BIOCHEMISTRY
including biophysical chemistry & molecular biology

ARTICLE

pubs.acs.org/biochemistry

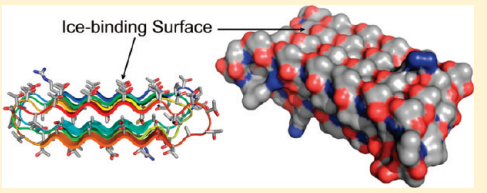
The Thr- and Ala-Rich Hyperactive Antifreeze Protein from Inchworm Folds as a Flat Silk-like β-Helix

Feng-Hsu Lin, Peter L. Davies,* and Laurie A. Graham

Department of Biochemistry, Queen's University, Kingston, ON K7L 3N6, Canada

Supporting Information

ABSTRACT: Inchworm larvae of the pale beauty geometer moth, *Campaea perlata*, exhibit strong (6.4 °C) freezing point depression activity, indicating the presence of hyperactive antifreeze proteins (AFPs). We have purified two novel Thr- and Ala-rich AFPs from the larvae as small (~3.5 kDa) and large (~8.3 kDa) variants and have cloned the cDNA sequences encoding both. They have no homology to known sequences in current BLAST databases. However, these proteins and the newly characterized AFP from the *Rhagium inquisitor* beetle both contain stretches rich in alternating Thr and Ala residues. On the basis of these repeats, as well as the discontinuities between them, a detailed structural model is proposed for the 8.3 kDa variant. This 88-residue protein is organized into an extended parallel-stranded β-helix with seven strands connected by classic β-turns. The alternating β-strands form two β-sheets with a thin core composed of interdigitating Ala and Ser residues, similar to the thin hydrophobic core proposed for some silks. The putative ice-binding face of the protein has a 4 × 5 regular array of Thr residues and is remarkably flat. In this regard, it resembles the nonhomologous Thr-rich AFPs from other moths and some beetles, which contain two longer rows of Thr in contrast to the five shorter rows in the inchworm protein. Like that of some other hyperactive AFPs, the spacing between these ice-binding Thr residues is a close match to the spacing of oxygen atoms on several planes of ice.



The ability of a protein to freely dissolve in aqueous solutions but then bind irreversibly to water in its solid state is the defining physical property of an antifreeze protein (AFP).[1] This unusual property has been illustrated, and exploited, with the development of ice affinity purification (IAP).[2,3] Via the slow growth of a mass of ice in a protein mixture, AFPs are successively bound and overgrown,[4] whereas non-AFPs (the vast majority of proteins) are excluded and concentrated in the liquid fraction. When using IAP to purify a recombinant AFP from *Escherichia coli* lysates, the fact that all endogenous *E. coli* proteins are excluded shows that the vast majority of proteins do not have an affinity for ice.[3] This and other variations of IAP have been used to purify recombinant AFP from plants, bacteria, and arthropods.[5−7]

AFPs are used by some animals (principally fish and insects) to prevent freezing.[8,9] Here the ability of the AFP to irreversibly adsorb to the surface of ice causes a nonequilibrium depression of the freezing temperature and a slight increase in the melting temperature[10] through the Gibbs−Thomson−Herring effect.[4] When the freezing temperature of the organism is depressed below that of the ambient temperature, the organism can resist freezing. In other organisms (principally plants and microorganisms), the AFPs help their host to tolerate freezing by suppressing the recrystallization of ice.[11] Because AFP is a misnomer in this situation where bulk freezing is not prevented, an alternative is to refer to all AFPs as ice-binding proteins (IBPs), although we continue to use the term AFP, where appropriate, throughout this work.[12]

The first IBP was described just more than 40 years ago.[13] Since then, at least a dozen remarkably different types have been described. Some are globular proteins with a mixture of secondary structures; others are repetitive proteins that are α-helical or β-helical or have polyproline type II helical structures.[14−16] Judging by the variety of IBPs and their scattered distribution in phyla, IBPs have independently evolved on numerous occasions, presumably as a mechanism by which their host could expand into a new niche or adapt to climate change.[17,18]

The mechanism by which IBPs bind irreversibly to ice has undergone several twists, turns, and reversals over the years and is still not well-defined. The original hydrogen bonding hypothesis[19] was largely disproved in favor of a mechanism that incorporated the hydrophobic effect.[20−22] The idea that constrained water is released on IBP binding to ice for a gain in entropy has been challenged by the idea that a hydrophobic ice-binding surface can organize surface-bound water into an icelike (clathrate) pattern that will fuse to ice.[23] Although this is an attractive hypothesis, the residence time of surface-bound water on the ice-binding site cannot be overly long and might be only marginally longer than that on other surfaces.[24]

The search for a unifying binding mechanism has been complicated by the variety of IBP structures and the many different ice planes they recognize.[25] Nevertheless, some common features and patterns are emerging. Ice-binding sites tend to be flat and relatively hydrophobic and occupy a large proportion

Received: March 1, 2011
Revised: April 11, 2011
Published: April 12, 2011

of the protein's surface area. We need to discover and characterize more IBP structures to elucidate a clearer pattern.

The application of IAP in combination with protein sequencing by tandem mass spectrometry (MS/MS) has greatly helped the search for new AFPs. A classic example of this approach is the discovery of the AFP in snow fleas (sfAFP). Two rounds of IAP applied to gram quantities of snow fleas produced pure sfAFP.[7] An amino acid composition and a mass were sufficient to identify the sequence of the abundant variant represented by four of 57 ESTs sequenced. Here we have applied a similar approach using the larvae of the pale beauty moth and have discovered and sequenced a novel AFP produced by this inchworm (iwAFP).

Some of the newly discovered AFPs are proving to be difficult to produce in sufficient quantities for structure determination. This is especially true of sfAFP,[7] AFP type Ih,[26] and the iwAFP reported here, which are thermolabile or difficult to fold. Therefore, we have invested considerable effort in deducing the structures of repetitive AFPs by an intuition-guided approach. To date, we have published three models of unique AFP folds: the IBP from perennial ryegrass, *Lp*AFP;[27] the AFP from snow fleas, sfAFP;[28] and the AFP from an Antarctic bacterium, *Mp*AFP.[29] Structures for sfAFP[16] and *Mp*AFP (C. Garnham, unpublished observations) are now available and have proven the models to be correct. Here we have modeled the newly discovered iwAFP as a novel β-helical fold that has some of the attributes of silk fibroin and have tentatively identified its ice-binding site as an unusually wide array of outward-pointing threonines on one face of the protein.

MATERIALS AND METHODS

Insect Collection. Inchworms (Figure 1a) were collected by hand during the month of March from 2003 through 2008 from snow-covered ground in various mixed deciduous woodlots within a 40 km radius of Kingston, ON. The larvae were stored on ice, sorted using a dissecting microscope, then frozen in liquid nitrogen, and stored at −80 °C.

Ice Affinity Purification. Larvae (2.5 g) were homogenized in 12 mL of ice-cold buffer [50 mM Tris-HCl (pH 7.9), 150 mM NaCl, 2 mM phenylthiocarbamide, 1 mM PMSF, 5 mM EDTA, and 2× complete, EDTA-free Protease Inhibitor Cocktail (Roche)]. Following a 30 min centrifugation at 12000*g*, the pellet was rehomogenized in an additional 12 mL of buffer and centrifuged again. The two supernatants were pooled and filtered through glass wool to remove lipid droplets (Figure 1b). The volume was increased to ∼100 mL using 50 mM NH_4HCO_3 (pH 8). Half the volume was gradually frozen around a coldfinger cooled from −0.8 to −2.4 °C over 36 h (Figure 1c). The surface of the ice was rinsed using several milliliters of ddH_2O to remove surface liquid. This ice fraction was then melted, and concentrated NH_4HCO_3 was added to a final concentration of 50 mM before a second round of IAP was performed in which half of the sample was frozen by being cooled from −0.8 to −2.4 °C over 43 h. Following rinsing and lyophilization of the ice mass, the dried residue was resuspended in 0.5 mL of 50 mM NH_4HCO_3. Amino acid compositional analysis was performed at the Advanced Protein Technology Centre (The Hospital for Sick Children, Toronto, ON). Thermal hysteresis (TH), used as a measure of antifreeze activity, was assayed using a nanoliter osmometer.[30]

Mass Spectrometry. Portions of the first and second ice fraction were subjected to MALDI MS with sinapinic acid using the Voyager DE-Pro instrument (Applied Biosystems) in linear mode. The same instrument was used in reflectron mode to obtain higher-resolution data for the lower-mass peptides. A portion of the second ice fraction was digested with porcine trypsin and subjected to MS/MS using a Waters qTOF ULTIMA GLOBAL mass spectrometer in electrospray mode. The spectrometer was equipped with a CapLC XE chromatography system using a Dionex C18 PepMap 100, 3 μm, 100 Å column. The mobile phase gradient was from 5 to 40% acetonitrile over 40 min. Data were processed using Water's MassLynx 4.0. Peaks were smoothed using the Savitzky Golay method with three channels and two smooths. The peaks were centroided using the top 80% of the peak height; +2 and +3 charge state peaks were deisotoped with a 1% noise threshold. These procedures were performed at the Protein Function Discovery

