Flies expand the repertoire of protein structures that bind ice

Koli Basu, Laurie A. Graham, Robert L. Campbell, and Peter L. Davies[1]

Protein Function Discovery Group and Department of Biomedical and Molecular Sciences, Queen's University, Kingston, ON, Canada K7L 3N6

An antifreeze protein (AFP) with no known homologs has been identified in Lake Ontario midges (Chironomidae). The midge AFP is expressed as a family of isoforms at low levels in adults, which emerge from fresh water in spring before the threat of freezing temperatures has passed. The 9.1-kDa major isoform derived from a preproprotein precursor is glycosylated and has a 10-residue tandem repeating sequence xxCxGxYCxG, with regularly spaced cysteines, glycines, and tyrosines comprising one-half its 79 residues. Modeling and molecular dynamics predict a tightly wound left-handed solenoid fold in which the cysteines form a disulfide core to brace each of the eight 10-residue coils. The solenoid is reinforced by intrachain hydrogen bonds, side-chain salt bridges, and a row of seven stacked tyrosines on the hydrophobic side that forms the putative ice-binding site. A disulfide core is also a feature of the similar-sized beetle AFP that is a β-helix with seven 12-residue coils and a comparable circular dichroism spectrum. The midge and beetle AFPs are not homologous and their ice-binding sites are radically different, with the latter comprising two parallel arrays of outward-pointing threonines. However, their structural similarities is an amazing example of convergent evolution in different orders of insects to cope with change to a colder climate and provide confirmation about the physical features needed for a protein to bind ice.

antifreeze protein | convergent evolution | midge | disulfide-rich | solenoid

S pecies that adapt to survive freezing temperatures may have access to environmental resources with less competition and predation from freeze-susceptible organisms. An example of this environmental bonanza is illustrated by the sudden proliferation and radiation of notothenioid fishes into the ice-laden Antarctic Ocean during the Cenozoic ice ages (1) after they acquired antifreeze glycoproteins (2). Antifreeze proteins (AFPs) bind to ice and prevent seed ice crystals from growing and damaging the host organism (3). In the absence of AFPs, the melting and freezing temperature of an aqueous medium is the same. However, when AFPs are present, they adsorb to the ice surface and lower the freezing temperature below the melting point (4). This temperature difference is referred to as thermal hysteresis (TH) and is the definitive measure of antifreeze activity.

AFPs are soluble at millimolar concentrations in aqueous solution. The mechanism by which these freely soluble AFPs leave the liquid state of water to irreversibly adsorb to the surface of ice has been a puzzle for many years. Computational studies have suggested that the ice-binding site of an AFP might organize water into an ice-like clathrate pattern with low translational energy (5–7). When these protein-bound waters contact the quasi-liquid layer at the interface with ice, the two ice-like water arrangements could fuse together and then turn to ice at temperatures in the TH range. Structure–function analyses have shown that AFPs collectively have a variety of different surfaces (ice-binding sites) that can organize water molecules into a clathrate (8, 9).

The crystal structures of four arthropod AFPs have been solved (10–13), and two more structures have been modeled with confidence (14, 15). The five arthropod AFPs are not homologous to each other. This diversity in AFP sequence and structure

suggests that AFPs arose fairly recently and independently from different progenitors to have the same function through convergent evolution. It is expected that discovery and characterization of other insect AFPs will reveal additional novel AFPs that will help identify structural features needed for binding ice, and provide clues about when insect AFPs arose. Also, it is notable that all of the arthropod AFPs above have high TH activity and other properties associated with hyperactive AFPs (16). This is fitting because these insects are terrestrial and encounter temperatures down to −20 to −30 °C. In contrast, marine fishes produce AFPs for the purpose of freeze resistance in a narrower thermal range of about 1 °C.

In this study, we have obtained the sequence and derived a structural model for an AFP from a midge. Nonbiting midges belong to the family Chironomidae and are distributed globally (17, 18). An AFP from a dipteran (two-winged fly) has been reported in the gall midge (*Thecodiplosis japonesis*) but was only characterized to the level of AFP activity and apparent molecular weight (19). The life cycle of the midge we collected in Southern Ontario is 1 y long (Fig. 1A), and the majority of that time is spent in the larval form (18). They pupate in spring and float to the surface of the lake to emerge as adults (Fig. 1B), which live for a ∼2-wk period during which time they swarm to mate. Eggs are laid on the water surface to continue the annual cycle. We predicted that midges might produce AFPs as a precautionary measure in the event of freezing temperature during their short, but very important, adult stage in the spring. Here, we have purified the AFP as a family of isoforms that have no known homologs and have determined its solenoid fold by molecular modeling. The putative ice-binding site is a flat region of

---

Significance

A previously unidentified insect antifreeze protein has been characterized in a fly. It is present in the adult stage of midges that emerge from fresh water in the spring and need protection from occasional night frosts. The mature protein has 79 residues and is small and repetitive enough to be reliably modeled as a tightly wound solenoid structure. Each tandem repeat of 10 residues forms a coil of the solenoid that is cross-braced by a disulfide bond. This fold orients a row of seven stacked tyrosine side chains to make a flat surface, which is the predicted ice-binding site of the protein. This protein has no homologs in the database and has, therefore, recently evolved to protect its host from freezing.

