Ta, including the signal sequence in italics, is indicated above the sequence, and all alterations that would be expected to disrupt either the reading frame or the function of the protein are highlighted in yellow. Deletions of a length inconsistent with the loss of a 12- to 13-amino acid repeat are noted and indicated below by triangles, and deleterious missense and nonsense mutations are also indicated below the sequence.

Pseudogenes. Of the large number of cDNA and genomic sequences obtained, most encoded complete AFPs. However, three genomic sequences appeared to be pseudogenes, with each sequence containing at least four aberrations (Fig. 6). These include highly conserved Cys or Thr substitutions, deletions of segments of length other than 36–39 bases (which would encode a single repeat), frameshifts, and mutations to a stop codon. Most of these alterations are in unique positions (Fig. 6), suggesting independent mutational events. The d_N/d_S ratio observed for these three pseudogenes did not differ from those for the other AFP genes (range, 0.29–2.86).

Discussion

Gene families. Insect AFPs adsorb to the repetitive crystalline surface of ice, so it is reasonable that they consist of repeats. In the beetles, these are encoded by 36- to 39-bp sequences that are reiterated, with modification, 6–10 times in each transcript. In turn, the genes are also repeated within the beetle genomes. In *T. molitor* there are 30–50 copies, including some alleles and a few pseudogenes. Southern analysis using *T. molitor* DNA from individuals indicated that these genes were linked (Liou et al. 1999), but perhaps not closely, since only a single copy of TmAFP was found in a 5.7-kb BAC library clone (DQ224368). Phylogenetic trees show that isoforms within each beetle species are more similar to one another than to the isoforms present in the other species (Fig. 3), which might suggest that a common ancestor contained a single AFP gene. However, we suspect that this is unlikely given that most AFPs are members of multiple-gene families. Changes in gene copy number are likely to occur in response to climate change, supported by the observation that in the fish, Macrozoarces americanus, individuals from somewhat warmer waters contained about a quarter of the 150 AFP genes found in those from higher latitudes (Hew et al. 1988). Similarly, numerous insect species have developed resistance to insecticides, often after only a few years of exposure, due to gene amplification (reviewed by Hemingway et al. 2004). Thus, multiple genes may have been present prior to divergence of the two beetle species, but concerted evolution (gene conversion and unequal crossover) and/or birth and death evolution (multiple rounds of gene loss and gene duplication) (Nei and Rooney 2005) may have led to higher within-species similarities. The clear separation between the genes of the two species is consistent with the former, whereas the presence of pseudogenes and variations in selective pressure are consistent with the latter. The greater similarity between isoforms within *T. molitor*, as well

as the poor resolution relative to the *D. Canadensis* clade, suggests that such events may have occurred more recently and more rapidly in *T. molitor* than in *D. Canadensis*. Regardless, these events have obscured the evolutionary history of this gene family, so rather than referring to them as orthologues or paralogues, we prefer to call them isofunctional homologues.

Variation in the length of isoforms. Analysis of natural and engineered TmAFP cDNAs expressed in bacteria shows that an increase in the number of repeats, up to a maximum of nine, results in a substantial increase in activity. Further increases actually reduce activity (Marshall et al. 2004). In *T. molitor*, the intermediate, seven- and eight-repeat isoforms are most abundant in the hemolymph (Liou et al. 1999), and it is curious that the more powerful ine-repeat isoform has not been found in either beetle species. A similar situation occurs with fish that express isoforms of different sizes, where, again, the shorter isoforms are less active. For example, the four-repeat isoforms of type I AFP in winter flounder represent a minor component in comparison to the three-repeat isoforms (Chao et al. 1996) and the longer isoforms (up to ~60 repeats) of the tripeptide-repeat antifreeze glycoproteins were at a much lower concentration than the shorter isoforms (as few as four repeats) in seven species tested (Jin and DeVries 2006). It is possible that the shorter isoforms are retained in the genome if the longer isoforms are more unstable and/or difficult to fold (as was found for the *T. molitor* AFP in vitro by Marshall et al. [2004]) or less soluble. An alternate hypothesis is offered by studies of large numbers of genes (Taylor et al. 2004) or pseudogenes (Zhang and Gerstein 2003) in which deletions were consistently found to outnumber insertions by up to threefold. Although the longest AFP isoforms we recovered from both beetle species appear to have arisen by the insertion of repeats, it is possible that a bias toward shorter isoforms is maintained by the constant generation of isoforms with deletions. The internal repeats appear to be largely interchangeable as indicated by their frequent duplication and deletion in various isoforms, and it is these repeats that give the most insight into the evolution of the beetle AFP genes. The majority (95%) of all repeat gains/losses seem to have occurred at the TCT ice-binding motif, and it is likely that unequal recombination or replicational slippage here gave rise to the variety of observed isoforms. A similar mechanism has been proposed for the expansion of fish AFGP (Logsdon and Doolittle 1997). The presence of an XCT or ACT motif instead of the TCT motif in repeats 1 and 2 may result in the reduction of the loss or gain of these two repeats in this manner. The similarity and short length of the repeats make it difficult to determine whether local events, such as replicational slippage, predominate. However, two repeats, found in the longest *D. Canadensis* isoform, are inconsistent with this mechanism since they appear to have originated from another isoform.