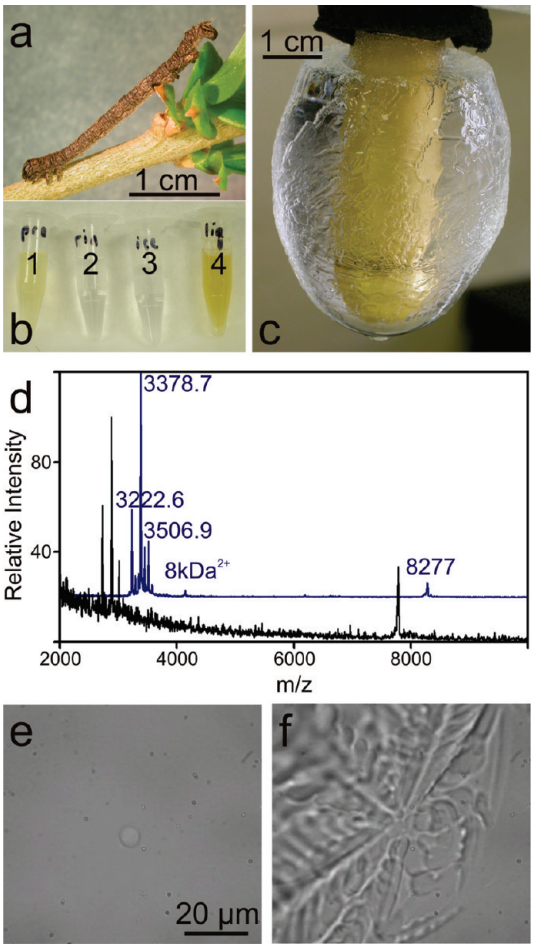


Figure 1. Purification of AFPs from inchworms. (a) Lab-reared *Campaea perlata* inchworm that was harvested in March. The inchworm shown here is approximately twice as long as those used directly from the field because it was reared to its final instar in the lab. (b) Fractions following one round of ice affinity purification (IAP): (1) the original diluted homogenate (pre), (2) material rinsed from the surface of the ice (rin), (3) the melted ice fraction (ice), and (4) the excluded liquid fraction (liq). (c) Faceted ice fraction following the first round of IAP. (d) Comparison of MALDI spectra of first- and second-round IAP fractions. For the sake of clarity, the first IAP spectrum is shifted left by m/z 500 and the second up 20%. (e) Morphology of a single ice crystal formed in the presence of dilute iwAFP. (f) "Burst" of this crystal immediately after the initiation of freezing. The *a*-axes lie within the plane of the image.

Facility at Queen's University. All MS/MS spectra were interpreted manually.

Circular Dichroism. Ice fraction 2 was dialyzed against 10 mM sodium phosphate buffer (pH 8), and CD spectra were recorded using an OLIS DSM1000 CD spectrophotometer and processed using OLIS Globalworks. Ten replicates of the raw CD spectra were collected for both the protein sample and the buffer alone, using the variable time per datum mode. The average spectrum for the protein sample was background corrected using the corresponding spectrum for the sample buffer, and the corrected spectrum was converted to molar ellipticity units using a mean residue molecular mass of 98 Da and a protein concentration of 24 μg/mL, determined by amino acid analysis (above).

cDNA Library Construction and EST Sequencing. Messenger RNA was extracted from 116 mg of inchworms (~20 individuals) by using the RNeasy Midi Kit followed by the Oligotex mRNA Midi Kit (Qiagen), yielding 2 μg of mRNA. A cDNA library [primary titer of 1.8×10^4 plaque forming units (pfu)] was constructed in λ-Zap II (Stratagene, which is now Agilent Technologies). Variable amounts (1−10 μL) of the amplified library (>10^4 pfu/μL) were mass excised, and then XL1-Blue MRF′ rather than SOLR cells were infected with aliquots of this phagemid stock. Following the addition of glycerol to a final concentration of 15%, the cell suspensions were frozen and shipped to Génome Québec Innovation Centre (McGill University, Montrèal, QC) where individual colonies were robotically selected from plates and subjected to high-throughput EST sequencing. All procedures were conducted according to the manufacturer's instructions except as otherwise mentioned.

AFP Gene Identification and Sequence Analysis. EST sequences were trimmed using a Phred quality score of 40, and individual bases with scores below 15 were masked and vector sequences removed. All accession numbers refer to the GenBank DNA sequence database. NCBI BLAST version 2.2.18 was run,[31] in house, with the EST sequences as the database. pBluescript (X52330) and lambda (J02459) sequences were used to detect contaminants. The sequences of two complete moth mitochondrial genomes, *Bombyx mori* (NC_002355) and the more closely related geometer moth, *Phthonandria atrilineata* (NC_010522), were used to detect mitochondrial inserts. The identity of the inchworms was determined by searching the Barcode of Life Data System (BOLD) database with EST sequences corresponding to the cytochrome *c* oxidase subunit I gene.[32] AFP sequences in the EST database were initially identified using TBLASTN with the MS/MS sequences as the query. Individual EST sequences were then used to detect multiple occurrences of sequences of interest using BLASTN. On-line BLAST searches at the NCBI Website were applied for most other searches. Sequence manipulations and comparisons were primarily conducted using DNAman version 6 (Lynnon Corp., Pointe-Claire, QC). ESTs are named according to plate number and position.

Polymerase Chain Reaction (PCR) Amplification and Cloning of AFP Sequences. The following primers were designed to amplify the 5′ region of the cDNAs for both variants by anchor PCR, 5′-AACAGGATATCTGCACTCGCTCTC-3′ (8 kDa variant) and 5′-GTTGCTATCTTCTGTCAGCTTCACG-3′ (3 kDa variant). These were used in combination with the T3 promoter PCR primer, 5′-CCCTCACTAAAGGGAACAAA-AGCTG-3′, in which 10 μL of a 1:1000 dilution of the cDNA library stock mentioned above was used as the template in 50 μL reaction mixtures. The same reaction conditions were employed for all PCRs unless otherwise mentioned. *Taq* DNA polymerase with 1.5 mM $MgCl_2$ and buffer with $(NH_4)_2SO_4$ (Fermentas) was used with the following cycling parameters: 5 min at 95 °C, 30 cycles of 1 min at 60 °C, 1 min at 72 °C, and 1 min at 95 °C, and a final extension of 20 min at 72 °C. Products were cloned using the TOPO TA Cloning Kit (Invitrogen) and sequenced. Two new primers were designed, 5′-CCAAGTTGTGAA-CATGTGGGCAC-3′ (8 kDa variant) and 5′-CGAGGCTG-AGTGATAGTTATGTGGA-3′ (3 kDa variant), which were used with the T7 promoter PCR primer, 5′-ACGACTCACTA-TAGGGCGAATTGG-3′, to amplify entire coding sequences. These were cloned as described above and sequenced. Sequences obtained by PCR are labeled FL (for full length) and numbered.

Recombinant Expression. The DNA sequence encoding the mature 8275 Da AFP was codon-optimized and synthesized by GeneArt (Regensburg, Germany). This synthetic DNA was inserted into the pET 28a vector to form a His-tagged construct that was expressed in *E. coli* BL21(DE3) cells at 16 °C. Pelleted cells from a 1 L expression were resuspended in 150 mM NaCl, 3 mM EDTA, 1 mM PMSF, and 50 mM Tris-HCl (pH 7.4) prior to sonication. After centrifugation, the cleared supernatant was subjected to nickel affinity chromatography. The eluant was concentrated to 50 μL and tested for TH.

Molecular Dynamics. The model for iwAFP was manually built using PyMOL 1.2r1.[33] The model was solvated in a 4.2 nm × 3.4 nm × 2.3 nm box containing 3202 water molecules. Eight sodium ions and seven chloride ions were added to neutralize the system charge and achieve a salt concentration of 0.1 M. The system was then subjected to energy minimization, position-restrained molecular dynamics to settle the water molecules, and unrestrained molecular dynamics simulations using GROMACS 4.04.[34] The simulations were performed at 4 °C using a time step of 2 fs for a total duration of 10 ns. The program uses a triclinic unit cell for its periodic boundary conditions. For calculation of short-range nonbonded interactions, only the nearest image was considered. Long-range electrostatics was treated with the particle mesh Ewald method. The simulations were conducted under isothermal conditions using V-rescale temperature coupling. The GROMOS96 43a1 force field was employed.

RESULTS

Protein Purification and Activity. A homogenate of inchworms (Figure 1a) at a 1:1 (w/v) ratio showed 6.4 °C of TH activity (freezing point depression). A diluted homogenate of 2.5 g of larvae was subjected to two rounds of ice affinity purification. After just a single round, the ice fraction (IAP1) was visibly clear (tube 3) and the bulk of the starting material was concentrated in the liquid fraction (tube 4) (Figure 1b). The growing ice was heavily faceted (Figure 1c). After this single round of IAP, there was already significant enrichment for 3−4 kDa (small) and 8.3 kDa (large) components (Figure 1d, bottom trace, which is shifted m/z 500 to the left for the sake of clarity). After the second round (top trace, which is shifted 20% upward for the sake of clarity), the background in the mass spectrograph was significantly decreased and only one large and several small components remained.