BIOCHEMISTRY

Fig. 1. Midge life cycle and AFP purification by ice affinity purification. (*A*) Schematic of the 1-y midge life cycle. (*B*) Adult. (*C*) A beaker containing 50 g of frozen midges. (*D*) The soluble fraction obtained by centrifugation of the insect homogenate. (*E*) The ice fraction after one round of ice affinity purification. (*F*) The ice fraction after three rounds of ice affinity purification.

stacked tyrosines on one side of the molecule that has the potential to organize ice-like clathrate waters.

Materials and Methods

Midge AFP was purified from the soluble fraction of adult midge homogenates by sequential ice affinity purification (20). Antifreeze activity (TH) and ice crystal morphology were monitored using a nanoliter osmometer (21). Protein masses and the number of disulfide bonds were determined by matrix-assisted laser desorption/ionization (MALDI). Trypsin digestion and tandem mass spectrometry was used for protein sequencing. A repeating sequence motif (C-X-G-X-X) was used to find AFP transcripts from an Illumina paired-end sequencing library made from adult midge mRNA (22–24). The entire coding region of the midge AFP gene was deduced, and its accuracy was confirmed by RT-PCR. Circular dichroism (CD) was used to measure secondary structure for comparison with other repetitive AFPs. The mature AFP was modeled as right- and left-handed solenoids using the *Tenebrio molitor* AFP (*Tm*AFP) structure (PDB ID code 1EZG) as a template, and the stability of each model was tested by molecular dynamics. Detailed experimental procedures are given in *SI Experimental Procedures*.

Results

The Midge Species Is Unassigned. The midges used in this study were overwhelmingly from one species based on sequence of the 5′-region of the mitochondrial cytochrome C oxidase subunit 1 (COI) gene (25). Transcriptome analysis showed but a single nucleotide polymorphism (26) in the population (GenBank accession no. KM102730). A search for the COI gene in the Barcode of Life Database resulted in matches of 99.38–99.84% identities to 386 Chironomidae sequences, obtained mostly from insects collected around Southern Ontario. None of the matches comes from an assigned genus or species; therefore, we are referring to the insect as the Lake Ontario midge.

Midge AFP Purified Solely by Affinity for Ice. Aqueous extracts from freshly caught homogenized adult midges (Fig. 1*B*) shaped ice crystals into a rounded hexagonal bipyramid and depressed the freezing point by 0.25 °C below the melting temperature of the ice, both of which are conclusive signs of the presence of an AFP. To scale up AFP isolation, midges (50 g) were blended in buffer (Fig. 1 *C* and *D*) and subjected to three rounds of ice affinity purification. The first ice mass grown from the homogenate supernatant in Fig. 1*D* was a translucent tan color with distinct faceting on the surface (Fig. 1*E*). After this fraction was reextracted two times by ice affinity purification, the new ice mass was colorless, completely clear, and still showed faceting (Fig. 1*F*). The total protein yield was ~50 μg.

Midges Produce a Family of AFP Isoforms with Multiple Pairs of Cysteines and Some Glycosylation. When analyzed by MALDI, the ice affinity-purified AFPs had masses ranging from 5.7 to 10.4 kDa, with a 9,074-Da species as the most prominent peak (Fig. 2*A*). This pattern was highly reproducible over multiple extractions. After reduction and alkylation by iodoacetamide, the main AFP peak had a mass of 10,005 Da (Fig. 2*B*). Because the combination of these two modifications reduces disulfide bonds and adds a 58-Da group to every cysteine, the mass increase in the major isoform is consistent with the opening of eight disulfide bonds. Other isoforms in the mass range of 7.3–10.0 kDa also appeared to have 16 cysteine residues, whereas only 12 are found in the 6,146-Da isoform.

Although some isoform diversity likely resides at the gene level, another possible source is posttranslational modification. The major isoform at 9,074 Da (±6) is 860 Da heavier than the 8,214 (±6)-Da isoform in the much smaller adjacent peak (Fig. 2*A*). This mass difference corresponds exactly to the molecular weight of the N-linked glycan $GlcNAc_2Man_1Fuc_2$, which has been found to modify another insect AFP (*Tm*AFP) (27).

The Midge AFP Shapes Ice Similarly to Hyperactive AFPs but Has Intermediate TH Activity. The morphology of a single ice crystal in purified midge AFP solution was that of a hexagonal bipyramid with rounded vertices, like the shape of a lemon (Fig. 2*C*). This is typical of hyperactive AFPs and most similar to the shape observed with *Tm*AFP. At the freezing point, the ice crystal "bursts" perpendicular to the *c* axis (Fig. 2*D* and *SI Experimental Procedures*), which is characteristic of hyperactive AFPs (16) and typically reflects AFP binding to the basal plane.

