Hyperactive Antifreeze Protein from Fish Contains Multiple Ice-Binding Sites†

Laurie A. Graham, Christopher B. Marshall,‡ Feng-Hsu Lin, Robert L. Campbell, and Peter L. Davies*

Department of Biochemistry and Protein Function Discovery Group, Queen's University, Kingston, Ontario, Canada K7L 3N6

Received October 10, 2007; Revised Manuscript Received November 27, 2007

ABSTRACT: Antifreeze proteins (AFPs) are produced to prevent freezing in many fish species that are exposed to icy seawater. There are a number of nonhomologous types of AFPs, diverse in both sequence and structure, which share the function of binding to ice and inhibiting its growth. We recently discovered a hyperactive AFP in the winter flounder and related species that is many-fold more active than other fish AFPs. Like the 3−4-kDa type I AFPs, it is alanine-rich and highly helical, but this 17-kDa protein is considerably larger and forms a dimer. We have sequenced the cDNA encoding this new AFP to gain insight into its structure and evolutionary relationship to the type I AFP family. The gene is clearly homologous to the righteye flounder type I AFP genes. Thus we have designated this protein "hyperactive type I AFP" (hyp-type I). The sequence of hyp-type I AFP supports a structural model in which two extended 195-amino acid alpha-helices form an amphipathic homodimer with a series of linked Ala- and Thr-rich patches on the surface of the dimer, each of which resembles ice-binding sites of type I AFPs. The superior activity of hyp-type I AFP may derive from the large combined surface area of the ice-binding sites, recognition of multiple planes of ice, and protection of the basal plane from ice growth.

Antifreeze proteins (AFPs[1]) are diverse in both sequence and structure but share the ability to bind to ice and prevent it from growing (*1*−*3*). In fishes, there are at least five different types of AFP that confer freeze resistance on their hosts (*4*). Type I AFPs, defined as small, Ala-rich (∼60% Ala) alpha-helices (*5*, *6*), are present in fishes from three taxonomic orders. These include pleuronectiformes (righteye flounders) such as the winter flounder (*7*) and yellowtail flounder (*8*), scorpaeniformes such as sculpins (*9*−*11*) and snailfish (*12*), and a perciforme, the cunner (*13*). Typically, the type I AFPs form amphipathic helices with a well-conserved Ala-rich surface opposite a less conserved, more hydrophilic side of the helix (*14*). Replacement studies have shown that even a single substitution of an Ala by a bulky residue on the Ala-rich surface can eliminate antifreeze activity of both the winter flounder and shorthorn sculpin AFPs (*14*, *15*). This is supported by NMR relaxation studies showing that Ala residues on this face of winter flounder isoform HPLC-6 interact more strongly with ice than those on the opposite surface (*16*). Other replacement studies, in which the Thr residues (spaced at 11-residue intervals) flanking the Ala-rich surface of HPLC-6 were replaced with Ser or Val, have illustrated the importance of the threonyl methyl group for activity (*17*, *18*). These data have led to the view that it is the hydrophobic surface that interacts with the ice/water interface (*19*), which is supported by models that show this face can be snugly fit to the experimentally determined planes of ice to which the AFPs bind (*3*, *15*, *20*).

In winter flounder, *Pseudopleuronectes americanus*, the species from which type I AFPs have been most extensively studied, the many isoforms produced can be grouped into two sets (*6*). Liver-specific isoforms are exported from this tissue into the blood, directed by the signal sequence mechanism, where they are thought to prevent ice growth in extracellular fluids. They are produced as proproteins that are processed at the N terminus by a dipeptidyl peptidase and at the C terminus by amidation of the penultimate residue (Arg) (*21*, *22*). The mature protein is typified by isoform HPLC-6, a 37-residue peptide containing three 11-amino-acid repeats, which forms an isolated alpha-helix stabilized by cap structures at both ends (*23*). The other set of isoforms is produced in skin, scales, fins and gills (*24*). These proteins have no signal sequence, and modification is limited to acetylation of the N-terminal Met. Immunohistochemistry indicates intracellular (*25*) or local extracellular accumulation (*26*), depending on tissue type. These isoforms are thought to provide the first line of defense against freezing from nucleation by ice crystals in the environment.

Until very recently it appeared that all fish AFPs had roughly comparable thermal hysteresis activity, which is defined as the ability to lower the freezing temperature below the melting point (*27*). AFP types I, II, III and IV and the antifreeze glycoproteins (AFGPs) produce ∼1 C° of thermal hysteresis at the relatively high AFP concentrations (10−40 mg/mL) found in their blood plasma. In most species, this level of freeze protection is sufficient to cover the difference between the freezing points of seawater (−1.9 °C) and unprotected fish plasma (−0.8 °C) (*5*). In contrast, insect AFPs are 10 to 100 times more effective, producing thermal

† This research was funded by a grant to P.L.D. from the Canadian Institutes for Health Research. P.L.D. holds a Canada Research Chair in Protein Engineering.