When the second ice fraction (IAP2) was concentrated, it showed 6.4 °C of TH activity at a concentration of 24 μg/mL. The ice crystals obtained during melting of such high-activity samples are very small, and the burst (onset of freezing) is extremely rapid. Therefore, a sample with a lower activity (2.55 °C)

Table 1. Comparison of the Amino Acid Composition (mole percent) of the Second Ice Affinity-Purified Fraction (IAP2) with Several Variants Deduced from cDNA Sequences (Figures 2 and 3)

amino acid	IAP2	small (FL#1)	large (FL#2)	large (FL#5)	large (P06_O24)	large (P01_F03)
Asx	6.7	5.9	4.6	4.6	5.7	4.5
Glx	7.4	0	1.1	1.1	2.3	2.3
Ser	9.7	5.9	12.5	11.4	11.4	11.4
Gly	10.4	2.9	2.3	2.3	2.3	2.3
His	1.0	2.9	0	0	0	0
Arg	3.8	8.8	2.3	2.3	0	3.4
Thr	20.5	38.2	38.6	38.6	34.1	36.4
Ala	17.8	26.5	22.7	23.9	27.3	25
Pro	2.0	0	2.3	2.3	2.3	1.1
Tyr	0.6	0[a]	0	0	1.1	1.1
Val	7.9	2.9	9.1	9.1	8	6.8
Met	4.3	0	0	0	1.1	0
Cys	0.4	0	1.1	1.1	0	0
Ile	2.7	0	2.3	2.3	2.3	1.1
Leu	3.2	2.9	1.1	1.1	2.3	3.4
Phe	1.0	0	0	0	0	1.1
Lys	0.5	2.9	0	0	0	0
total	100.0	99.8	100.0	100.1	100.2	99.9

[a] The three residues absent in the two variants are found in some of the other large isoforms.

was imaged. The ice crystals were shaped like hexagonal plates with slightly outward-curving surfaces, and this form was attained during melting (Figure 1e). The shape did not change during cooling within the thermal hysteresis gap (not shown). Once the nonequilibrium freezing point was exceeded, dendritic growth occurred in the general direction of the *a*-axes (Figure 1f), which is asymmetric in this image as it was initiated from the top left quadrant of the crystal (not shown). This activity was completely lost after 2 h at room temperature (22 °C), indicating that the AFPs are thermolabile.

The amino acid composition of the IAP2 fraction (Table 1) was highly biased toward small side chain amino acids, particularly Ala and Thr at 18 and 21%, respectively. Other AFPs deficient in larger amino acids, such as *Tm*AFP and sfAFP, migrate anomalously on gels and stain poorly with both Coomassie blue[35,36] and silver stain (not shown). Because of this experience, and because the amount of sample was limited and the AFP did not react positively with Bradford reagent, gel electrophoresis was not attempted. Instead, MALDI MS was used to determine the purity and complexity of the ice-binding components.

EST Library. Because the mRNAs encoding AFPs are typically abundant during winter, a cDNA library was constructed from inchworms gathered in March. An EST sequencing approach was undertaken in an attempt to identify the AFP(s). A total of 2356 ESTs were sequenced. Of these, 10% failed or produced poor quality sequence, 14% did not contain an insert, 2% encoded mitochondrial transcripts, and a few contained cloning artifacts. It is difficult to identify geometer moths at the larval stage, but six ESTs encoding the mitochondrial cox I protein were recovered and used to search the BOLD database.[32] These ESTs were exact or >99% matches to the barcode sequences of *C. perlata* specimens, allowing a conclusive identification of the species. Multiple copies of individual sequences were present among the 1724 remaining sequences, but none encoded predicted proteins matching the mass or composition of the AFPs determined above.

Mass Spectrometry and Sequence Characterization. To obtain more information about the AFPs, the IAP2 fraction was digested with trypsin and subjected to MS/MS. A variety of peptide sequences were obtained that were rich in Thr and Ala residues, some with stretches in which these residues alternated (Figures 2 and 3). This was consistent with the biased amino acid composition (Table 1). Matching sequences were found within the EST sequences, but surprisingly, they were all located near the C-terminal end of two sequences predicted to encode ~40 kDa proteins (Figure S1 of the Supporting Information). Of the 1724 ESTs presumed to originate from nuclear transcripts, 12 encoded large (~8 kDa) AFPs and five encoded the small (3−4 kDa) AFP. This corresponds to 0.7 and 0.3% of the transcripts, respectively. Anchor PCR, using a combination of vector and AFP specific primers, generated a full-length coding sequence for the small AFP and additional sequences encoding isoforms of the large variant.

The complete cDNA sequences encode proteins predicted to begin with an 18-residue secretory signal peptide that is highly conserved (15 of 18 identical residues) between the small and large AFPs (Figure S2 of the Supporting Information). This is followed by a region of unknown function that is 28−36 kDa, nonrepetitive, and of fairly average amino acid composition and has no similarity to known sequences. Despite the presence of many basic residues in these regions that would yield fragment sizes suitable for MS/MS sequencing, no such fragments were observed in the IAP2 fraction. This N-terminal region is longer in the sequence encoding the small AFP and is 47% identical to the large AFP across the overlapping region.

As mentioned above, the MS/MS sequences and tryptic fragment masses that were observed corresponded only to the C-terminal portions of the coding sequences. In the case of the small AFP, the four sequences obtained encode identical mature AFPs (data not shown). The mass variability observed in the 3−4 kDa range (Figure 1d) appears to be caused primarily by

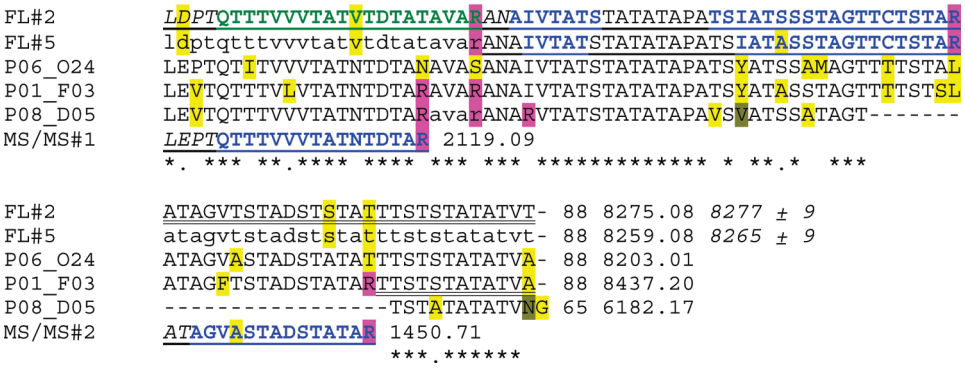


Figure 2. Alignment of five mature large AFP isoforms deduced from cDNA sequences. Differences between sequences are highlighted in yellow, dark yellow, or pink (Arg residues). The number of residues and the predicted masses of the isoforms (daltons) are indicated at their C-termini. The corresponding masses observed by MALDI MS are in italics. Tryptic fragments that were partially sequenced by MS/MS are underlined; conclusively determined residues are colored blue (or green in the case of adjacent fragments) in boldface. Italics indicate residue pairs consistent with either the mass of the observed b2 ions or the abundant ion generated by cleavage on the N-terminal side of Pro. Fragments shown in lowercase are identical to those in another isoform. Double underlining indicates fragments that were not sequenced but for which a consistent mass was observed in the whole digest ion spectrum. The sequence and mass of two tryptic fragments (MS/MS#1 and MS/MS#2) that contain polymorphisms relative to the known sequences are also shown. cDNA clones obtained by PCR are numbered and denoted FL (for full-length), whereas sequences obtained by sequencing of ESTs are named by plate number and position. Conserved residues are denoted with asterisks and conservative substitutions with periods. Sequences have been deposited in GenBank with the following accession numbers: JF436943 for FL#2, JF436944 for FL#5, JF436940 for P06_O24, JF436941 for P01_F03, and JF436942 for P08_D05.

Figure 3. Sequences of the single small AFP deduced from cDNA. The four different isoforms (#1−#4) were generated by variable processing of the precursor (GenBank entry JF436945) at the two termini. Arg and Lys residues are highlighted in pink, and the number of residues and the predicted masses (daltons) are indicated at the end of each sequence. The corresponding mass observed by MALDI MS is in italics. Tryptic fragments that were partially or wholly sequenced by MS/MS are underlined with conclusively determined residues colored blue (or green in the case of adjacent fragments) and in boldface. Nonunderlined residues were identified by MS/MS of the undigested AFP. Italics indicate residue pairs consistent with either the mass of the observed b2 or y2 ions.

variable processing of the mature protein at both the N- and C-termini (Figure 3). The most abundant isoform (3379 Da) begins with two Thr residues and ends with an Arg residue. The 3584 Da isoform includes two additional residues at the N-terminus, Ile and Ser, which is well supported by the mass and MS/MS sequence of a tryptic fragment. The 3223 Da isoform is consistent with loss of the C-terminal Arg and is also well-supported by MS/MS data. The third isoform (3507 Da) is predicted by whole mass alone to contain an additional C-terminal Lys residue beyond the penultimate Arg. A C-terminal 22−23-amino acid portion that is also removed to generate the AFP has an amino acid bias different from that of the mature sequence, including several aromatic residues (Figure S1 of the Supporting Information).

Unlike the small AFP, the large AFP is not processed at the C-terminal end, and sequence polymorphisms, rather than processing differences, were observed. A single cleavage site was observed at the N-terminus, immediately following two Phe residues (Figure S1 of the Supporting Information). The C-terminal tryptic fragment lacked a basic residue and was not sequenced by MS/MS, but fragments matching the expected masses based on isoform FL#2 and P01_F03 (Figure 2) were observed. A total of five unique mature AFP sequences (Figure 2) were encoded by 11 cDNA sequences (not shown). Whole masses and MS/MS sequences corresponding to two (FL#2 and FL#5) were observed. Isoform P06_O24 would not generate any fragments as it is devoid of basic residues. Two additional MS/MS sequences that differed from a known sequence at a single position were obtained. One sequence encoded an isoform containing a deletion of 24 amino acids, but a corresponding mass was not observed by MALDI MS.

Expression of the Large AFP. To conclusively demonstrate that the sequences characterized and cloned were responsible for thermal hysteresis activity in the inchworm, the AFP-encoding portion of the most abundant large isoform (FL#2) was codon optimized and expressed in *E. coli*. Both the soluble and insoluble fractions were analyzed by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) following sonication of the cell pellet, but a new protein band was not seen, which was not unexpected given the biased amino acid composition of this protein. Following nickel affinity chromatography, the partially purified fraction showed 0.48 °C of TH activity. Such low activity was not unexpected given the thermolability of iwAFP and its low yield.