At a concentration of 0.08 mM, the antifreeze-active sample from the adult midges had a TH activity of 1.2 (±0.04) °C (*SI Experimental Procedures*). Hyperactive insect AFPs have TH values of ~5.2 °C (spruce budworm AFP isoform 501) and ~3.2 °C (*Tm*AFP) at the same concentration (16). In contrast, the moderately active fish AFPs show ~0.25 °C (type I), and ~0.35 °C (type II and III AFPs) of TH activity, again at 0.1 mM concentration. Thus, the antifreeze-active components from the midges have a TH activity well above that of AFPs from fishes and plants but less than that shown by most hyperactive AFPs.
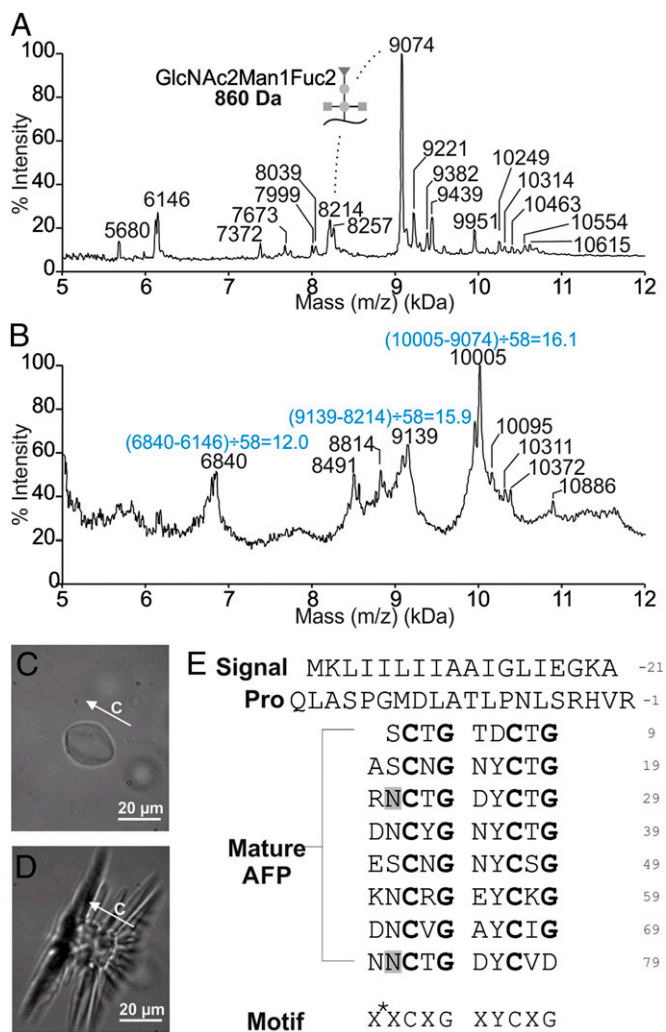
Fig. 2. Midge antifreeze protein (AFP) is expressed as a family of hyperactive isoforms with multiple cysteines and shapes ice in a manner similar to hyperactive AFPs. (*A*) MALDI mass analysis of the third ice affinity purification fraction containing AFPs purified from adult midges. The difference in mass between the 9,074- and 8,214-Da species corresponds to a single GlcNAc$_2$Man$_1$Fuc$_2$ modification. (*B*) MALDI mass analysis following reduction and alkylation of cysteine residues. The mass differences divided by 58 give the number of Cys in each of the major isoforms. (*C*) Morphology of a single ice crystal in a solution of midge AFP. (*D*) Explosive growth of the ice crystal when the temperature is lowered below the freezing temperature. The *c* axis and scale bar are indicated in *C* and *D*. (*E*) Midge AFP sequence (GenBank accession no. KM102729). The signal peptide, pro sequence, and mature protein regions of the midge AFP are labeled. The consensus sequence motif is presented below, with X* representing an acidic or basic residue, and X representing any residue. Glycosylation sites are highlighted in gray. The strict cysteine and glycine repetitions are highlighted in bold.

Tandem Mass Spectrometry Reveals a Repeating C-X-G-X-X Sequence in the Midge AFP. Two peptides, CTGASCNGNYCTGR and DNNCVGAYC(L or I)GNNCR, from the reduced, alkylated, and trypsin-treated AFP were sequenced by tandem mass spectrometry. The sequence C-X-G-X-X, where X is any residue, appears three times in tandem in the deduced sequences, with the last repeat being truncated by trypsin cleavage.

The Major AFP Isoform Comprises Tandem 10-Residue Repeats. Total RNA extracted from frozen adult insects using the TRIzol method showed no sign of degradation. However, a brown contaminant, which was likely melanin, copurified with the RNA

and was largely removed by a previously published method developed for mosquito RNA (23). Illumina paired-end sequencing on the bridge-amplified mRNA yielded ~110 million reads. The pentapeptide repeating motif was used as a query to search for RNA fragments corresponding to AFP gene transcripts. Using Seedtop+, 491 reads from the entire RNA-seq database were found to encode a [CXGXX]$_4$C series, but 65 of these were discarded as they contained stop codons and were likely spurious matches. Using random Seedtop+ hits as seed sequences, the AFP transcript was built by de novo assembly in an iterative approach (with sequential extraction of additional overlapping fragments from the database) and a single transcript was built with confidence using 90 reads that spanned from start to stop codons. Of these 90 reads, 56 were found in the set of 426 Seedtop+ hits. Many of the other matches likely correspond to transcripts encoding other AFP isoforms.

