# Supporting Information

## Basu et al. 10.1073/pnas.1422272112

### SI Experimental Procedures

**Collection and Storage of Insects.** Adult midges were collected en masse from our building air intake filters in late April and early May. Most of the insects were alive when collected and were stored at −80 °C in 200-g aliquots until use.

**Purification of Insect Total RNA.** Total RNA was extracted from adult midges using the TRIzol method, as described for *Drosophila* (1), with a slight modification; the insects were pulverized under liquid nitrogen in a prechilled ceramic mortar and pestle. A brown contaminant, which purified along with the RNA, was removed by two rounds of polyacrylamide bead purification (Bio-Rad; BioGel P-60) as described previously (2), and one round of silica-gel–based membrane filtration (Qiagen; RNeasy Midi Kit). The amount of RNA was quantified by nanospectrophotometry (Thermo Scientific; NanoDrop 2000), and the quality was assessed by agarose gel electrophoresis in the presence of ethidium bromide.

**Species Analysis.** PCR was used to amplify the 5′-region of the mitochondrial cytochrome C oxidase (COI) subunit 1 gene to determine the most abundant species of midge in the collection. Because of the high copy number of mitochondrial DNA (mtDNA) per animal cell, and small size of the mtDNA genome, it is likely that some mtDNA was present in the total RNA preparation, and, therefore, the total RNA preparation sufficed as the mtDNA source. COI primer pair (5′-TTTCTACAAATCATAAAGATATTGG-3′) and (5′-TAAACTTCAGGGTGACCAAAAAATCA-3′) (3) were used for PCR in a total volume of 50 μL with final concentrations of 47 ng/μL RNA preparation, 1× Taq buffer (Thermoscientific; $+(NH_4)SO_4$, $-MgCl_2$), 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.4 μM of each primer, and 0.05 units/μL of Taq polymerase (Thermoscientific). The PCR thermocycling program consisted of 95 °C for 5 min, then 30 cycles at 95 °C for 1 min, 50 °C for 1 min, and 72 °C for 2 min, followed by a 10-min 72 °C incubation for TA cloning. PCR products were cloned into the pCR 2.1-TOPO vector using the TOPO TA cloning kit (Invitrogen) and transformed into One Shot Top10, chemically competent *Escherichia coli* (Invitrogen). Plasmids from 15 colonies were purified by the GeneJet Plasmid Miniprep Kit and sequenced using the T7 promoter at the Robarts–London Regional Genomic Centre (London, ON, Canada). The RNA-seq library was searched with BLASTn with the partial PCR amplified COI as the query, and all reads were aligned onto the COI using Trinity. Presence of polymorphisms in the COI was assessed using Integrated Genome Viewer.

**Purification of Midge Antifreeze Protein.** Frozen adult midges (50 g) were homogenized at 4 °C in a blender (Sunbeam; Osterizer 8) with 150 mL of 50 mM ammonium bicarbonate (pH 7.9), 2 mM phenylthiocarbamide, into which 3 Complete protease inhibitor mixture tablets (Roche) were dissolved. The homogenate was centrifuged for 1 h at 48,254 × *g*. The supernatant was saved, and the pellet was rehomogenized in the blender with 80 mL of 50 mM ammonium bicarbonate and recentrifuged under the same conditions. The supernatants were pooled, filtered through glass wool to remove lipids, and brought to 500 mL with 50 mM ammonium bicarbonate (pH 7.9) in preparation for ice affinity purification. Ice affinity purification was performed as described previously (4) at a cooling rate of 1 °C per day from a starting temperature of −0.7 °C and was stopped when at least 50% of the solution was incorporated into the ice fraction (~4 d). Ice affinity purification was conducted three times, with the second and third rounds being performed on the preceding ice fraction. After each round, the ice fraction was melted slowly while 1 M ammonium bicarbonate (pH 7.9) was added stepwise to replenished the lost buffer and maintain a concentration of ~50 mM. The volume was again increased to 500 mL before subsequent rounds of ice affinity purification. The last melted ice fraction was concentrated by lyophilization or by using a centrifugal concentrator (Millipore) with a 3-kDa molecular weight cutoff. Protein concentration was determined by absorbance measurements at 280 nm and an extinction factor of 12,920 $M^{-1}{\cdot}cm^{-1}$ predicted from the sequence of the major mature AFP sequence using the Expasy ProtParam tool (web.expasy.org/protparam).

**Reduction and Alkylation of the AFP Sample.** To 4 μL of 0.05 mg/mL AFP in 50 mM ammonium bicarbonate (pH 7.9), 2.5 μL of 4% (vol/vol) RapiGest SF detergent (Waters) also in 50 mM ammonium bicarbonate (pH 7.9) and 1 μL of 100 mM DTT were added. The mixture was incubated for 30 min at 67 °C, cooled to room temperature, and then 1 μL of 1 M iodoacetamide and 1.5 μL of water were added. This mixture was incubated in the dark at room temperature for 30 min.