Sequence Characteristics and Circular Dichroism. The AFP domains of the large and small variants cannot be aligned with any certainty because of their differing lengths and repetitive nature. Nevertheless, the two AFPs are clearly homologous given the sequence identity (47%) between their cleaved N-terminal domains, as mentioned above. The two mature AFP sequences

Table 2. Bias in Amino Acid Content by Position in the Large and Small iwAFP[a]

	large (FL#2)		small (#1)	
amino acid	position 1	position 2	position 1	position 2
Thr	70	7	59	18
Ala		45	6	47
Ser	2	23	6	6
Val	5	14		6
Asp	5	2	12	
Gly	5			6
Ile	5			
Pro	2	2		
Arg	5		6	12
Leu		2	6	
Gln		2		
Asn	2			
Cys		2		
Lys			6	
His				6
total	101	99	101	101

[a] Each AFP was divided into sequential dipeptides, with the first residue in each dipeptide denoted as position 1. The percent abundance of each amino acid at each position was calculated. The register of the dipeptides was different between the two isoforms, beginning with the first residue of the small isoform and with the second residue of the large isoform. Zeros have been omitted for the sake of clarity.

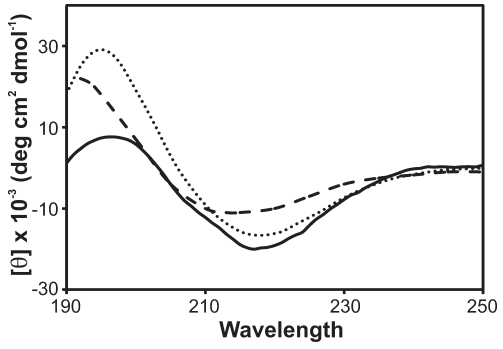


Figure 4. CD spectrum of the second IAP fraction. The spectrum at 4 °C of iwAFP after exclusion of other proteins and solutes by two rounds of IAP (—) is compared to those from two other β-helical AFPs: sbwAFP (– – –) and region IV of *Mp*AFP (· · ·).[29,37] All spectra exhibited high β-strand content with minima close to 220 nm.

do share a similar amino acid composition and patterning, however. If they are simply broken into amino acid pairs, a distinct dipeptide bias is observed (Table 2). Thr is the predominant residue at the first position, whereas Ala is predominant at the second. The other residues prevalent at the second position tend to be small and neutral, such as Ser and Val in the large AFP. Therefore, both variants are likely to adopt similar secondary structures.

A likely secondary structure adopted by a protein with a dipeptide repeat pattern would be a β-strand. To test this hypothesis, the CD spectrum of the IAP2 fraction, which contains both the large and small AFPs, was obtained (Figure 4). It is

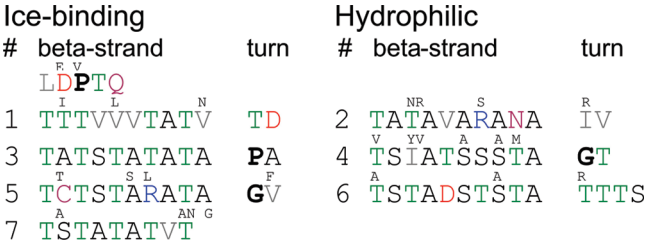


Figure 5. Sequence of the large iwAFP isoform, FL#2. The sequence is arranged as β-strand segments based on the Thr-Ala/Ser dipeptide repeat pattern. Outward-pointing residues are colored in gray. The putative ice-binding face, which is dominated by outward-facing Thr residues, and the hydrophilic face are labeled as such. Regions in which this Thr-Ala/Ser pattern is disrupted, often by Pro or Gly, are shown at the right. Polymorphisms present in other isoforms are shown in smaller font above the variable residue.

similar to the spectra of β-strand-rich proteins with a single minimum at 217 nm and a maximum at 196 nm and is qualitatively similar to the spectra of two nonhomologous AFPs with β-helical structures.[29,37]

Modeling the Structure of the Large iwAFP Variant. In addition to the dipeptide pattern and CD spectrum mentioned above, the sequence of the large AFP provided additional clues that facilitated the development of a structural model. First, other Thr-rich AFPs display rows of Thr residues on the outer surface of β-strands,[15,38] forming a Thr-X-Thr motif. A similar motif, where X is usually Ala, Ser, or Val, is observed in iwAFP (Figure 2). However, here it is extended to have as many as five Thr residues in a row (Thr-X-Thr-X-Thr-X-Thr-X-Thr), rather than occurring just once at regular intervals, as seen with the 12- or 13-amino acid spacing found in the beetle AFP from *Tenebrio molitor* (*Tm*AFP) and *Dendroides canadensis* (*Dc*AFP).[39−42] Second, this sequential pattern showed six periodic deviations at 12-amino acid intervals throughout the length of the protein (Figure 5), allowing the sequence to be divided into seven segments. These deviations that break up the Thr-X-Thr-X-Thr-X-Thr-X-Thr tracts are typically two amino acids long, and three of them contain either Pro or Gly, which would be expected to disrupt secondary structure. Third, the shortest sequence (P08_D05) contains an internal deletion of 24 residues, which would precisely remove two segments. Therefore, we concluded that the protein is likely composed of seven β-strands connected by turns, with outward-facing Thr residues clustered on one face of the protein as ice-binding residues. β-Strands often contain alternating hydrophilic and hydrophobic residues, with the hydrophobic residues facing inward. In this case, the first position, occupied by the putative ice-binding Thr residues, is more hydrophilic than the second position, which is occupied by Ala, Ser, and Val. Also, all of the larger residues found within the seven segments, such as Asn, Ile, and especially Arg, are found at the first (exterior) position. Only small residues, including a single Cys and Thr in addition to Ala, Ser, and Val, are found at the second (interior) position. Additionally, although each position has eight or nine polymorphic sites (Figure 5), Ser/Ala changes are found at five of these in the second position, while the most drastic change is a Thr to Ile substitution in strand 1. The changes in the first position of the dipeptide repeat are more extreme in character, including Thr to Met or Arg, Arg to Ser, and Ile to Tyr mutations. These observations support the notion that the residues at the second position face inward as they are more

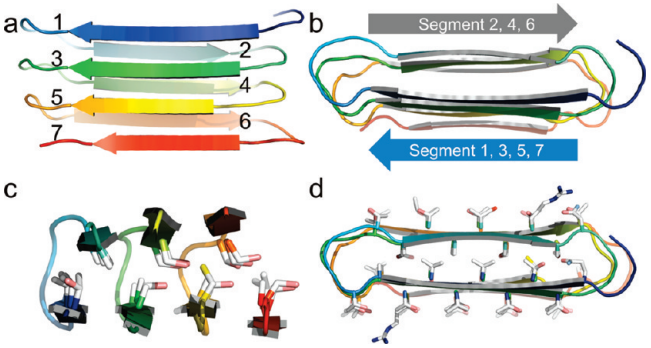


Figure 6. Predicted iwAFP fold. The backbone is represented in cartoon format with color changing from blue to orange from the N-terminus to the C-terminus. (a) Folding scheme looking down at the putative ice-binding face. (b) Folding scheme with the N-terminal end forward, looking down the axis of the β-helix with the ice-binding face pointing downward and the hydrophilic face pointing upward. (c) View into the core with the model rotated 90° clockwise in the horizontal plane relative to panel b. The rightward tilt of the β-carbons of the interdigitating core residues can been seen from this angle. (d) Same view as panel b showing the interdigitation of the core residues rotated 90° relative to panel c. The side chains of the residues on the hydrophilic (top) and ice-binding (bottom) surfaces are also shown.

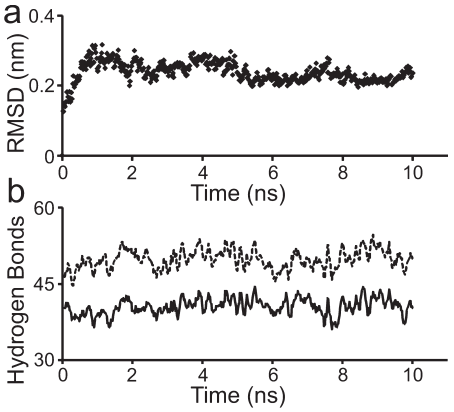


Figure 7. Stability of the iwAFP model. (a) In this molecular dynamics analysis, the RMSD of the backbone atoms is plotted as a function of time. (b) Hydrogen bond count as a function of time during the molecular dynamics simulation. Intramolecular hydrogen bonds were assigned during model building and then counted during the course of the simulation. The solid line represents the number of hydrogen bonds formed between backbone atoms. The dashed line also includes hydrogen bonds formed to the core Ser side chains. Data were smoothed using a moving average across five data points spanning 0.1 ns.

hydrophobic and, because of their uniform size, are less likely to disrupt packing of the core.

To form a stable β-strand structure with position 2 facing inward, at least two β-sheets would be required. Because there are only seven strands, multilayer stacking is not likely. Also, because the seven segments are quite regular and repetitive, and the regions with the deviations are short, the segments likely fold into a simple, double-layer, parallel or antiparallel arrangement.