Conservation of residues within repeats. The conservation of residues within the 12-amino acid repeat consensus, TCTxSxxCxxAx, correlates well with their roles in the structure and function of the protein. The four amino acids that occupy the inside of the β-helix, and that are essential to maintain its structure (two Cys, Ala, and Ser; Fig. 1) are almost invariant. The two Cys form internal disulfide bonds that divide the interior of the β-helix into two compartments. One compartment contains the conserved Ala side chain plus a tightly bound water molecule (Liou et al. 2000). The other contains the Ser side chain. Both Ser and the water participate in an internal H-bonding network. The ice-binding Thr residues (Marshall et al. 2002) are also well conserved, particularly at the second position of each TCT motif. Small residues (Ala and Ser) sometimes substitute at the first position (Fig. 2). Because the first and second repeats are unique and are linked by an additional disulfide bond, they likely have an important role in the folding and/or stabilization of the protein. The C-terminal loop of the *C. fumiferana* AFP, which alters direction relative to the other loops (Fig. 1D) (Graether et al. 2000), is thought to have a similar role. The residues denoted x in the repeat consensus are variable, with charged or polar residues predominating as expected, since these residues are solvent exposed. Variability is highest at the 12th position, flanking the ice-binding site (Figs. 1 and 2). A comparison cannot be made with the corresponding residue of the *C. fumiferana* AFP since it is buried rather than solvent exposed (Fig. 1E). However, similar residues (Asp, Asn, and Gly) predominate at the other flanking residue of all three species (Fig. 2) (Doucet et al. 2002).

Codon bias. It is evident that the codon usage in AFP genes in both beetle species is markedly biased toward a low GC content at the third codon position (Fig. 5). This position is more GC-rich in the *C. fumiferana* AFPs but is still below that of other genes from this species. *T. molitor* genes that are expressed in the same tissue and developmental stage do not share this trait. Nor is it the result of the biased amino acid composition of the AFPs, since codons for the commonly used amino acids show the same trend (Supplementary Table S1). In another insect, the honey bee (*Apis mellifera*), GC content in the third codon position is positively correlated with the GC content of the surrounding genomic sequence (Jørgensen et al. 2006). This does not appear to be the case for *T. molitor* (Supple-

mentary Table S2), at least for the nine genes examined (Supplementary Table S2). Although global (rather than gene-specific) biases are not evident in the coding sequences of Archaea and bacteria that live at different temperatures (Das et al. 2006), the GC content of functional RNAs (tRNA and rRNA) is negatively correlated with decreasing temperature. Therefore, we postulate that the bias in the AFP sequences might facilitate transcription or translation at lower temperatures. An analysis of the ratio of synonymous and nonsynonymous changes indicates that only one of the residues within the repeat (residue 12; Fig. 1), adjacent to the ice-binding face, might be under positive selection. This position is certainly the most variable, even within repeats in the same position in the isoforms. However, if there is strong selection on the third codon position, mutations here cannot be considered truly silent. Therefore, although Swanson and Aquadro (2002) concluded that *T. molitor* AFPs are undergoing positive Darwinian evolution for isoform diversity at this position, another possibility is that the residue is neutral but that the codon is under selection for AT content. It is also possible that the low GC content at the third codon position on these genes, selected due to temperature or conformational considerations, has coincidentally also facilitated their evolution by replicational slippage.