**MALDI.** The masses of the purified proteins, both unmodified and modified by alkylation, were determined by MALDI mass spectrometry using an α-cyano-4-hydroxycinnamic acid matrix and the dried droplet method. The instrument used was a SCIEX Voyager DE Pro in Linear mode/delayed extraction, and data were processed using Applied Biosystems Data Explorer (version 4.0).

**Thermal Hysteresis Measurements.** Thermal hysteresis measurements to assess AFP activity were performed as described previously (5). All measurements were done with a 1-min hold at the melting temperature, and a cooling rate of 0.01 °C every 4 s.

**MS/MS Protein Sequencing.** Reduced and alkylated samples were digested overnight at a 100:1 ratio of protein to Promega trypsin V5111. Tandem mass spectrometry was conducted using Waters QTOF Global with a Waters CapLC XE chromatography system (LC Packings C18 PepMap100, 3 μm, 100-Å column) in data-dependent acquisition electrospray ionization, positive, mode. Waters MassLynx, version 4.0, was used for data acquisition and analysis, and the raw data were processed using Waters Protein Lynx, version 2.1.5.

**Transcriptome Analysis.** Midge total RNA was sent to The Centre for Applied Genomics (Hospital for Sick Children, Toronto, ON, Canada) for Illumina paired-end sequencing of the mRNA. A total of 112 million reads, of length 101 bp each, were obtained. The raw data were translated in all six reading frames and searched using the BLAST Seedtop+ function, which looks for a specified pattern. The seedtop query was $(C\text{-}X\text{-}G\text{-}X\text{-}X)_4\text{-}C$, where C is a cysteine, X is any amino acid, G is glycine, and 4 is the number of contiguous repetitions of the pattern searched. Seedtop+ hits were manually curated to remove all spurious hits. Spurious hits included reads with stop codons within the repeating pattern and three or more adjacent cysteine codons, which was not expected for the AFP. The raw data were also converted to a BLAST database that was searchable with the BLASTn command using the Seedtop+ hits as queries. Read hits and their mates were assembled at the nucleotide level using the de novo assembly software, Trinity (6). The quality of the read

alignments to the contig was assessed visually using Integrated Genome Viewer (IGV) (7). Portions of the contigs with good quality alignments (high read coverage and few discrepancies between read sequences) were used as BLASTn queries for the next round of assembly. Contigs with good quality alignments were assembled together, and using this iterative process the major isoform of the midge AFP transcript was built from the start codon to the stop codon.

**Confirmation of AFP Transcript.** A midge cDNA pool was made using the Invitrogen Thermeoscript RT-PCR System, oligo(dT)$_{20}$ primers, and total midge RNA. AFP primer pair (5′-GCAGTCGAAGTTATCAGAATC-3′) and (5′-GCATTAAAGATGGGTTCGAG-3′) were used for PCR using the Roche, Expand High Fidelity PCR System. The PCR thermocycling program consisted of 30 cycles at 95 °C for 1 min, 53 °C for 1 min, and 72 °C for 2 min, followed by a 10-min 72 °C incubation for TA cloning. PCR products were TA cloned and sequenced exactly like the COI subunit 1 gene above.

**Secondary Structure Estimation.** Circular dichroism (CD) was performed at 4 °C using a Chirascan spectrophotometer (Applied Photophysics). The sample was prepared in 10 mM sodium phosphate (pH 8) at a protein concentration of 12 μM. The condition was scanned six times and corrected for the buffer scan. Proviewer Software (Applied Photophysics) was used for preliminary data processing, and OLIS Spectralworks (On-line Instruments) was used for deconvolution.