We argue that in inchworm AFP, the fold should be parallel for a number of reasons. First, once the seven strands were constructed, we noticed that the Thr residues on the four odd-numbered segments were better conserved. Only two positions show variation, with a Val in the first segment and an Arg in the fifth. The three even-numbered segments contained six variations. Therefore, the odd-numbered segments should be on one side of the protein, whereas the even-numbered segments should be on the other side, which can only be done with a parallel β-strand fold (Figure 6a,b). Second, when the polymorphisms were examined (Figures 2 and 4), six were observed on the outward-facing residues of the even-numbered strands and only two on the odd-numbered strands. Third, the parallel β-helix is a common tertiary structure for AFPs that is adopted by four mutually nonhomologous repetitive β-rich proteins, *Tm*AFP, *Mp*AFP, sbwAFP, and *Lp*AFP.[15,27,29,38] Fourth, the very regular spacing of water molecules on the ice surface would be best matched by a regular spacing of ice-binding residues.

The last parameter to be addressed is the handedness of the structure. A right-handed model was constructed because the inward-facing Val residues on the first strand appear to pack best in this orientation. The methyl groups do not appear to pack favorably if they are oriented toward the core. When they are oriented toward the solvent, they are less exposed in the right-handed model as the β-carbon is angled toward the core (Figure 6c). If this conjecture is incorrect, it has minimal bearing on the salient features of the model as the core would be little changed and the surfaces of the sheets would essentially be mirror images of one another.

Molecular Dynamics. The credibility of the 8 kDa variant model was tested by subjecting it to a molecular dynamics simulation. Because iwAFP functions at low temperatures and we know the protein is thermolabile, the simulation was performed at 4 °C. Because it is unlikely that the ϕ and ψ angles of a manually constructed model equate to those of the actual structure, the initial stage of the simulation gives the model a chance to adjust these to more realistic values. Subsequently, the model should stabilize if the overall architecture is sound. This was indeed the case as indicated by the overall fluctuation over 10 ns that is represented by the root-mean-square deviation (RMSD) of all backbone atoms from the initial structure (Figure 7a). Using the initial frame as a reference, the RMSD increased rapidly to 0.3 nm in the first nanosecond. The RMSD decreased gradually to 0.22 nm and stabilized by 5 ns. In addition, the number of intramolecular hydrogen bonds was expected to be large because it is one of the stabilizing forces of a β-helical fold. Therefore, it was necessary to examine the hydrogen bonding within the model. Two groups of hydrogen donors and receptors were selected for this analysis: the peptide backbone and Ser residue side chains within the core. The number of hydrogen bonds between backbone atoms fluctuated somewhat but did not change significantly, suggesting that the β-helix integrity was well maintained through the course of the simulation (Figure 7b). When the Ser residues were also considered, the total number of hydrogen bonds remained consistently higher than the previous value, suggesting that these core Ser residues also help maintain the structure by forming hydrogen bonds with main chain atoms. The overall structure remained largely unchanged during the simulation with only minor fluctuations in the turns (Figure S3 of the Supporting Information). The only exception was the N-terminal sequence, LDP, which appeared to be very flexible.

Characteristics of the Modeled 8 kDa AFP Structure. The core of the structural model (Figure 6c,d) is formed by interdigitation of the inward-facing Ala/Ser residues, giving a rather short distance between the backbone atoms of the two sheets (6.5 Å). Unlike models of the type II (crystalline) silk from

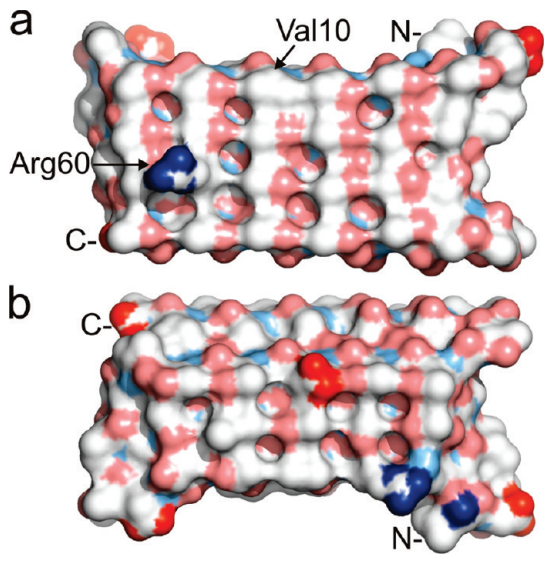


Figure 8. Surface representation of the iwAFP model. Atoms are colored pink (uncharged oxygen), red (negatively charged oxygen), light blue (uncharged nitrogen), blue (charged nitrogen), and white (other). The N- and C-termini are indicated by N- and C-, respectively: (a) putative ice-binding face and (b) non-ice-binding face.

B. mori,[43] the sheets are formed by parallel rather than antiparallel strands, forming a β-helix. The AFP is shaped like a rectangular prism with the two larger surfaces formed by the opposing β-sheets having dimensions of approximately 18 Å × 45 Å and with a thickness of approximately 13 Å (Figure 8). The larger face (Figure 8a), composed of the four odd-numbered strands, displays a 4 × 5 array of side chains consisting almost entirely of Thr, which, because of its flatness and Thr content, is presumed to be the ice-binding face. Of the two residues that are not Thr, Val10 is similar in size to Thr. Arg60 is a much larger residue, which is substituted with Leu in two isoforms, but is located near one corner of the ice-binding face where its flexibility may allow it to angle away to avoid steric interference with ice binding. The smaller face (Figure 8b) consists of the three even-numbered strands; for 8 of 14, the outward-facing residues are also Thr residues. However, this surface is more hydrophilic and noticeably less flat because of the presence of residues with charged and/or longer side chains, such as Arg, Asp, and Ile. The first five residues (LDPTQ) were modeled to fold over the N-terminal end of the β-helix to form a cap. However, ~8 ns into the molecular dynamics simulation, the LDP portion moved away from this position to form hydrophobic contacts with the two residues (I and V) in the turn linking strands 2 and 3, which contributed to a high root-mean-square fluctuation (Figure S3 of the Supporting Information). The C-terminal end of the β-helix is stabilized by the inward-pointing residues of segments 6 and 7, which are a mixture of Ala and Ser. The hydroxyl groups of these Ser residues are close enough to form hydrogen bonds with the neighboring backbone or other Ser residues. These extra hydrogen bonds from Ser, in combination with the backbone hydrogen bonds, lock the C-terminal backbone in place.

DISCUSSION

Discovery of New Antifreeze Proteins. Ice-binding proteins (IBPs) are probably the most structurally diverse group of proteins that share a common function, adsorbing to ice. One of the objectives in discovering new IBPs is to find common attributes within this diversity that will elucidate the mechanism by which they bind to ice. The identification of new IBPs has become much easier since the introduction of ice affinity purification (IAP). This is a particularly important methodology for organisms caught in the wild, such as these inchworms, which were not plentiful, and the tiny snow fleas used in a previous study.[7] We were able to start with just gram quantities of animals and obtain a fraction highly enriched for AFPs after only two rounds of IAP that, considering the starting material was a crude whole-body homogenate, is a graphic demonstration of the ability of IAP to separate proteins that bind to ice from those that do not.

A second property of IBPs that we intended to exploit is the fact that their mRNAs are often present at very high levels in the cold.[44] We reasoned that the sequences would be well represented in a cDNA library and so undertook a limited EST sequencing project. In the snow flea example, the sequence of the small AFP was identified simply by matching a mass and amino acid composition to the most abundant transcript that represented 7% of the ESTs. However, this straightforward approach was unsuccessful with iwAFP for two reasons. First, although the sequences contained a secretory signal peptide as expected, they also contained a large, unrecognizable region (>80% of the coding sequence) of unknown function that is absent from the mature AFP. Therefore, none of the predicted sequences matched the observed mass or composition of the processed AFP. Second, the transcripts were less abundant than expected, representing <1% of the complete EST library, and comparable in abundance to many other sequences. Therefore, MS/MS sequencing was essential in identifying the correct ESTs, while providing additional information such as the variable processing observed in the small variant.

As secreted proteins, IBPs are subject to at least some post-translational modification such as removal of the signal peptide and disulfide bond formation. Two fish AFPs, type II from sea raven[45] and type I from winter flounder,[46] also contain removable pro-sequences. The convergently evolved antifreeze glycoproteins of northern cod[12] and Antarctic cod[47] are synthesized as polyproteins that are cleaved at different amino acid spacer sequences of four and three residues, respectively. The proteolytic mechanism behind processing of the iwAFP is not apparent given the variable cleavage sites (FF|LD, VA|IS|TT, and SS|R|K|AL), but the cleavage may be occurring in an exposed linker between pro-domains and mature domains. It is also not clear why the pro-region needs to be removed. Additional domains are not an impediment to ice binding as shown by fusion IBPs (AFP-GFP and AFP-calpain) being efficiently recovered by IAP.[48,49] They even increase the TH activity of the fusion protein,[50] unless of course the other domain occludes the ice-binding site.

An additional benefit of sequencing a moderate number of ESTs is that most AFPs are members of multigene families.[17] In some cases, homologous AFP variants of different sizes lack sufficient sequence similarity for their cDNAs to cross-hybridize.[7,51] This was also the case with inchworm, and by sequencing ~2000 ESTs, we recovered 15 sequences encoding the two variants. An additional benefit was that the isoforms of the large variant provided important clues to assist in the modeling of its structure.

Conclusive proof that the AFP was correctly identified was obtained when the portion of the cDNA encoding a mature large isoform was expressed in *E. coli*. Ice shaping and 0.48 °C of

antifreeze activity were observed, sufficient to demonstrate that the sequence encodes an AFP. Not unexpectedly, the recombinant protein was not observed via SDS−PAGE, probably because of the lack of Coomassie blue-binding residues. A similar problem was encountered with *Tm*AFP, another Thr-rich AFP with a bias toward smaller residues.[52] However, the poor activity suggested that the protein either did not express well in *E. coli* or was not well folded, even when expressed at 16 °C, because of its thermolability. This approach did not afford us sufficient material with which to pursue structural studies.