The assembled AFP transcript codes for 117 residues (Fig. 2*E*). It begins with an 18-residue N-terminal signal peptide for export of the AFP out of the cell. The signal peptide is followed by a 20-residue pro sequence, which is not part of the mature protein because its molecular weight is not accounted for in any of the MALDI peaks (Fig. 2*A*). The mature protein is 79 residues long and has a theoretical mass of 8,197 Da when all 16 cysteines are disulfide bonded, which is 17 Da less than the mass (8,214 ± 6 Da) of the major isoform (Fig. 2*A*). The difference likely corresponds to a posttranslation modification such as methylation or hydroxylation (28). The C-X-G-X-X repeat found in the sequences determined by tandem mass spectrometry is duplicated in a 10-residue repeat of X*-X-C-X-G-X-Y-C-X-G, where X* are typically acidic or basic residues. This 10-residue repeat occurs eight times and contains two Cys and Gly residues each spaced at 5-residue intervals and one Tyr at 10-residue intervals. The only exceptions are in the first repeat, which is one residue shorter, where Asp substitutes for Tyr, and in the last repeat where the C-terminal residue is Asp instead of Gly. Elsewhere, the sequence is highly biased toward small polar (Asn, Thr, Ser) or charged (Lys, Arg, Asp, Glu) residues with only two aliphatic residues (Val and Ile) present. Two possible N-linked glycosylation sites are highlighted in gray (Fig. 2*E*).

Unusual CD Spectrum. The CD spectrum of midge AFP has a maximum molar ellipticity at 185 nm and a minimum at 205 nm, after which there is a gradual return to zero from 205 to 260 nm with an inflection point at 224 nm (Fig. 3*A*). The spectrum is unusual, and its deconvolution using the structural database does not correlate with the common secondary structures of α-helix, β-strand, β-turn, or random coil. It is instructive, however, to compare it to the CD spectrum of the tightly coiled *Tm*AFP (Fig. 3*B*) (29). A trough at 205 nm is seen in both midge AFP and *Tm*AFP spectra, whereas the inflection point in the midge spectrum at 224 nm deepens to a trough in the *Tm*AFP spectrum, which shows a small amount of β-structure. Other coiled AFPs like *Marinomonas primoryensis* AFP (*Mp*AFP) and *Rhagium inquisitor* AFP (*Ri*AFP), with larger average coils of 19 and 21 residues, respectively, both have appreciable β-content that shows in their CD profiles (Fig. 3 *C* and *D*) (30, 31).

Midge AFP Has a Solenoid Fold Braced by Disulfide Cross-Bridges. The 10-residue tandem repeats of the midge AFP, with cysteines at every fifth position, are reminiscent of the beetle AFP (*Tm*AFP) sequence (32). *Tm*AFP comprises 12-residue tandem repeats with disulfide-bonded cysteines present at 6-residue intervals. Its crystal structure (PDB ID code 1EZG) is a right-handed β-solenoid (33) with 12-residue coils that are bisected and stabilized by internal disulfide bridges (Fig. 4*A*) (10). A single parallel β-sheet runs the length of the solenoid and comprises the ice-binding site formed from two parallel ranks of outward-facing Thr derived from the TCT ice-binding motifs. On one side of the
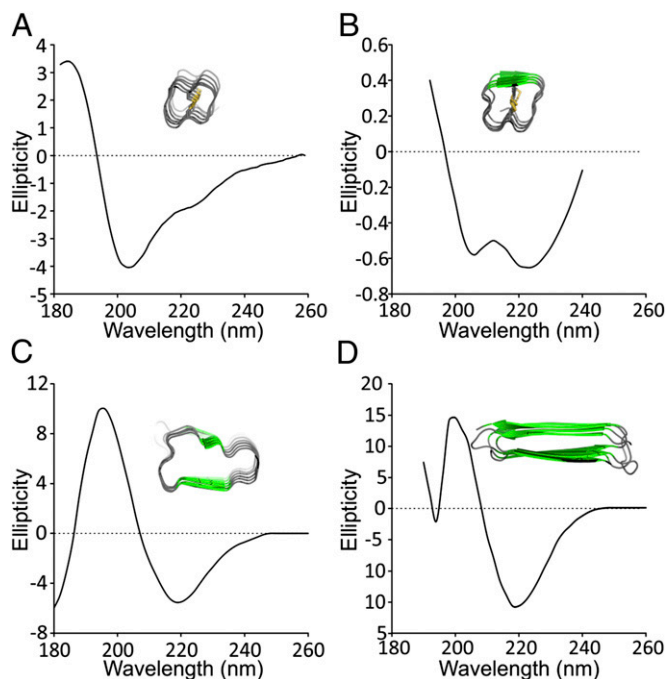
Fig. 3. Unusual CD spectrum of the midge AFP and comparison with other repetitive AFPs. (*A*) Midge AFP [molar ellipticity (in degrees square centimeters per decimole times 10⁻³)]. (*B*) *Tm*AFP [mean residue molar ellipticity (in degrees square centimeters per decimole times 10⁻⁴)]. Reprinted from ref. 29 with permission from Elsevier; www.sciencedirect.com/science/journal/10465928. (*C*) *Mp*AFP [ellipticity (in millidegrees)]. Data from ref. 30. (*D*) *Ri*AFP [mean residue ellipticity (in degrees square centimeters per decimole times 10⁻³)]. Data from ref. 31. *Insets* are cross-sections of the helical AFPs showing β-structure as green arrows.