**Model Building and Molecular Dynamics.** The sequence of the midge AFP has short tandem repeats containing regularly spaced cysteine residues similar to *Tm*AFP, and therefore its structure was modeled based on the *Tm*AFP crystal structure (8). *Tm*AFP is a right-handed β-helix with a disulfide core. The repeating sequence in *Tm*AFP is 12 residues long and makes up one loop of the solenoid. The repeating sequence of the midge AFP is 10 residues long. Therefore, using PyMOL (www.pymol.org), two outward-pointing residues were deleted per loop from the *Tm*AFP structure, the inward-pointing Ala and Ser were changed to Gly, and the remaining residues were mutated to the midge AFP sequence. The model was energy minimized using GROMACS. The minimized model was refined to ensure all peptide bonds were in *trans* confirmation using the PyMOL sculpt function and that the ϕ and ψ angles of the peptide backbone were favorable when evaluated on a Ramachandran plot using COOT. Using GROMACS, the refined model was solvated in a box of water and energy minimized, followed by constant-volume and constant-pressure position-restrained molecular dynamics runs, each 0.1 ns in duration. The position-restrained dynamics runs were followed by 20 ns of unrestrained molecular dynamics simulation. The temperature variable for all GROMACS dynamics runs was set to 298 K. The temperature and pressure were maintained with the v-rescale and Parrinello–Rahman protocols, respectively. To make the left-handed model of the midge AFP, PyMOL was used to turn four loops of the *Tm*AFP structure into a left-handed solenoid by breaking two peptide bonds per loop (Fig. S3*A*), rebuilding the bonds in adjacent loops (Fig. S3*B*), then renumbering the residues so the numbering follows the new left-handed turn (Fig. S3*C*). GROMACS was used to energy minimize the left-handed partial-*Tm*AFP (Fig. S3*D*), and it was then used as a template to build the left-handed midge AFP model, as was done for the right-handed model. Model refinement, energy minimization, position-restrained dynamics, and molecular dynamics of the left-handed model were conducted exactly as described for the right-handed model. The root-mean-square deviation (rmsd) of the model during the simulation was calculated by the GROMACS g_rms function on the main chain including hydrogen atoms.


1. 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.
2. 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.
3. 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.
4. 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.
5. Braslavsky I, Drori R (2013) LabVIEW-operated novel nanoliter osmometer for ice binding protein investigations. *J Vis Exp* 2013(72):e4189.
6. Grabherr MG, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29(7):644–652.
7. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* 14(2): 178–192.
8. 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.

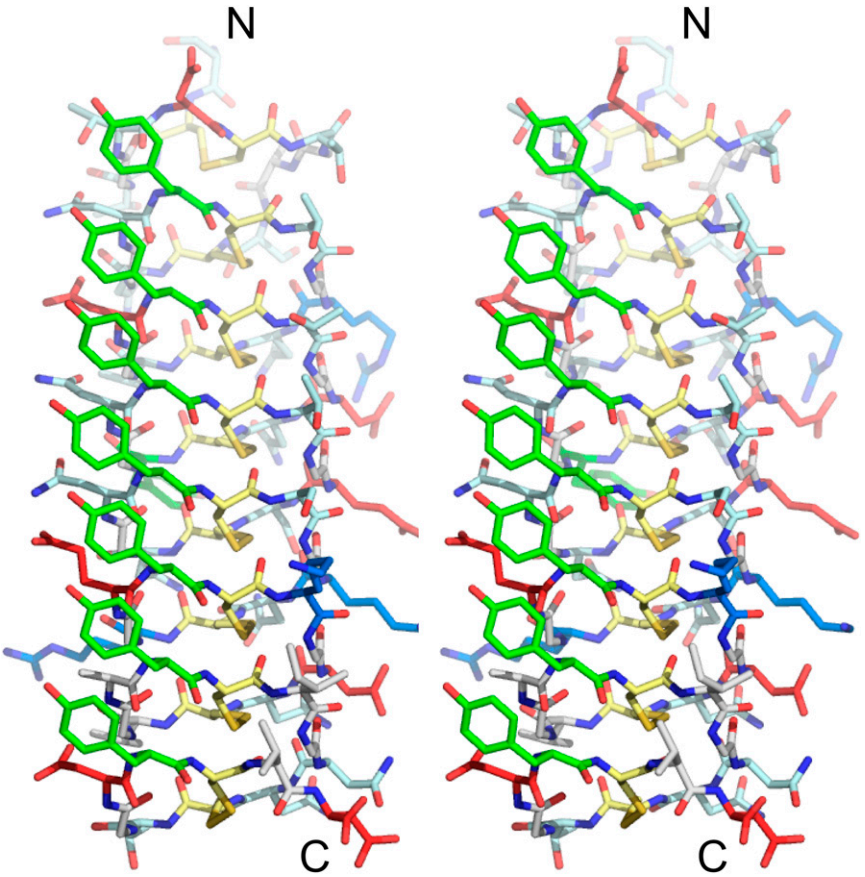



**Fig. S1.** A stereoview of the midge AFP model. Tyr are green, the acidic residues are red, basic residues are blue, and Cys are yellow. Coordinates of the model are available for inspection at pldserver1.biochem.queensu.ca/~rlc/MidgeAFP/.