* Author for correspondence. Tel: 613-533-2983. Fax: 613-533-2497. E-mail: peter.davies@queensu.ca.

‡ Current address: Division of Signaling Biology, Ontario Cancer Institute, Toronto, ON, Canada M5G 1L7.

[1] Abbreviations: AFP, antifreeze protein; hyp-type I AFP, type I hyperactive AFP; AFGP, antifreeze glycoprotein.

10.1021/bi7020316 CCC: $40.75
Published on Web 01/29/2008

hysteresis values over 3 C° at less than one-tenth the concentration (*28*). These AFPs are described as hyperactive (*29*) and their evolution in over-wintering insects could be rationalized by the requirement for these terrestrial organisms to adapt to far colder temperatures than marine fish ever encounter. Thus the discovery of a hyperactive AFP in the winter flounder was surprising (*30*), particularly because the type I isoforms produced by this species had been recognized and extensively characterized over the last 30 years (*6*). The hyperactive AFP provides ~1.1 C° of thermal hysteresis activity at its circulating concentration of 0.1 mg/mL, suggesting that this protein alone might be adequate to protect the flounder at the freezing point of seawater. On the other hand, it appears that the fish would not be protected by type I alone, even though it is present at ~100-fold higher concentration. Subsequently, we have found homologues of this hyperactive AFP in species from two other genera of righteye flounders, yellowtail flounder (*Limanda ferruginea*) and American plaice (*Hippoglossoides platessoides*) (*31*). In the latter species, the hyperactive AFP appears to be the only antifreeze present in the blood.

The hyperactive and type I AFPs of winter flounder exhibit some intriguing differences and similarities. Type I AFPs have been extensively characterized as monomeric, 3−4-kDa peptides whereas the hyperactive AFP is a highly elongated 32-kDa homodimer. However, they all share high Ala content (~60%) and are entirely alpha-helical (*23*, *32*, *33*). Previously, we cloned a genomic sequence named "*5a*" with a putative exon encoding an exceptionally long Ala-rich sequence. It seemed likely that *5a* was a pseudogene since no corresponding protein nor mRNA was detected (*34*). The N-terminal sequence, amino acid composition and molecular weight of the newly discovered hyperactive AFP resembled those of the putative protein encoded by *5a*, thus the hyperactive AFP was referred to as "*5a*-like", even though differences in all three properties made it clear that the protein was not encoded by *5a* (*30*). Despite the similarities, the evolutionary relationship between the hyperactive AFP, the *5a* genomic sequence, and the small type I AFPs was unclear. To further characterize the hyperactive flounder AFP and elucidate its origins, we have sequenced peptides derived from the purified protein and used this information to clone the corresponding cDNA. The sequence clearly shows that the hyperactive AFP is homologous to the smaller flounder AFPs and has enabled us to develop a model for its structure and putative ice-binding surface(s).

MATERIALS AND METHODS

Tryptic Digestion. Hyp-type I AFP was purified by size-exclusion chromatography and ice affinity purification as previously described (*30*, *32*). Approximately 0.1 mg of protein was digested by trypsin in 50 mM ammonium bicarbonate (adjusted to pH 7.0 with HCl immediately before use), with 0.1 mM $CaCl_2$ added to the trypsin stock, again just before use. The initial digest was performed using 1 μg of bovine trypsin (Sigma-Aldrich, St. Louis, MO) at 13 °C for 1 h, followed by sequential 1 h incubations at increasing temperatures (18 °C, 21 °C, 37 °C), with addition of 0.4 μg of fresh trypsin each time. Digestion was monitored by MALDI-TOF MS and SDS−PAGE, and the cleavage products were concentrated by lyophilization for resuspension in 1/10 volume of 10 mM ammonium bicarbonate adjusted to pH 7.0 with HCl immediately before use.

Protein/Peptide Sequencing. Sequence data were primarily derived from LC−MS/MS on a Q-TOF Ultima Global (Waters, Milford, MA). Digested hyp-type I (500 fmol) was introduced into the mass spectrometer using a CapLC liquid chromatography system (Waters). One fragment (the largest at 3886.1) was analyzed by MALDI-MS/MS methods. For MALDI Q-TOF analysis, hyp-type I AFP (2 pmol) digest was mixed with 5 mg/mL alpha-cyano matrix in 70% acetonitrile and 0.1% trifluoroacetic acid. The masses of intact tryptic fragments were determined by the same method.

cDNA Library Construction. Winter flounder were collected in March from Chapel's Cove, Newfoundland, Canada, and the livers were rapidly harvested and flash frozen. Liver tissue (0.13 g) was ground under liquid nitrogen, and total mRNA was isolated using the RNeasy midi kit (Qiagen, Hilden, Germany). The polyadenylated RNA fraction was isolated using the Oligotex mRNA midi kit (Qiagen). A cDNA library of approximately 1×10^4 pfu was constructed using the ZAP-cDNA library synthesis kit (Stratagene, Santa Clara, CA) and amplified according to the manufacturer's instructions.