Modeling. The repetitive nature of the sequence and the information provided by the different isoforms facilitated modeling the structure of the large iwAFP variant. Both the dipeptide repeat pattern and CD data suggested that the predominant secondary structure was β-strand. Thr-Ala/Ser repeats were not observed in the structural database, but more than 70% of the 391 kDa *B. mori* silk fibroin heavy chain is composed of Gly-Ala/Ser repeats, most commonly Gly-Ala-Gly-Ala-Gly-Ser.[53] When this protein undergoes the transition to the crystalline form (hereafter called silk), it adopts a β-sheet structure with a rectangular unit cell recently refined to be 9.38, 6.98, and 9.49 Å by X-ray fiber diffraction.[54] The first value is thought to represent twice the spacing between adjacent strands and the second the repeat distance along a strand (the distance separating every second residue). The third is the sum of the distances across pairs of stacked strands. To achieve this spacing, opposing strands are thought to be staggered in two dimensions. Residues are not opposite one another but are offset by half a residue, and the strands are also staggered rather than directly opposite one another. This allows the core Ala/Ser side chains to interdigitate, fitting into the empty space between the side chains on the opposing strands. One model of this structure is available (Figure S4a of the Supporting Information).[55] We have adopted a similar arrangement for the Ala-rich core of the iwAFP (Figure 6c,d and Figure S4b of the Supporting Information), which is tightly packed and very regular because of the complete absence of larger residues within the central strands.

The nature of the sequence of iwAFP suggests that a parallel β-helix is the likeliest arrangement. Although silk is generally thought to be antiparallel, in part, on the basis of fiber diffraction studies of crystallized synthetic peptides,[56] recent experiments using synthetic poly-Ala,[57] fiber diffraction and modeling,[54] and closer examination of the sequence[53] have called the original model into question. If silk is parallel, then the distances measured are even more likely to be relevant to iwAFP and will be discussed in more detail later.

The β-helix is turning out to be a recurring tertiary structure adopted by a number of mutually nonhomologous AFPs. *Ri*AFP, the recently discovered beetle antifreeze protein, has also been suggested to form a β-helix.[58] Both *Tm*AFP and sbwAFP,[15,38] as well as the model of *Lp*AFP,[27] are tightly wound parallel β-helices with 12−15 residues per loop (Figure S5 of the Supporting Information). At 24 residues, the loops in iwAFP are significantly longer. The identified ice-binding faces of the *Tm*AFP and sbwAFPs are found on a narrow parallel β-sheet that occupies one side of the protein (Figure 9a,b). The bacterial antifreeze protein, *Mp*AFP, is a β-helix with a loop of intermediate length, typically containing 19 residues.[29] Its ice-binding face is located on the Ca^{2+}-stabilized turns of the protein that produce parallel rows of Thr and Asn/Asp residues. In each of these four published examples, the tight winding of the helical coil or the position of the ice-binding face is such that its width is limited to

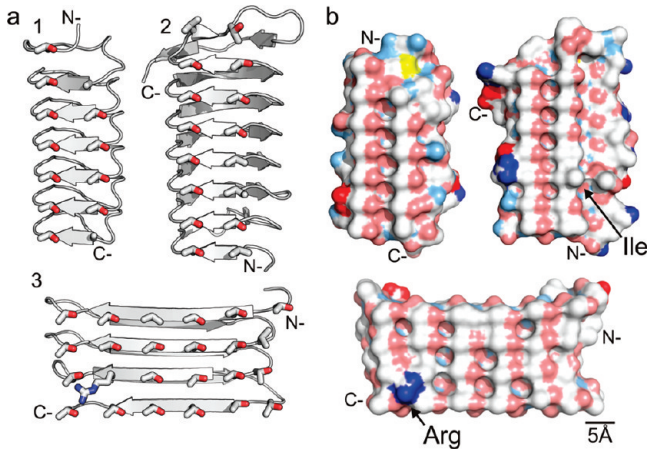


Figure 9. Distribution of residues on the Thr-rich ice-binding surfaces of β-helical AFPs. (a) Three different β-helical AFP structures are presented in ribbon format with key side chains on the ice-binding face displayed in stick format. Oxygen atoms are colored red to indicate the rotamer configuration. The ice-binding face is oriented such that the directionality of β-sheet elements goes from right to left: (1) *Tm*AFP (Protein Data Bank entry 1EZG), (2) sbwAFP (Protein Data Bank entry 1M8N), and (3) inchworm AFP 8k model. The precise orientation of the long side chain of the Arg residue was not predicted. The N- and C-termini are indicated as N- and C-, respectively. (b) The same structures in the same orientation are represented in space-filling format. Atoms are colored as indicated in the legend of Figure 8.

two outward-pointing residues (typically Thr) on either side of one inward-pointing residue. Increases in activity can be derived from additional loops,[59,60] which lengthen the ice-binding face without increasing its width.

With iwAFP, the longer loops facilitate the formation of a much wider ice-binding face (Figure 9a,b), containing a total of five ice-binding residues (again, typically Thr). Not surprisingly, the length of this surface is shorter, with only four loops versus 5−10 for the other β-helical AFPs. In this regard, the ice-binding surface of iwAFP is more similar to that of the sfAFP, which is also a parallel helical protein but with polyproline type II helical elements rather than β-strands. These coils in iwAFP are also wide enough to accommodate at least five ice-binding residues. An Arg (Leu in some isoforms) is present at one corner of the putative ice-binding face. This location and the flexibility of the Arg side chain may allow it to fold away from the ice-binding site to prevent steric hindrance. Bulkier residues are sometimes tolerated on the ice-binding surfaces of other AFPs, such as the Ile on the highly active sbwAFP (Figure 9b).[59] The modeled structure of iwAFP suggests two options for increasing its antifreeze activity by protein engineering. One is to add more coils; the other is to further extend the width of each coil by adding TA/S dipeptide repeats.

Other Characteristics of the Ice-Binding Face. Two of the common features of AFPs are that they tend to be amphipathic and have flat ice-binding sites. The ice-binding faces are typically rich in Thr and/or Ala. The other sides of the protein are often much more variable, containing both large and charged residues as seen with *Tm*AFP, *Dc*AFP, and sbwAFP.[38,40,59] A similar pattern is seen with iwAFP as the odd-numbered strands, which make up the putative ice-binding face, contain primarily outward-facing Thr residues (18 of 20), whereas the even-numbered strands, although still rich in Thr (9 of 15), contain a greater variety of residues (Arg, Asp, Asn, Val, Ile, and Ser).

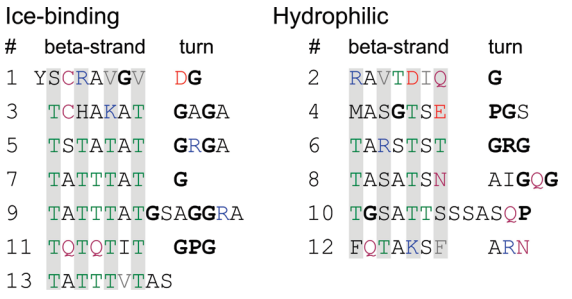


Figure 10. Predicted segmentation pattern for *Ri*AFP. The *Ri*AFP sequence is segmented on the basis of the iwAFP folding scheme. The sequence is arranged as β-strand segments based on the Thr-Ala/Ser dipeptide repeat pattern. The putative ice-binding face and hydrophilic face are categorized and labeled. Shorter regions in which this pattern is disrupted, often by Pro or Gly, are shown to the right of the putative strands. Outward-pointing residues are highlighted in gray.

***Rhagium inquisitor* AFP.** The recently reported novel AFP from *R. inquisitor* has an important bearing on the evolution of AFPs.[58] *Ri*AFP is the second distinct AFP type found in beetles,[61,62] which means that not all beetles have the *Tm*AFP-type structure seen in *T. molitor* and *D. canadensis*. As suggested for AFP evolution in fishes,[17] insects appear to have undergone their radiation and divergence before being challenged more recently with the need to have an AFP. A second example of unrelated AFPs appearing in one branch of insects is iwAFP, the second AFP discovered in lepidopterans (butterflies and moths) after the spruce budworm AFP.[63] Again, these two AFPs clearly indicate that adaptation to an environmental stress drove the two moth species to evolve different AFPs after they had differentiated from a common ancestor species.

Although *Ri*AFP (beetle) and iwAFP sequences are not homologous, both share a similar dipeptidic pattern. The Thr-Ala/Ser repeats in iwAFP are nearly continuous through the whole sequence, resulting in numerous Thr residues on both the putative ice-binding face and the hydrophilic face (Figure 5). In contrast, *Ri*AFP shows a clearer clustering of Thr-Ala repeats on one side of the β-helix than does iwAFP, which made it easier to recognize the segments of its structure (Figure 10). Short sequences composed of Gly and Pro are also present in *Ri*AFP and break the dipeptidic repeats as they do in iwAFP. There are several other differences between the two AFPs. *Ri*AFP segments tend to have four Thr residues per strand[58] instead of the five found in iwAFP . Alanines in *Ri*AFP are often replaced with Gly or Gln in addition to Ser found in iwAFP; however, it could still possess a silklike core[56] similar to that of iwAFP. Interestingly, *Ri*AFP is predicted to make an N-terminal disulfide bond if it follows the iwAFP folding scheme. On the basis of the aforementioned findings (Figure 10), the two AFPs may be structural analogues and provide another striking example of convergent evolution.