disulfide ladder, there is a row of inward-pointing Ala and waters. The other side of the divided core has a row of internal Ser. The other eight residues face outward. Although the coils in *Tm*AFP are already incredibly tightly wound, we reasoned that a smaller cross-section could be obtained with a 10-residue repeat if the disulfide-bonding pattern was maintained but the internal-pointing Ala and Ser were replaced by Gly (Fig. 4*B*). The smaller and more flexible Gly residues could potentially accommodate the strain of making a 10- rather than a 12-residue disulfide-bonded loop. Based on this reasoning, midge AFP was modeled on the *Tm*AFP structure as both right-handed and left-handed solenoids. Only the left-handed model (Fig. 4*C* and *SI Text*) was stable during 20 ns of molecular dynamics simulations at 25 °C (Fig. 4*D*). The stability of the model was qualitatively assessed by observing visual changes in the fold and quantitatively by the root-mean-square deviation (rmsd) of the model from the first frame over the length of the simulation. The right-handed model changed drastically in the first 2 ns of molecular dynamics corresponding to a sudden increase in rmsd of 1.8 Å. Over the remainder of the simulation, the model slowly unraveled and the rmsd gradually increased to 5 Å. The left-handed model moved very little throughout the simulation, and the rmsd remained nearly constant, averaging 0.7 (±0.1) Å, indicating that the left-handed model is stable.

The Ramachandran plot of the left-handed model shows that all residues have allowed backbone conformations (Fig. 4*E*). Although several of the residues are in the β-region of the Ramachandran plot, and the midge AFP structure was modeled using a β-solenoid fold, there is no β-structure in the resulting model based on DSSP calculations (34, 35). This is likely due to the tightness of the coils preventing extended backbone

conformations of adjacent residues. The model is therefore referred to simply as a solenoid. In the final model, the 10-residue repeat makes up one coil of the solenoid (Fig. 4*B*). Replacement of the internal Ala and Ser relative to *Tm*AFP with Gly eliminates any core other than the disulfide bridges and these are aligned without twist, even following molecular dynamics. The model is stabilized along one side by pi stacking of seven Tyr and on the opposite side by salt bridging of adjacent acidic and basic residues on neighboring loops. Ice-binding faces are usually the most regular, flattest, and most hydrophobic surfaces on AFPs. The Tyr-rich surface meets all these criteria (Fig. 5 *A* and *C*), and therefore, it likely forms the ice-binding surface. The opposite
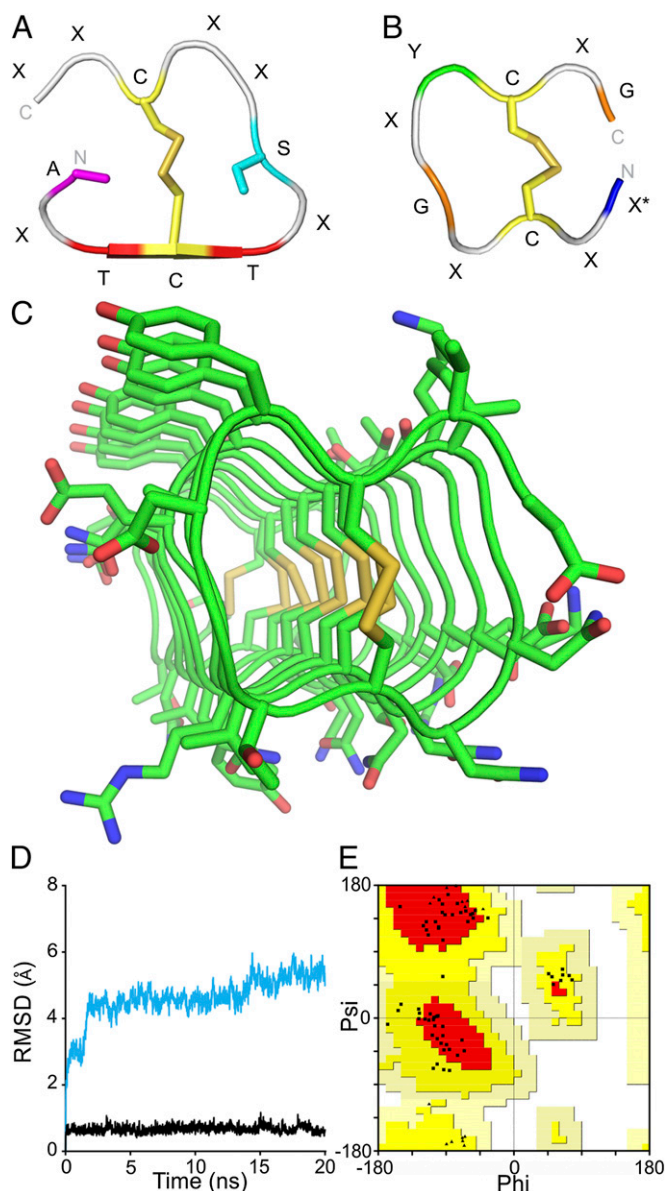