Cloning of Hyp-type I cDNA. All PCR amplifications were done using Taq DNA polymerase (Qiagen) with Q-solution (to amplify GC-rich templates) following the manufacturer's instructions. Approximately 20,000 pfu of the amplified library or 2 μL of previous PCR products (for reamplifications) was used as the template. Degenerate and nondegenerate primers were used at concentrations of 1 μM and 0.1 μM, respectively, in 50 μL reactions. Unless otherwise mentioned, reaction conditions were as follows: initial denaturation of 5 min at 95 °C, hot start at 80 °C, 30 cycles of 95 °C for 1 min, 55 °C for 1 min and 72 °C for 2 min, followed by a 20 min final extension.

Partially degenerate primers were designed and named based on the peptide sequences: IDPAARA = 5′-ATYG-AYCCHGCMGCMAGRGC-3′ (N-terminal sequence), IATINA = 5′-ATHGCMACHATYAAYGC-3′ (internal tryptic peptide), DNAAAAK antisense = 5′-YTTKGCKGCKGCK-GCRTTRTC-3′ (internal tryptic peptide). The IDPAARA and T7 extended primer (5′-CCCTCACTAAAGGGAACAAA-AGCTG-3′) were used in the first reaction as above but with a 56 °C annealing temperature. The IATINA and T7 extended primers were used in the first reamplification and the IATINA and DNAAAAK antisense primers in the second with an annealing temperature of 46 °C for 25 cycles for both. The single band of ~270 bp that was obtained was subcloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA) according to manufacturer's instructions and sequenced (Cortec, Kingston, Ontario, Canada).

The sequence of the internal fragment was used to design a nondegenerate primer (MAAKDTA antisense = 5′-GGCAGTGTCCTTGGCTGCCAT-3′) that contained mismatches relative to the *5a* gene sequence. The upstream portion of the cDNA was amplifed by anchor PCR using this primer and the T3 extended = 5′-CCCTCACTAAAGG-GAACAAAAGCTG-3′ primer. Then, a primer specific to the 5′ UTR was designed (5′ UTR = 5′-GTAGTGAAC-CAGTGCTCCCTAA-3′) and used with the T7 extended primer to obtain a contiguous hyp-type I coding sequence

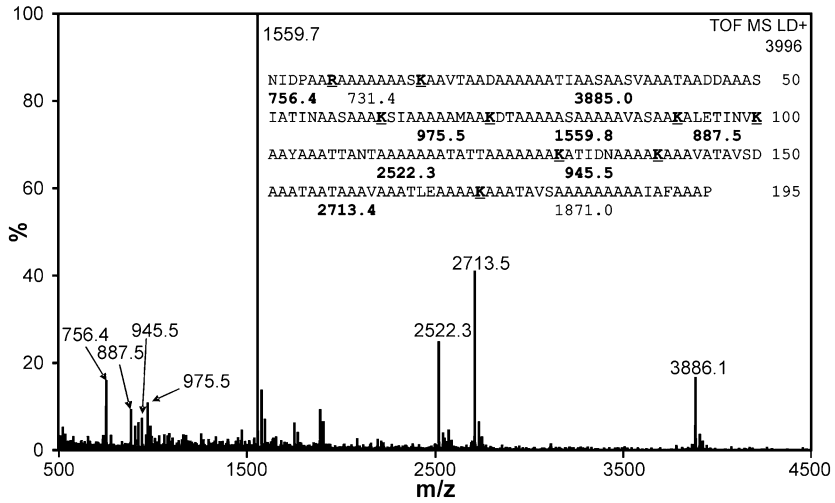


FIGURE 1: MALDI-MS spectrum of tryptic digestion of hyp-type I. Inset shows the linear sequence with basic residues in boldface and underlined with the predicted peptide masses (those observed in spectrum in bold) shown underneath the tryptic fragments.

from the library. All products were subcloned and sequenced as above.

Library Screening for a 5a cDNA. To determine whether the *5a* "pseudogene" is expressed at the mRNA level, primers that exactly matched the sequence of *5a* were designed. Mismatches relative to the newly acquired sequence of hyp-type I are underlined. Signal peptide = 5′-TCACTCT-TCACTGTCGGACAATTC-3′, IASINANT = 5′-CATAG-CAAGCATCAACGCCAACAC-3′, VAKTTIDK antisense = 5′-CTTGTCGATGGTTGTTTTGGCTAC-3′. The first reactions were performed as described above, using either the signal-peptide primer or IASINANT with the T7 primer. All possible combinations were reamplified with the same primers or with IASINANT to nest at the 5′ end and VAKTTIDK antisense to nest at the 3′ end. Products were subcloned and sequenced as above.