Conclusion

To recapitulate, multiple AFP gene copies in beetles give flexibility in transcriptional regulation, with some isoforms synthesized in different tissues or expressed at different times during the winter (Andorfer and Duman 2000), as well as presumably contributing to the high level of TH and the consequent freeze-avoidance survival strategy. The genes themselves have been assembled from repetitive units, which have given rise to flexibility in TH activity, with longer forms having superior activity. Codon bias may have allowed genomic flexibility, perhaps enhancing rates of recombination and gene conversion or possibly providing an evolutionary advantage for expression at low temperatures. The resulting AFP family encodes isoforms that are remarkably well conserved in their stable, β-helix structure with a flat ice-binding motif along one face that forms a complementary surface with an ice crystal. What is even more astounding is that moths have evolved AFPs with similar ice-binding motifs on a β-helix scaffold and encoded by a family of internally repetitive isoforms within the last few million years. The beetle and the moth AFPs are fine examples of convergent evolution, and both appear to have been assembled from core repeats, reminiscent of the modules suggested to play an important role in the evolution of genes in the early eukaryotes (Gō 1981).

Acknowledgments. CHIR and NSERC (Canada) are acknowledged for grants to P.L.D and V.K.W., respectively. The authors thank Dr. D. Forsdyke for comments on the manuscript. The comments made by the reviewers were invaluable and helped us to dramatically improve our analysis and interpretation of the results.

References

Andorfer CA, Duman JG (2000) Isolation and characterization of cDNA clones encoding antifreeze proteins of the pyrochroid beetle *Dendroides Canadensis*. J Insect Physiol 46:365–372

Chao H, Hodges RS, Kay CM, Gauthier SY, Davies PL (1996) A natural variant of type I antifreeze protein with four ice-binding repeats is a particularly potent antifreeze. Protein Sci 5:1150–1156

Das S, Paul S, Bag SK, Dutta C (2006) Analysis of *Nanoarchaeum equitans* genome and proteome composition: indications for hyperthermophilic and parasitic adaptation. BMC Genomics 7:186

Doucet D, Tyshenko MG, Kuiper MJ, Graether SP, Sykes BD, Daugulis AJ, Davies PL, Walker VK (2000) Structure-function relationships in spruce budworm antifreeze protein revealed by isoform diversity. Eur J Biochem 267:6082–6088

Doucet D, Tyshenko MG, Davies PL, Walker VK (2002) A family of expressed antifreeze protein genes from the moth, *Choristoneura fumiferana*. Eur J Biochem 269:38–46

Duman JG, Li N, Verleye D, Goetz FW, Wu DW, Andorfer CA, Benjamin T, Parmelee DC (1998) Molecular characterization and sequencing of antifreeze proteins from larvae of the beetle *Dendroides canadensis*. J Comp Physiol [B] 168:225–232

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evol 39:783–791

Gō M (1981) Correlation of DNA exonic regions to protein structural units in haemoglobin. Nature 291:90–92

Graether SP, Sykes BD (2004) Cold survival in freeze-intolerant insects: the structure and function of beta-helical antifreeze proteins. Eur J Biochem 271:3285–3296

Graether SP, Kuiper MJ, Gagne SM, Walker VK, Jia Z, Sykes BD, Davies PL (2000) Beta-helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect. Nature 406:325–328

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

Graham LA, Walker VK, Davies PL (2000) Developmental and environmental regulation of antifreeze proteins in the mealworm beetle *Tenebrio molitor*. Eur J Biochem 267:6452–6458

Hayes PH, Scott GK, Ng NF, Hew CL, Davies PL (1989) Cysteine-rich type II antifreeze protein precursor is initiated from the third AUG codon of its mRNA. J Biol Chem 264:18761–18767

Hemingway J, Hawkes NJ, McCarroll L, Ranson H (2004) The molecular basis of insecticide resistance in mosquitoes. Insect Biochem Mol Biol 34:653–665

Hew CL, Wang NC, Joshi S, Fletcher GL, Scott GK, Hayes PH, Buettner B, Davies PL (1988) Multiple genes provide the basis for antifreeze protein diversity and dosage in the ocean pout, *Macrozoarces americanus*. J Biol Chem 263:12049–12055

Jia Z, Davies PL (2002) Antifreeze proteins: an unusual receptor-ligand interaction. Trends Biochem Sci 27:101–106

Jin Y, DeVries AL (2006) Antifreeze glycoprotein levels in Antarctic notothenioid fishes inhabiting different thermal environ-

ments and the effect of warm acclimation. Comp Biochem Physiol B Biochem Mol Biol 144:290–300