Fig. 4. Midge AFP structural model. (*A* and *B*) One coil of the (*A*) *Tm*AFP and (*B*) left-handed midge AFP repeating structure. Only internal-helix side chains are shown in *A* and *B*. (*C*) Left-handed midge AFP model. (*D*) rmsd of right-handed (cyan) and left-handed (black) midge AFP models during molecular dynamics simulation. (*E*) Ramachandran plot of the left-handed midge AFP model showing all nonglycyl residues as squares and all glycyl residues as triangles. Most favored regions are in red, additionally allowed regions are in yellow, generously allowed regions are in peach, and disallowed regions are in white.
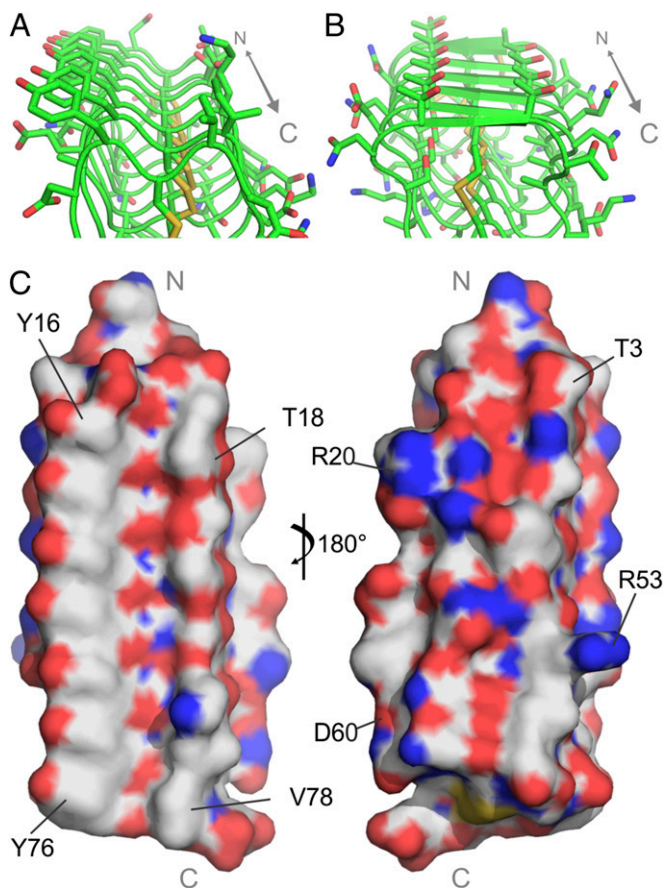
Fig. 5. The midge AFP has ice-binding residues not observed in other AFPs. (*A*) Predicted ice-binding site (IBS) of midge AFP. (*B*) IBS of *Tm*AFP. The C-terminal ends are to the front in *A* and *B*. (*C*) Surface representation of the midge AFP model showing the putative flat ice-binding surface (*Left*) and the opposing charged, uneven surface (*Right*). The N and C termini are indicated in gray.

surface has several bulky and charged side chains, making the surface much more polar and uneven (Fig. 5*C*).

Discussion

We have had considerable success modeling the previously unidentified folds of AFPs that have repetitive sequences and no homologs in the database. The AFP from snow fleas (sfAFP) was modeled as a set of tightly packed type-II polyproline helices stabilized by backbone-to-backbone hydrogen bonding and two disulfide bridges (36). The crystal structure of sfAFP solved a year later showed that the packed coil region had an rmsd of ~1 Å from the model (12). The plant AFP, *Lolium perenne* AFP (*Lp*AFP), was modeled as a β-solenoid with internal asparagine ladders and a small hydrophobic core of apposed valines (37). All these crucial structural features were validated by the crystal structure done more than 10 y later, albeit the modeled solenoid was of the opposite handedness (38). The bacterial ice adhesion domain of *Mp*AFP was correctly modeled as a β-solenoid based on the RTX repeats of an alkaline protease, which predicted that one of the two rows of internal Ca^{2+} ions was absent and replaced by hydrophobic packing (8, 30). Another AFP structure predicted de novo is that of the inquisitor beetle (*Ri*AFP), which, along with inchworm AFP (iwAFP), was modeled as a flat, silk-like β-solenoid (14). Although the iwAFP structure has not yet been experimentally determined, the recently solved structure of the similar *Ri*AFP confirms the accuracy of the modeled

structure (13). In all of these examples, and in the modeling of the midge AFP here, sequence repetition in the protein allows for the prediction of the fold of a unit motif, such as the coil of a solenoid, which can be replicated to form the entirety of the molecule. Molecular dynamics used to assess stability in a simulated environment offers evidence of either the accuracy of the modeled structure or the need to revise the model. A good example of this is the handedness of the β-solenoids. With some helices like *Lp*AFP, there is little to distinguish the stability of a right- versus a left-handed coil. However, here, there is clear evidence from molecular dynamics that the midge AFP structure is only stable as a left-handed structure. Using some features from a nonhomologous AFP with a similar CD spectrum, intuition, and molecular dynamics, we have modeled the Lake Ontario midge AFP structure with confidence as a 10-residue repeating solenoid, the novelty of which is the lack of any β-structure within an incredibly tight amphipathic coil and a row of stacked, surface-exposed Tyr.