Molecular Modeling. The CD spectrum indicates that the secondary structure is completely alpha-helical, and the analytical ultracentrifugation data indicate a dimer with an axial ratio of 18:1 (*32*). An initial model of the structure was built with phi and psi angles set at −64° and −44° respectively, based on the angles measured, after excluding three to four residues at each terminus, in type I isoform HPLC-6 (−64° and −44°) and the tetramerization domain of tetrabrachion (−63° and −43°) (*23*, *35*). The angles were then slightly modified (−62° and −42°) to more precisely align every eleventh residue. The resulting helix showed a single discontinuity in the alignment of the 11-amino-acid repeats, about two-thirds of the way down the molecule, where an additional two residues interrupted the repeat pattern. Using PyMOL 0.98 (*36*), with the sculpting function activated to minimize clashes and unreasonable geometries, the coiling of the helix was increased slightly in this 13-residue region so that the 11-amino-acid repeats aligned on either side. This resulting amphipathic monomer was duplicated and docked, in an antiparallel orientation along an Ala-rich face devoid of large or charged residues, thereby placing the putative ice-binding faces of both subunits on the same surface.

RESULTS

Complete tryptic cleavage of hyp-type I AFP requires digestion in stages. Hyp-type I AFP is thermolabile and aggregates at room temperature (*32*) potentially reducing its susceptibility to tryptic digestion by decreasing the accessibility of protease cleavage sites. The use of denaturants or detergents to prevent this was avoided as these materials could interfere with ionization in mass spectrometry (MS). Therefore, the initial digest of purified hyp-type I AFP was performed at a low temperature (13 °C) at which this AFP is fully stable (*32*), causing partial cleavage of the long helices (data not shown), and presumably producing fragments that are less prone to aggregation. The second and third incubations were performed within the temperature range at which the first thermal denaturation occurs (18 °C and 21 °C) so that tryptic cleavage of newly exposed sites would compete with aggregation. The final stage was carried out at 37 °C, and digestion was judged to be complete as

Table 1: MS/MS Fragment Masses and Sequences Compared with Deduced Masses and Sequence of Hyp-type I

fragment mass			fragment sequences deduced from cDNA
observed[a]	predicted[b]	residues[c]	bolded residues sequenced by MS/MS[d]
756.4	756.4	1−7	N*IDPAAR*[d]
nd	731.4	8−16	*AAAAA*AASK
3886.1	3885.0	17−62	AAVTAADAAAAAATIAA**SAA-SVAAATAADDAAA**SIA TINAASAAAK (MALDI)
975.5	975.5	63−73	SI**AAAAAMAAK**
1559.7 (100)	1559.8	74−92	**DTAAAAASAAAAAVASAAK**
887.5	887.5	93−100	ALETINVK
2522.3	2522.3	101−130	AAYAAATTAN**TAAAAAAAT-ATTAAAAAAAK**
945.5	945.5	131−140	**ATIDNAAAAK**
2713.5	2713.4	141−172	AAAVATAVSDAAATAATA**AA-VAA**ATLEAAAAK
1871.0	1871.0	173−195	AAATAVSAA**AAAAAAAIAF-AAA**P polymorphic fragment[e]
1589.8	1589.8		*dtaa***AAASAAATAVASAAK**

[a] Fragment masses determined by MALDI except for those indicated in italics which were determined by LC−ESI-MS. [b] Predicted masses were determined using PEPTIDEMASS (*58*). [c] Amino acid residue numbers as in Figure 1. [d] Residues conclusively determined prior to cloning of the cDNA are in bold (MS/MS) or italics (Edman degradation). [e] Sequence of a minor species with residues deduced by similarity to hyp-type I in lowercase italics and the polymorphism underlined.