Johnston SL, Lee RE Jr (1990) Regulation of supercooling and nucleation in a freeze intolerant beetle (*Tenebrio molitor*). Cryobiology 27:562–568

Jørgensen FG, Schierup MH, Clark AG (2006) Heterogeneity in regional GC content and differential usage of codons and amino acids in GC-poor and GC-rich regions of the genome of *Apis mellifera*. Mol Biol Evol [Dec. 5; Epub ahead of print]

Korber B (2000) HIV signature and sequence variation analysis. In: Rodrigo AG, Learn GH (eds) Computational analysis of HIV molecular sequences. Kluwer Academic, Dordrecht, Netherlands, pp 55–72

Leinala EK, Davies PL, Jia Z (2002a) Crystal structure of beta-helical antifreeze protein points to a general ice binding model. Structure (Cambr) 10:619–627

Leinala EK, Davies PL, Doucet D, Tyshenko MG, Walker VK, Jia Z (2002b) A beta-helical antifreeze protein isoform with increased activity. Structural and functional insights. J Biol Chem 277:33349–33352

Li N, Andorfer CA, Duman JG (1998) Enhancement of insect antifreeze protein activity by solutes of low molecular mass. J Exp Biol 201:2243–2251

Liou YC, Thibault P, Walker VK, Davies PL, Graham LA (1999) A complex family of highly heterogeneous and internally repetitive hyperactive antifreeze proteins from the beetle Tenebrio molitor. Biochemistry 38:11415–11424

Liou YC, Tocilj A, Davies PL, Jia Z (2000) Mimicry of ice structure by surface hydroxyls and water of a beta-helix antifreeze protein. Nature 406:322–324

Logsdon JM Jr., Doolittle WF (1997) Origin of antifreeze protein genes: a cool tale in molecular evolution. Proc Natl Acad Sci USA 94:3485–3487

Marshall CB, Daley ME, Graham LA, Sykes BD, Davies PL (2002) Identification of the ice-binding face of antifreeze protein from Tenebrio molitor. FEBS Lett 529:261–267

Marshall CB, Daley ME, Sykes BD, Davies PL (2004) Enhancing the activity of a beta-helical antifreeze protein by the engineered addition of coils. Biochemistry 43:11637–11646

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

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 39:121–152

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818

Ramsey JA (1964) The rectal complex of the mealworm Tenebrio molitor, L. (Coleoptera, Tenebrionidae). Phil Trans R Soc Lond B 248:279–314

Raymond JA, DeVries AL (1977) Adsorption inhibition as a mechanism of freezing resistance in polar fishes. Proc Natl Acad Sci USA 74:2589–2593

Scott GK, Hew CL, Davies PL (1985) Antifreeze protein genes are tandemly linked and clustered in the genome of the winter flounder. Proc Natl Acad Sci USA 82:2613–2617

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

Swanson WJ, Aquadro CF (2002) Positive Darwinian selection promotes heterogeneity among members of the antifreeze protein multigene family. J Mol Evol 54:403–410

Swofford DL (2002) PAUP*: phylogenetic analyses using parsimony and other methods. Version 4.0b10. Sinauer Associates, Sunderland, MA

Taylor MS, Ponting CP, Copley RR (2004) Occurrence and consequences of coding sequence insertions and deletions in Mammalian genomes. Genome Res 14:555–566

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Tyshenko MG, Doucet D, Davies PL, Walker VK (1997) The antifreeze potential of the spruce budworm thermal hysteresis protein. Nat Biotechnol 15:887–890

Walker VK, Kuiper MJ, Tyshenko MG, Doucet D, Graether SP, Liou YC, Sykes BD, Jia Z, Davies PL, Graham LA (2001) Surviving winter with antifreeze proteins: Studies on budworms and beetles. In: Denlinger DL, Giebultowiz J, Saunder DS (eds) Insect timing: circadian rhythmicity to seasonality. Elsevier Science, Amsterdam, pp 199–212

Author Query Form

Journal : JMEV

Article No. : 02563

Disk Usage : ☐ Yes ☐ No

☐ Incompatible file format ☐ Virus infected

☐ Discrepancies between electronic file and hard copy

☐ Other: . . . . . .

☐ Manuscript keyed in ☐ Files partly used (parts keyboarded.)

Author Queries

Sr. No.	Query	Author's Remarks
1	Pls. update Jørgensen et al. 2006 if possible. Thank you.	