Although the midge AFP model was inspired by the crystal structure of *Tm*AFP, similarities in their structures do not arise from homology but are the result of convergent evolution to a similar fold (39). *Tm*AFP, which is a β-solenoid with seven 12-residue loops, six disulfide bonds through the core, and two disulfide bonds involved in a capping structure, was remarkable for being the tightest experimentally observed β-helix (40). Midge AFP, which has 10-residue loops in our model, is even more tightly coiled than *Tm*AFP. The sequence of the two proteins cannot be aligned unless two gaps are inserted per loop in the midge AFP to force alignment of the Cys, resulting in less than 20% of the non-Cys residues matching in any register. Unlike the *Tm*AFP fold, secondary structure analyses of midge AFP do not find any β-strand content in the model, and therefore the latter should not be classified as a β-solenoid. It can, however, be described as a solenoid due to the regularity of the protein backbone dihedral angles at each position of the 10-residue motif. The model was validated by the Ramachandran plot and by the midge AFP's CD profile, which looks like that of *Tm*AFP but without any β-structure. The tight and stable fold would result in an extremely rigid structure, reduced freedom of motion of side chains, and an increased potential for water molecule ordering on the protein surface, which seems to be a requirement for ice binding.

Distances between adjacent loops are ~4.8 Å, which is typical for a solenoid fold. Indeed, it is this distance, seen in the β-helical AFPs, which closely matches the 4.5-Å critical repeating distance of oxygen atoms on both the primary prism and basal planes of ice (41). Binding to both these planes can explain hyperactivity (16, 42). We predict that the Tyr ladder will play a role in ice binding through formation of clathrate waters around the phenyl groups with "anchoring" to the Tyr hydroxyls and backbone peptide groups. The Tyr-rich ice-binding site is unlike those of other solenoid AFPs, which are typically rich in Thr (Fig. 5*B*).

The data collected by sequencing of the Lake Ontario midge transcriptome confirmed that the complexity of the AFP species in the midge was not only due to posttranslational modifications but also resulted from multiple isoforms of the protein. The major isoform has an N-terminal signal peptide and pro sequence, which are cleaved from the mature AFP. Cleavage of this pro sequence from the mature AFP occurs at the C-terminal side of an arginine, which is a known cleavage site for trypsin-like proteases that are present in fly hemolymph (43). Pro sequences are present in the winter flounder type I AFP, sea raven type II AFP, and iwAFP but their functions are unknown (14, 44, 45). They may assist in proper protein folding and then become cleaved during or after export from the cell.

AFPs are typically expressed by terrestrial insects in their overwintering form, when they may face temperatures as low as −30 °C for weeks at a time (46). Because fresh water freezes at

0 °C, it is not surprising that there is no sign of antifreeze activity in the overwintering bloodworm stage of the Lake Ontario midges, which would have a lower freezing point due to salts and other solutes within their tissues. Instead, the antifreeze-producing stage of the midges is the adult fly, which emerges from fresh water in the spring. Spring temperatures are on average above freezing, but there are occasional night frosts. Therefore, unlike other insect AFPs, which must be both highly expressed and extremely potent in order for the insect to survive overwintering, the Lake Ontario midges produce a small quantity of AFPs to counter short subzero periods, consistent with the microgram quantities of AFPs purified from 50 g of midges by ice affinity purification. It will be interesting to examine midge populations from colder regions to see whether the amount of AFP they produce is higher. This study has provided the tools to do this analysis.

The Lake Ontario midge is the fifth insect and sixth arthropod to yield a novel AFP. We hypothesize that characterization of other insect AFPs will be fruitful for the discovery of additional uncharacterized AFP folds.