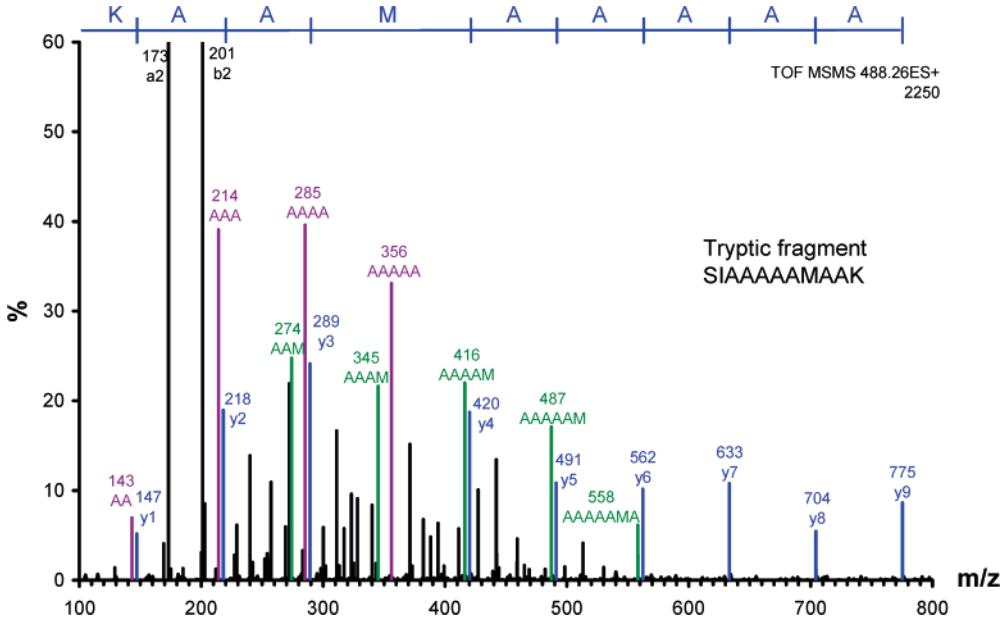


FIGURE 2: Representative MS/MS profile. The y ion series is shown in blue and was used to deduce the sequence of all but the first two residues. The strongest signals correspond to the b2 and a2 ions containing the first two residues (SI). Two internal ion series are indicated in green and purple. The other predominant signals correspond to internal or satellite fragments (data not shown).

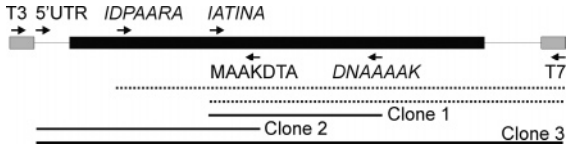


FIGURE 3: Hyp-type I AFP cloning strategy. The vector arms are represented by gray bars, coding sequence is represented by the black bar and noncoding sequences are represented with lines. Primers within the coding region are named for the corresponding amino acid residues with degenerate primers italicized. The dashed lines indicate the expected products of the first two PCR amplifications with cloned fragments shown as black lines.

masses corresponding to partial digestion products were not seen (Figure 1 and Table 1). The N-terminal sequence of hyp-type I AFP was previously determined by Edman degradation.

Sequencing of Tryptic Peptides Covering 43% of Hyp-type I AFP. Tryptic fragments were subjected to LC−ESI-MS/MS as well as MALDI-MS/MS with many fragments producing ion series of sufficient quality to provide sequence information (Table 1 and Figure 2). Manual interpretation of these series was required due to the Ala-richness of this AFP, since internal fragments of the same mass were often generated by different cleavages. For example, three different fragments of identical mass, containing one Met and three Ala (AAAM, AAMA, AMAA) could be produced from the peptide SIAAAAAMAAK (Figure 2). Three of the four fragments that are longer by one residue are also of equal mass as they contain an additional Ala, and so forth. Therefore, these peaks were often more intense than those of the y series. Nevertheless, conclusive sequence was obtained covering 43% of the protein.

Sequence of the 195-Residue Hyp-type I AFP Is 60% Ala. Both the N-terminal and internal MS/MS-derived sequences (Table 1) were used to design degenerate oligonucleotide primers for the purpose of amplifying a cDNA encoding hyp-type I AFP. This strategy was complicated by the high Ala content of the peptides, the abundance of Ser (which can be encoded by six possible codons) and the presence of isobaric residues (Ile and Leu). Primers were designed to regions of maximum amino acid sequence complexity that did not contain Ser residues. Where MS/MS data indicated the presence of isobaric residues, Ile was favored for two reasons: the *5a* sequence encodes Ile at equivalent positions, and the amino acid analysis indicated that hyp-type I AFP contained eight Ile but only two Leu residues (Table 2 and (*32*)). Degeneracy was further reduced by excluding those codons that are rarely found in either *5a* or type I AFPs (*34*).

Table 2: Composition of Hyp-type I AFP Determined by Amino Acid Analysis (aaa) of the Purified Protein and Deduced from the cDNA Sequence

residue	aaa	cDNA
Asp	13	7
Asn		5
Glu	3	2
Gln		0
Ser	9	10
Gly	1	0
His	0	0
Arg	1	1
Thr	18	18
Ala	120	121
Pro	3	2
Tyr	1	1
Val	8	8
Met	1	1
Ile	8	8
Leu	2	2
Phe	1	1
Lys	6	8
Cys	nd	0
Trp	nd	0
total	195	195
average mass	obs: 16683	predicted: 16685

Three nested PCR reactions, employing combinations of vector primers and degenerate hyp-type I primers (Figure 3), amplifed an internal <300 bp cDNA fragment from a library constructed from the liver of a single flounder collected in late winter, a period of peak AFP production.