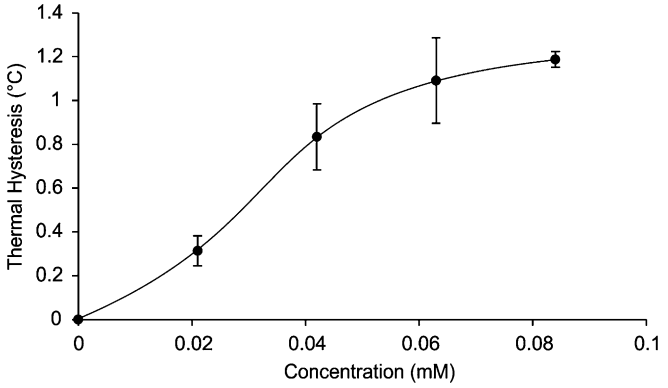



**Fig. S2.** Thermal hysteresis activity (in degrees Celsius) of the midge AFP as a function of protein concentration (millimolar). Readings were done in triplicate, and the error bars represent the SD of each measurement set.

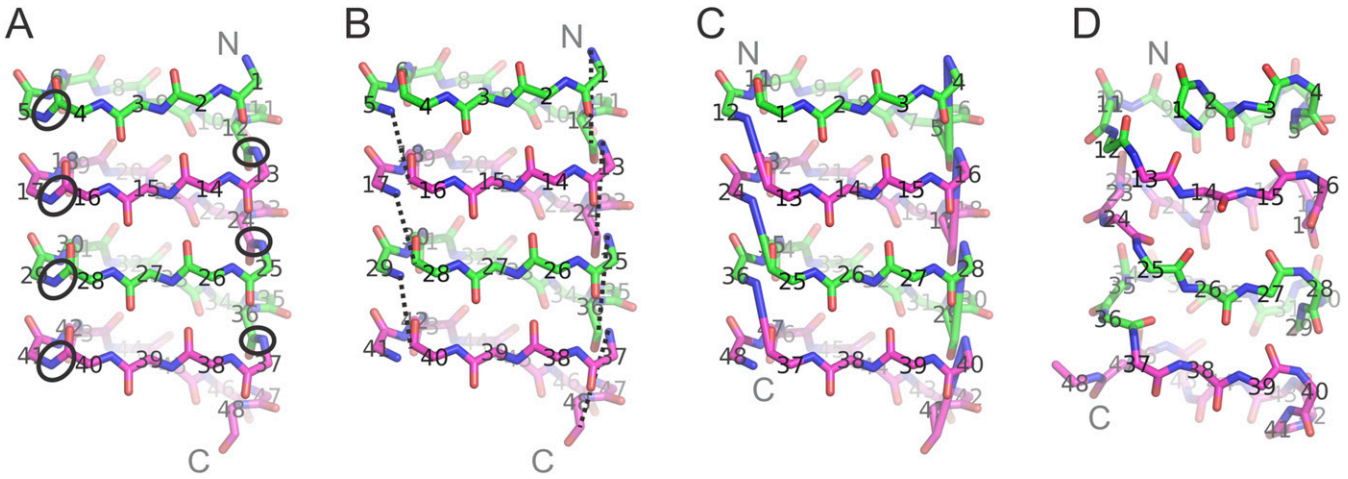
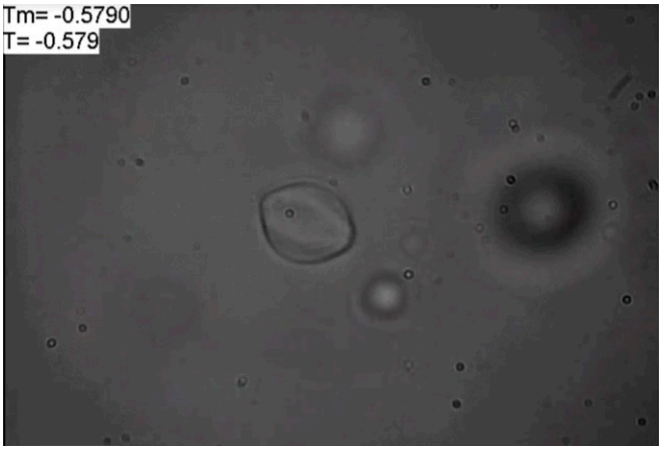



**Fig. S3.** Converting four coils of the *Tm*AFP structure (PDB ID code 1EZG) from a right-handed helix to a left-handed helix. Main chain coils are alternatively colored in green and magenta, and N and C termini are labeled in each panel in gray. (*A*) The bonds that were broken are circled. (*B*) The bonds that were built are indicated by dotted lines. (*C*) Left-handed structure after renumbering. (*D*) Left-handed structure after energy minimization.




**Movie S1.** Thermal hysteresis assay of midge AFP. A single ice crystal formed in the presence of midge AFP is monitored as the AFP solution is gradually cooled from the melting temperature to just below the freezing temperature.

Movie S1