Hyperactivity. On the basis of the recent hypothesis by Scotter et al.,[64] the burst pattern at the end of the thermal hysteresis gap suggests that the higher activity of certain (hyperactive) AFPs is dependent on binding to the basal plane as well as to a second plane, such as the prism plane. iwAFP is clearly hyperactive as it shows >6 °C of activity, and its burst pattern, showing *a*-axial dendritic growth rather than spicular growth from the basal plane, is consistent with dual-plane binding. It is also possible to predict the planes of ice to which the 8 kDa iwAFP binds on the basis of the model. Because of the similarity in Thr richness, iwAFP is expected to share a similar ice binding mechanism with *Tm*AFP and sbwAFP.[15,38] The TxTxTxTxT motif is a longer version of the TxT motif found on sbw- and *Tm*AFPs and the TxTxTxT motif found on the hyperactive *Ri*AFP.[58] The distance between every second residue along a strand in silk is thought to be 6.98 Å,[54] intermediate between the 6.7 and 7.4 Å spacing observed in the crystal structures of sbwAFP and *Tm*AFP, respectively;[15,38] our model has a distance of 6.6 Å, similar to that of sbwAFP. The spacing between Thr residues on adjacent strands is close to 4.5 Å for all three insect AFPs. Therefore, dual-plane binding would be expected, as clearly demonstrated for sbwAFP.[15]

SUMMARY

Larvae of the pale beauty moth, *C. perlata*, collected in winter were shown to have freezing point depression activity exceeding 6 °C. Two AFP variants (~3 and ~8 kDa) that contribute to this activity were purified by IAP and their corresponding cDNA sequences cloned. Both sequences encode larger ~40 kDa proteins from which the bulk of the sequence is cleaved, leaving the highly repetitive mature AFPs. Both variants are dominated by Thr-Ala dipeptide repeats, and CD suggests that the predominant secondary structure is β-strand. The larger variant was modeled as a parallel β-helix using an intuitive approach based upon the repetitive nature of the sequence, breaks within the dipeptide pattern, and the polymorphisms found in the different isoforms. The resulting seven-stranded β-helix is stabilized by a core consisting primarily of interdigitating Ala and Ser residues, similar to that proposed for silk. Although both outward-facing surfaces are rich in Thr residues, the side containing the four odd-numbered strands is proposed to be the ice-binding surface as it is much flatter and more regular. All but two of the residues are Thr residues, forming a surface reminiscent of the ice-binding surfaces of the AFPs from spruce budworm (sbwAFP) and beetles (*Tm*AFP). These latter AFPs contain a double row of Thr on the β-sheet that serves as the ice-binding site, whereas iwAFP contains five ice-binding residues per β-strand.

ASSOCIATED CONTENT

Supporting Information. Sequence alignments of the unprocessed iwAFPs, root-mean-square fluctuations of the iwAFP model, and cross-sectional comparisons with the silk model and other AFPs. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Telephone: (613) 533-2983. Fax: (613) 533-2497. E-mail: peter.davies@queensu.ca.

Funding Sources

This work was supported by Canadian Institutes of Health Research Grant 6199. F.-H.L. was supported by an E. G. Bauman Fellowship and an Ontario Graduate Scholarship. P.L.D. holds a Canada Research Chair in Protein Engineering.

ACKNOWLEDGMENT

We thank David McLeod and Kim Munro of the Protein Function and Discovery Facility at Queen's University for MS

and CD analyses, respectively. We also thank David Roquis and Pierre LePage at the Génome Québec Innovation Centre for EST sequencing. David Comartin and Andrew Sage contributed by preparing phagemid stocks and analyzing EST sequences. We are grateful to many past and present members of the Davies lab for combing the woodlands in winter to gather inchworms. Robert Campbell assisted with data analysis, and Sherry Gauthier was invaluable in providing assistance with much of the cloning and PCR work. We thank both Robert Campbell and Christopher Garnham for their critical comments on the manuscript.

■ ABBREVIATIONS

AFP, antifreeze protein; BOLD, Barcode of Life Data System; *Dc*AFP, *D. canadensis* AFP; EST, expressed sequence tag; IAP, ice affinity purification; IBP, ice-binding protein; iwAFP, inchworm AFP; *Lp*AFP, *Lolium perenne* AFP; *Mp*AFP, *Marinomonas primoryensis* AFP; mOsm, milliosmole; MS/MS, tandem mass spectrometry; pfu, plaque forming units; *Ri*AFP, *R. inquisitor* AFP; RMSD, root-mean-square deviation; rmsf, root-mean-square fluctuation; sbwAFP, spruce budworm AFP; sfAFP, snow flea AFP; TH, thermal hysteresis; *Tm*AFP, *T. molitor* AFP.

■ REFERENCES

(1) Jia, Z., and Davies, P. L. (2002) Antifreeze proteins: An unusual receptor-ligand interaction. *Trends Biochem. Sci. 27*, 101–106.

(2) Raymond, J. A. (2000) Distribution and partial characterization of ice-active molecules associated with sea-ice diatoms. *Polar Biol. 23*, 721–729.

(3) Kuiper, M. J., Lankin, C., Gauthier, S. Y., Walker, V. K., and Davies, P. L. (2003) Purification of antifreeze proteins by adsorption to ice. *Biochem. Biophys. Res. Commun. 300*, 645–648.

(4) Knight, C. A., Cheng, C. C., and DeVries, A. L. (1991) Adsorption of α-helical antifreeze peptides on specific ice crystal surface planes. *Biophys. J. 59*, 409–418.

(5) Zhang, C., Zhang, H., Wang, L., Zhang, J., and Yao, H. (2007) Purification of antifreeze protein from wheat bran (*Triticum aestivum* L.) based on its hydrophilicity and ice-binding capacity. *J. Agric. Food Chem. 55*, 7654–7658.

(6) Raymond, J. A., Fritsen, C., and Shen, K. (2007) An ice-binding protein from an Antarctic sea ice bacterium. *FEMS Microbiol. Ecol. 61*, 214–221.

(7) Graham, L. A., and Davies, P. L. (2005) Glycine-rich antifreeze proteins from snow fleas. *Science 310*, 461.

(8) Duman, J. G. (2001) Antifreeze and ice nucleator proteins in terrestrial arthropods. *Annu. Rev. Physiol. 63*, 327–357.

(9) Fletcher, G. L., Hew, C. L., and Davies, P. L. (2001) Antifreeze proteins of teleost fishes. *Annu. Rev. Physiol. 63*, 359–390.

(10) Celik, Y., Graham, L. A., Mok, Y. F., Bar, M., Davies, P. L., and Braslavsky, I. (2010) Superheating of ice crystals in antifreeze protein solutions. *Proc. Natl. Acad. Sci. U.S.A. 107*, 5423–5428.

(11) Knight, C. A., DeVries, A. L., and Oolman, L. D. (1984) Fish antifreeze protein and the freezing and recrystallization of ice. *Nature 308*, 295–296.

(12) Chen, L., DeVries, A. L., and Cheng, C. H. (1997) Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc. Natl. Acad. Sci. U.S.A. 94*, 3811–3816.

(13) DeVries, A. L., and Wohlschlag, D. E. (1969) Freezing resistance in some Antarctic fishes. *Science 163*, 1073–1075.

(14) Sicheri, F., and Yang, D. S. (1995) Ice-binding structure and mechanism of an antifreeze protein from winter flounder. *Nature 375*, 427–431.

(15) Graether, S. P., Kuiper, M. J., Gagne, S. M., Walker, V. K., Jia, Z., Sykes, B. D., and Davies, P. L. (2000) β-Helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect. *Nature 406*, 325–328.

(16) Pentelute, B. L., Gates, Z. P., Tereshko, V., Dashnau, J. L., Vanderkooi, J. M., Kossiakoff, A. A., and Kent, S. B. (2008) X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers. *J. Am. Chem. Soc. 130*, 9695–9701.

(17) Scott, G. K., Fletcher, G. L., and Davies, P. L. (1986) Fish Antifreeze Proteins: Recent Gene Evolution. *Can. J. Fish. Aquat. Sci. 43* (5), 1028–1034.

(18) Cheng, C. H. (1998) Evolution of the diverse antifreeze proteins. *Curr. Opin. Genet. Dev. 8*, 715–720.

(19) Devries, A. L., and Lin, Y. (1977) Structure of a peptide antifreeze and mechanism of adsorption to ice. *Biochim. Biophys. Acta 495*, 388–392.

(20) Sonnichsen, F. D., DeLuca, C. I., Davies, P. L., and Sykes, B. D. (1996) Refined solution structure of type III antifreeze protein: Hydrophobic groups may be involved in the energetics of the protein-ice interaction. *Structure 4*, 1325–1337.

(21) Chao, H., Houston, M. E., Jr., Hodges, R. S., Kay, C. M., Sykes, B. D., Loewen, M. C., Davies, P. L., and Sonnichsen, F. D. (1997) A diminished role for hydrogen bonds in antifreeze protein binding to ice. *Biochemistry 36*, 14652–14660.

(22) Haymet, A. D., Ward, L. G., Harding, M. M., and Knight, C. A. (1998) Valine substituted winter flounder 'antifreeze': Preservation of ice growth hysteresis. *FEBS Lett. 430*, 301–306.

(23) Nutt, D. R., and Smith, J. C. (2008) Dual function of the hydration layer around an antifreeze protein revealed by atomistic molecular dynamics simulations. *J. Am. Chem. Soc. 130*, 13066–13073.

(24) Modig, K., Qvist, J., Marshall, C. B., Davies, P. L., and Halle, B. (2010) High water mobility on the ice-binding surface of a hyperactive antifreeze protein. *Phys. Chem. Chem. Phys. 12*, 10189–10197.