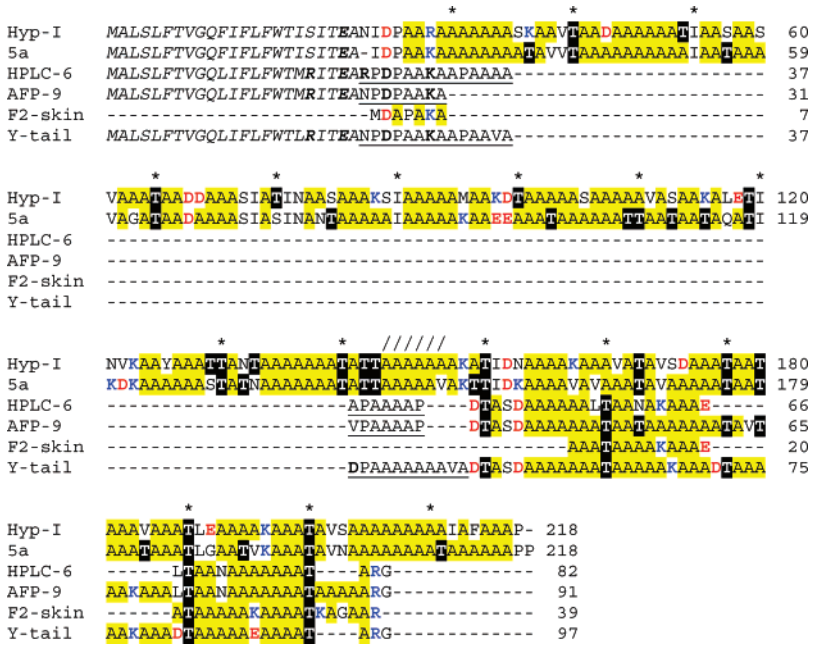


FIGURE 4: Alignment of five winter flounder and one yellowtail flounder AFP sequences. Secretory signal-peptides are in italics, and the pro-peptide regions are underlined. Some sequences contain a C-terminal Gly that is removed, leaving an amidated terminus. Residues within the mature proteins are colored as follows: Ala highlighted yellow, Thr highlighted black, acidic residues in red text and basic residues in blue text. The alignment mirrors the mRNA alignment. Asterisks are used to denote the Thr spaced at 11-residue intervals in the shorter isoforms. This pattern can be extended in the longer isoforms (with the exception of the break in the pattern indicated by slashes), with Thr at up to 10 of the 17 positions. Y-tail is the yellowtail flounder AFP sequence. Hyp-I is the hyp-type I AFP sequence.

Table 3: Ala Codon Usage and Third Codon Position G or C Content (%) in Non-Ala Codons of AFPs Compared to Non-AFPs[a]