1. Near TJ, et al. (2012) Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proc Natl Acad Sci USA* 109(9):3434–3439.
2. Chen L, DeVries AL, Cheng CH (1997) Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *Proc Natl Acad Sci USA* 94(8):3817–3822.
3. DeVries AL (1971) Glycoproteins as biological antifreeze agents in Antarctic fishes. *Science* 172(3988):1152–1155.
4. Raymond JA, DeVries AL (1977) Adsorption inhibition as a mechanism of freezing resistance in polar fishes. *Proc Natl Acad Sci USA* 74(6):2589–2593.
5. Nutt DR, Smith JC (2008) Dual function of the hydration layer around an antifreeze protein revealed by atomistic molecular dynamics simulations. *J Am Chem Soc* 130(39):13066–13073.
6. Yang C, Sharp KA (2004) The mechanism of the type III antifreeze protein action: A computational study. *Biophys Chem* 109(1):137–148.
7. Smolin N, Daggett V (2008) Formation of ice-like water structure on the surface of an antifreeze protein. *J Phys Chem B* 112(19):6193–6202.
8. Garnham CP, Campbell RL, Davies PL (2011) Anchored clathrate waters bind antifreeze proteins to ice. *Proc Natl Acad Sci USA* 108(18):7363–7367.
9. Sun T, Lin FH, Campbell RL, Allingham JS, Davies PL (2014) An antifreeze protein folds with an interior network of more than 400 semi-clathrate waters. *Science* 343(6172):795–798.
10. 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(6793):322–324.
11. Leinala EK, Davies PL, Jia Z (2002) Crystal structure of beta-helical antifreeze protein points to a general ice binding model. *Structure* 10(5):619–627.
12. Pentelute BL, et al. (2008) X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers. *J Am Chem Soc* 130(30):9695–9701.
13. Hakim A, et al. (2013) Crystal structure of an insect antifreeze protein and its implications for ice binding. *J Biol Chem* 288(17):12295–12304.
14. Lin FH, Davies PL, Graham LA (2011) The Thr- and Ala-rich hyperactive antifreeze protein from inchworm folds as a flat silk-like β-helix. *Biochemistry* 50(21):4467–4478.
15. Guz N, Toprak U, Dageri A, Oktay Gurkan M, Denlinger DL (2014) Identification of a putative antifreeze protein gene that is highly expressed during preparation for winter in the sunn pest, *Eurygaster maura*. *J Insect Physiol* 68:30–35.
16. Scotter AJ, et al. (2006) The basis for hyperactivity of antifreeze proteins. *Cryobiology* 53(2):229–239.
17. Oliver D (1971) Life history of the Chironomidae. *Annu Rev Entomol* 16(1):211–230.
18. Pinder L (1986) Biology of freshwater Chironomidae. *Annu Rev Entomol* 31(1):1–23.
19. Li Y, Gong H, Park HY (2000) Purification and partial characterization of thermal hysteresis proteins from overwintering larvae of pine needle gall midge, Thecodiplosis japonensis (Diptera: cecidomiidae). *Cryo Lett* 21(2):117–124.
20. Kuiper MJ, Lankin C, Gauthier SY, Walker VK, Davies PL (2003) Purification of antifreeze proteins by adsorption to ice. *Biochem Biophys Res Commun* 300(3):645–648.
21. Braslavsky I, Drori R (2013) LabVIEW-operated novel nanoliter osmometer for ice binding protein investigations. *J Vis Exp* 2013(72):e4189.
22. Bogart K, Andrews J (2006) Extraction of total RNA from *Drosophila*. *CGB Technical Report* (The Center for Genomics and Bioinformatics, Indiana University, Bloomington, IN), 2006-10.
23. Satyamoorthy K, Li G, Van Belle PA, Elder DE, Herlyn M (2002) A versatile method for the removal of melanin from ribonucleic acids in melanocytic cells. *Melanoma Res* 12(5):449–452.
24. Grabherr MG, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29(7):644–652.
25. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 3(5):294–299.
26. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178–192.
27. 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(35):11415–11424.
28. Wilkins MR, et al. (1999) High-throughput mass spectrometric discovery of protein post-translational modifications. *J Mol Biol* 289(3):645–657.
29. Liou YC, et al. (2000) Folding and structural characterization of highly disulfide-bonded beetle antifreeze protein produced in bacteria. *Protein Expr Purif* 19(1):148–157.
30. Garnham CP, et al. (2008) A Ca²⁺-dependent bacterial antifreeze protein domain has a novel beta-helical ice-binding fold. *Biochem J* 411(1):171–180.
31. Hakim A, Thakral D, Zhu DF, Nguyen JB (2012) Expression, purification, crystallization and preliminary crystallographic studies of Rhagium inquisitor antifreeze protein. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 68(Pt 5):547–550.
32. Graham LA, Liou YC, Walker VK, Davies PL (1997) Hyperactive antifreeze protein from beetles. *Nature* 388(6644):727–728.
33. Kobe B, Kajava AV (2000) When protein folding is simplified to protein coiling: The continuum of solenoid protein structures. *Trends Biochem Sci* 25(10):509–515.
34. Joosten RP, et al. (2011) A series of PDB related databases for everyday needs. *Nucleic Acids Res* 39(Database issue):D411–D419.
35. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22(12):2577–2637.
36. Lin FH, Graham LA, Campbell RL, Davies PL (2007) Structural modeling of snow flea antifreeze protein. *Biophys J* 92(5):1717–1723.
37. Kuiper MJ, Davies PL, Walker VK (2001) A theoretical model of a plant antifreeze protein from *Lolium perenne*. *Biophys J* 81(6):3560–3565.
38. Middleton AJ, et al. (2012) Antifreeze protein from freeze-tolerant grass has a beta-roll fold with an irregularly structured ice-binding site. *J Mol Biol* 416(5):713–724.
39. Liou YC, Davies PL, Jia Z (2000) Crystallization and preliminary X-ray analysis of insect antifreeze protein from the beetle *Tenebrio molitor*. *Acta Crystallogr D Biol Crystallogr* 56(Pt 3):354–356.
40. Kajava AV, Steven AC (2006) Beta-rolls, beta-helices, and other beta-solenoid proteins. *Adv Protein Chem* 73:55–96.
41. Graether SP, et al. (2000) Beta-helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect. *Nature* 406(6793):325–328.
42. Pertaya N, Marshall CB, Celik Y, Davies PL, Braslavsky I (2008) Direct visualization of spruce budworm antifreeze protein interacting with ice crystals: Basal plane affinity confers hyperactivity. *Biophys J* 95(1):333–341.
43. Gorman MJ, Andreeva OV, Paskewitz SM (2000) Molecular characterization of five serine protease genes cloned from Anopheles gambiae hemolymph. *Insect Biochem Mol Biol* 30(1):35–46.
44. Davies PL, Roach AH, Hew CL (1982) DNA sequence coding for an antifreeze protein precursor from winter flounder. *Proc Natl Acad Sci USA* 79(2):335–339.
45. Ng NF, Trinh KY, Hew CL (1986) Structure of an antifreeze polypeptide precursor from the sea raven, *Hemitripterus americanus*. *J Biol Chem* 261(33):15690–15695.
46. Kristiansen E, et al. (2005) Isolation and characterization of hemolymph antifreeze proteins from larvae of the longhorn beetle *Rhagium inquisitor* (L.). *Comp Biochem Physiol B Biochem Mol Biol* 142(1):90–97.