(25) Davies, P. L., Baardsnes, J., Kuiper, M. J., and Walker, V. K. (2002) Structure and function of antifreeze proteins. *Philos. Trans. R. Soc. London, Ser. B 357*, 927–935.

(26) Marshall, C. B., Chakrabartty, A., and Davies, P. L. (2005) Hyperactive antifreeze protein from winter flounder is a very long rod-like dimer of α-helices. *J. Biol. Chem. 280*, 17920–17929.

(27) Kuiper, M. J., Davies, P. L., and Walker, V. K. (2001) A theoretical model of a plant antifreeze protein from *Lolium perenne*. *Biophys. J. 81*, 3560–3565.

(28) Lin, F. H., Graham, L. A., Campbell, R. L., and Davies, P. L. (2007) Structural modeling of snow flea antifreeze protein. *Biophys. J. 92*, 1717–1723.

(29) Garnham, C. P., Gilbert, J. A., Hartman, C. P., Campbell, R. L., Laybourn-Parry, J., and Davies, P. L. (2008) A Ca^{2+}-dependent bacterial antifreeze protein domain has a novel β-helical ice-binding fold. *Biochem. J. 411*, 171–180.

(30) Chakrabartty, A., and Hew, C. L. (1991) The effect of enhanced α-helicity on the activity of a winter flounder antifreeze polypeptide. *Eur. J. Biochem. 202*, 1057–1063.

(31) Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res. 25*, 3389–3402.

(32) Ratnasingham, S., and Hebert, P. D. (2007) bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Mol. Ecol. Notes 7*, 355–364.

(33) DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA.

(34) Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput. 4*, 435–447.

(35) Liou, Y. C., Daley, M. E., Graham, L. A., Kay, C. M., Walker, V. K., Sykes, B. D., and Davies, P. L. (2000) Folding and structural characterization of highly disulfide-bonded beetle antifreeze protein produced in bacteria. *Protein Expression Purif. 19*, 148–157.

(36) Mok, Y. F., Lin, F. H., Graham, L. A., Celik, Y., Braslavsky, I., and Davies, P. L. (2010) Structural basis for the superior activity

of the large isoform of snow flea antifreeze protein. *Biochemistry 49*, 2593–2603.

(37) Gauthier, S. Y., Kay, C. M., Sykes, B. D., Walker, V. K., and Davies, P. L. (1998) Disulfide bond mapping and structural characterization of spruce budworm antifreeze protein. *Eur. J. Biochem. 258*, 445–453.

(38) Liou, Y. C., Tocilj, A., Davies, P. L., and Jia, Z. (2000) Mimicry of ice structure by surface hydroxyls and water of a β-helix antifreeze protein. *Nature 406*, 322–324.

(39) Graham, L. A., Liou, Y. C., Walker, V. K., and Davies, P. L. (1997) Hyperactive antifreeze protein from beetles. *Nature 388*, 727–728.

(40) Duman, J. G., Li, N., Verleye, D., Goetz, F. W., Wu, D. W., Andorfer, C. A., Benjamin, T., and Parmelee, D. C. (1998) Molecular characterization and sequencing of antifreeze proteins from larvae of the beetle *Dendroides canadensis*. *J. Comp. Physiol., B 168*, 225–232.

(41) Liou, Y. C., Thibault, P., Walker, V. K., Davies, P. L., and Graham, L. A. (1999) A complex family of highly heterogeneous and internally repetitive hyperactive antifreeze proteins from the beetle *Tenebrio molitor*. *Biochemistry 38*, 11415–11424.

(42) Andorfer, C. A., and Duman, J. G. (2000) Isolation and characterization of cDNA clones encoding antifreeze proteins of the pyrochroid beetle *Dendroides canadensis*. *J. Insect Physiol. 46*, 365–372.

(43) Asakura, T., and Yao, J. (2002) ^{13}C CP/MAS NMR study on structural heterogeneity in *Bombyx mori* silk fiber and their generation by stretching. *Protein Sci. 11*, 2706–2713.

(44) Lin, Y., and Long, D. J. (1980) Purification and characterization of winter flounder antifreeze peptide messenger ribonucleic acid. *Biochemistry 19*, 1111–1116.

(45) Hayes, P. H., Scott, G. K., Ng, N. F., Hew, C. L., and Davies, P. L. (1989) Cystine-rich type II antifreeze protein precursor is initiated from the third AUG codon of its mRNA. *J. Biol. Chem. 264*, 18761–18767.

(46) Davies, P. L., Roach, A. H., and Hew, C. L. (1982) DNA sequence coding for an antifreeze protein precursor from winter flounder. *Proc. Natl. Acad. Sci. U.S.A. 79*, 335–339.

(47) Hsiao, K. C., Cheng, C. H., Fernandes, I. E., Detrich, H. W., and DeVries, A. L. (1990) An antifreeze glycopeptide gene from the antarctic cod *Notothenia coriiceps neglecta* encodes a polyprotein of high peptide copy number. *Proc. Natl. Acad. Sci. U.S.A. 87*, 9265–9269.

(48) Pertaya, N., Marshall, C. B., DiPrinzio, C. L., Wilen, L., Thomson, E. S., Wettlaufer, J. S., Davies, P. L., and Braslavsky, I. (2007) Fluorescence microscopy evidence for quasi-permanent attachment of antifreeze proteins to ice surfaces. *Biophys. J. 92*, 3663–3673.

(49) Ravulapalli, R., Campbell, R. L., Gauthier, S. Y., Dhe-Paganon, S., and Davies, P. L. (2009) Distinguishing between calpain heterodimerization and homodimerization. *FEBS J. 276*, 973–982.

(50) DeLuca, C. I., Comley, R., and Davies, P. L. (1998) Antifreeze proteins bind independently to ice. *Biophys. J. 74*, 1502–1508.

(51) Graham, L. A., Marshall, C. B., Lin, F. H., Campbell, R. L., and Davies, P. L. (2008) Hyperactive antifreeze protein from fish contains multiple ice-binding sites. *Biochemistry 47*, 2051–2063.

(52) Liou, Y. C., Daley, M. E., Graham, L. A., Kay, C. M., Walker, V. K., Sykes, B. D., and Davies, P. L. (2000) Folding and structural characterization of highly disulfide-bonded beetle antifreeze protein produced in bacteria. *Protein Expression Purif. 19*, 148–157.

(53) Zhou, C. Z., Confalonieri, F., Medina, N., Zivanovic, Y., Esnault, C., Yang, T., Jacquet, M., Janin, J., Duguet, M., Perasso, R., and Li, Z. G. (2000) Fine organization of *Bombyx mori* fibroin heavy chain gene. *Nucleic Acids Res. 28*, 2413–2419.

(54) Takahashi, Y., Gehoh, M., and Yuzuriha, K. (1999) Structure refinement and diffuse streak scattering of silk (*Bombyx mori*). *Int. J. Biol. Macromol. 24*, 127–138.

(55) Fossey, S. A., Nemethy, G., Gibson, K. D., and Scheraga, H. A. (1991) Conformational energy studies of β-sheets of model silk fibroin peptides. I. Sheets of poly(Ala-Gly) chains. *Biopolymers 31*, 1529–1541.

(56) Lotz, B., and Colonna Cesari, F. (1979) The chemical structure and the crystalline structures of *Bombyx mori* silk fibroin. *Biochimie 61*, 205–214.

(57) Suzuki, Y., Okonogi, M., Yamauchi, K., Kurosu, H., Tansho, M., Shimizu, T., Saito, H., and Asakura, T. (2007) High-field ^{1}H MAS and ^{15}N CP-MAS NMR studies of alanine tripeptides and oligomers: Distinction of antiparallel and parallel β-sheet structures and two crystallographically independent molecules. *J. Phys. Chem. B 111*, 9172–9178.

(58) Kristiansen, E., Ramlov, H., Hojrup, P., Pedersen, S. A., Hagen, L., and Zachariassen, K. E. (2011) Structural characteristics of a novel antifreeze protein from the longhorn beetle *Rhagium inquisitor*. *Insect Biochem. Mol. Biol. 41*, 109–117.

(59) Leinala, E. K., Davies, P. L., Doucet, D., Tyshenko, M. G., Walker, V. K., and Jia, Z. (2002) A β-helical antifreeze protein isoform with increased activity. Structural and functional insights. *J. Biol. Chem. 277*, 33349–33352.

(60) Marshall, C. B., Daley, M. E., Sykes, B. D., and Davies, P. L. (2004) Enhancing the activity of a β-helical antifreeze protein by the engineered addition of coils. *Biochemistry 43*, 11637–11646.

(61) Walters, K. R., Serianni, A. S., Sformo, T., Barnes, B. M., and Duman, J. G. (2009) A nonprotein thermal hysteresis-producing xylomannan antifreeze in the freeze-tolerant Alaskan beetle *Upis ceramboides*. *Proc. Natl. Acad. Sci. U.S.A. 106*, 20210–20215.

(62) Tomchaney, A. P., Morris, J. P., Kang, S. H., and Duman, J. G. (1982) Purification, composition, and physical properties of a thermal hysteresis "antifreeze" protein from larvae of the beetle *Tenebrio molitor*. *Biochemistry 21*, 716–721.

(63) Tyshenko, M. G., Doucet, D., Davies, P. L., and Walker, V. K. (1997) The antifreeze potential of the spruce budworm thermal hysteresis protein. *Nat. Biotechnol. 15*, 887–890.

(64) Scotter, A. J., Marshall, C. B., Graham, L. A., Gilbert, J. A., Garnham, C. P., and Davies, P. L. (2006) The basis for hyperactivity of antifreeze proteins. *Cryobiology 53*, 229–239.