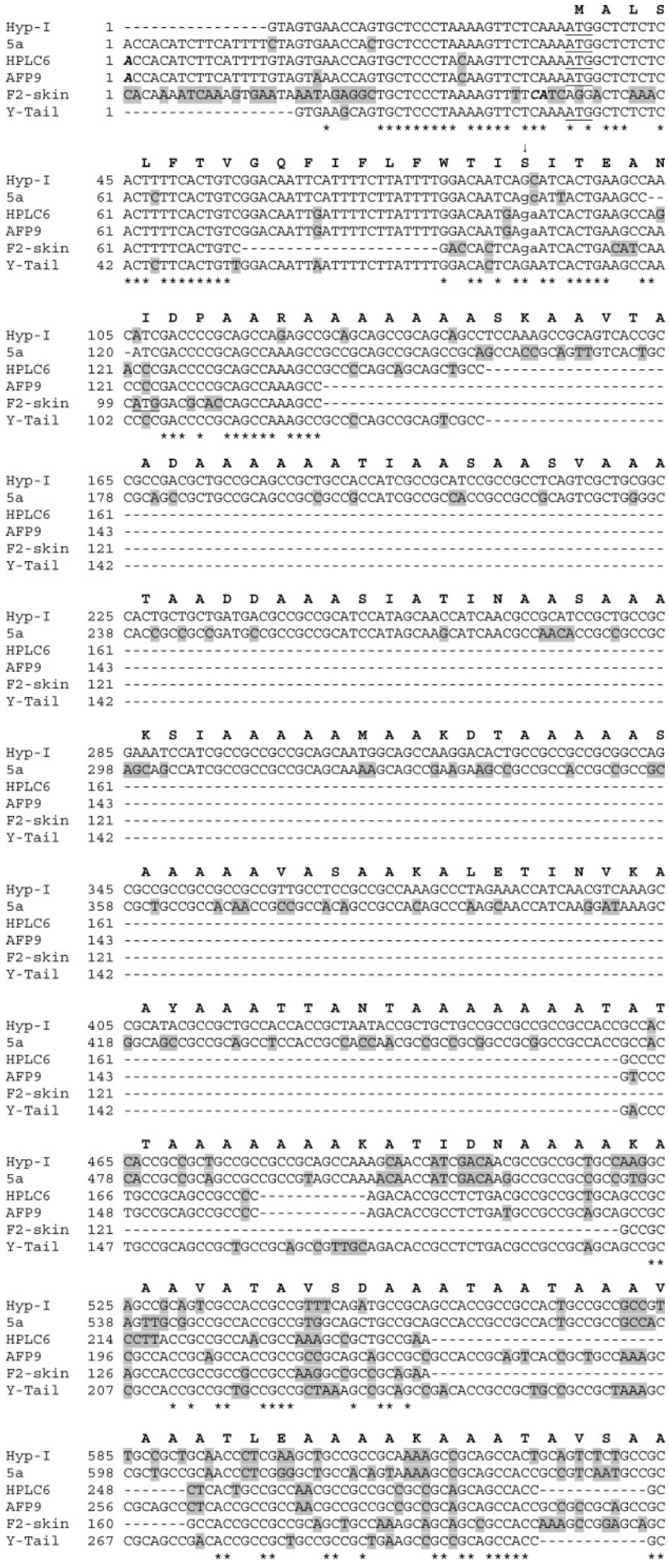
	Ala codons (%)[b]				codon no.		%	
	GCC	GCA	GCT	GCG	Ala	all	third pos. G[c]	third pos. C[c]
AFPs								
hyp-I	**63**	20	15	*2*	123	219	2 (3)	58
5a	**70**	20	6	*3*	128	219	4 (2)	52
HPLC6	**71**	18	11	*0*	38	83	2 (3)	38
AFP9	**70**	26	4	*0*	46	92	2 (3)	49
F2-skin	**69**	27	4	*0*	26	40	8 (1)	31
Y-tail	**65**	20	16		51	98	0 (2)	47
non-AFPs[c]								
opsin	30	23	**38**	*10*	40	358	17 (28)	21
zona pel.	21	**38**	33	*8*	24	510	24 (7)	18
THR	**57**	*0*	13	30	23	417	41 (22)	38
pepsinogen	**40**	16	**40**	*4*	25	379	21 (16)	47
NaCl trans.	**46**	*9*	32	12	74	1024	31 (44)	37
β-tubulin	**55**	14	28	*3*	29	446	28 (23)	34
amylase	**61**	7	29	*4*	28	513	25 (35)	42
aminopep.	**37**	17	33	*12*	81	975	28 (47)	31

[a] Complete coding sequences with the following accession numbers were obtained from GenBank; winter flounder AFPs (hyp-I = EU188795, *5a* = M63476 + M63477, HPLC6 = X07506, AFP9 = X53718, F2-skin = M63479), yellowtail flounder AFP (Y-tail = X06356), winter flounder non-AFPs (opsin = AY631039, zona pel. = U03674, THR = AY794223, pepsinogen = AF156787, NaCl trans. = L11615, β-tubulin = X74492). [b] The % Ala encoded by each codon is shown along with the total number of both Ala and all codons. [c] Total % G or % C in the third codon position, after exclusion of Ala codons and those that must end in G (Met and Trp, number in brackets).

Nondegenerate primers were designed to this sequence and used with vector primers to obtain the 5′ end of the cDNA. The third step was to use a primer designed to the 5′ UTR, together with a downstream vector primer, to amplify the entire cDNA sequence in one segment.

Hyp-type I AFP Is a Secreted Protein. The resulting cDNA contains an open reading frame in which the first 23 residues (MALSLFTVGQFIFLFWTISITEA) are predicted by SignalP (*37*) to encode a secretory signal peptide (Figure 4). The sequence following the predicted signal peptide cleavage site exactly matches the N-terminal sequence of the purified mature protein as determined by Edman degradation (NIDPAARAAAAA) (*32*). The predicted average mass of the mature 195-residue protein encoded by this cDNA is 16,686 Da, consistent with the mass of 16,685 ± 3 Da observed by MALDI-MS.

The tryptic fingerprint obtained for the purified protein matches that predicted from the translation of the cDNA sequence (Table 1 and Figure 1) as nine of the ten predicted fragments were observed. In addition, the predicted and observed amino acid compositions agree remarkably well (Table 2). Partial or complete peptide sequences were determined for the nine observed fragments and are in complete agreement with the deduced sequence (Table 1).


M A L S
Hyp-I 1 ----------GTAGTGAACCAGTGCTCCCTAAAAGTTCTCAAAATGGCTCTCTC
5a 1 ACCACATCTTCATTTTCTAGTGAACCACTGCTCCCTAAAAGTTCTCAAAATGGCTCTCTC
HPLC6 1 ACCACATCTTCATTTTGTAGTGAACCAGTGCTCCCTACAAGTTCTCAAAATGGCTCTCTC
AFP9 1 ACCACATCTTCATTTTGTAGTAAACCAGTGCTCCCTACAAGTTCTCAAAATGGCTCTCTC
F2-skin 1 CACAAATCAAAGTGAATAAATAGAGGCTGCTCCCTAAAAGTTTCATCAGGACTCAAAC
Y-Tail 1 ----------GTGAAGCAGTGCTCCCTAAAAGTTCTCAAAATGGCTCTCTC
L F T V G Q F I F L F W T I S I T E A N
Hyp-I 45 ACTTTTCACTGTCGGACAATTCATTTTCTTATTTTGGACAATCAGCATCACTGAAGCCAA
5a 61 ACTCTTCACTGTCGGACAATTCATTTTCTTATTTTGGACAATCAgcATTACTGAAGCC--
HPLC6 61 ACTTTTCACTGTCGGACAATTGATTTTCTTATTTTGGACAATGAgaATCACTGAAGCCAG
AFP9 61 ACTTTTCACTGTCGGACAATTGATTTTCTTATTTTGGACAATGAgaATCACTGAAGCCAA
F2-skin 61 ACTTTTCACTGTC----------GACCACTCAgaATCACTGACATCAA
Y-Tail 42 ACTCTTCACTGTTGGACAATTAATTTTCTTATTTTGGACACTCAGAATCACTGAAGCCAA
I D P A A R A A A A A A A S K A A V T A
Hyp-I 105 CATCGACCCCGCAGCCAGAGCCGCAGCAGCCGCAGCAGCCTCCAAAGCCGCAGTCACCGC
5a 120 -ATCGACCCCGCAGCCAAAGCCGCCGCAGCCGCAGCCGCAGCCACCGCAGTTGTCACTGC
HPLC6 121 ACCCGACCCCGCAGCCAAAGCCGCCCCAGCAGCAGCTGCC----------
AFP9 121 CCCCGACCCCGCAGCCAAAGCC----------
F2-skin 99 CATGGACGCACCAGCCAAAGCC----------
Y-Tail 102 CCCCGACCCCGCAGCCAAAGCCGCCCCAGCCGCAGTCGCC----------
A D A A A A A A T I A A S A A S V A A A
Hyp-I 165 CGCCGACGCTGCCGCAGCCGCTGCCACCATCGCCGCATCCGCCGCCTCAGTCGCTGCGGC
5a 178 CGCAGCCGCTGCCGCAGCCGCCGCCGCCATCGCCGCCACCGCCGCCGCAGTCGCTGGGGC
HPLC6 161 ----------
AFP9 143 ----------
F2-skin 121 ----------
Y-Tail 142 ----------
T A A D D A A A S I A T I N A A S A A A
Hyp-I 225 CACTGCTGCTGATGACGCCGCCGCATCCATAGCAACCATCAACGCCGCATCCGCTGCCGC
5a 238 CACCGCCGCCGATGCCGCCGCCGCATCCATAGCAAGCATCAACGCCAACACCGCCGCCGC
HPLC6 161 ----------
AFP9 143 ----------
F2-skin 121 ----------
Y-Tail 142 ----------
K S I A A A A A M A A K D T A A A A A S
Hyp-I 285 GAAATCCATCGCCGCCGCCGCAGCAATGGCAGCCAAGGACACTGCCGCCGCCGCGGCCAG
5a 298 AGCAGCCATCGCCGCCGCCGCAGCAAAAGCAGCCGAAGAAGCCGCCGCCACCGCCGCCGC
HPLC6 161 ----------
AFP9 143 ----------
F2-skin 121 ----------
Y-Tail 142 ----------
A A A A A V A S A A K A L E T I N V K A
Hyp-I 345 CGCCGCCGCCGCCGCCGTTGCCTCCGCCGCCAAAGCCCTAGAAACCATCAACGTCAAAGC
5a 358 CGCTGCCGCCACAACCGCCGCCACAGCCGCCACAGCCCAAGCAACCATCAAGGATAAAGC
HPLC6 161 ----------
AFP9 143 ----------
F2-skin 121 ----------
Y-Tail 142 ----------
A Y A A A T T A N T A A A A A A A T A T
Hyp-I 405 CGCATACGCCGCTGCCACCACCGCTAATACCGCTGCTGCCGCCGCCGCCGCCACCGCCAC
5a 418 GGCAGCCGCCGCAGCCTCCACCGCCACCAACGCCGCCGCGGCCGCGGCCGCCACCGCCAC
HPLC6 161 ----------GCCCC
AFP9 143 ----------GTCCC
F2-skin 121 ----------
Y-Tail 142 ----------GACCC
T A A A A A A A K A T I D N A A A A K A
Hyp-I 465 CACCGCCGCTGCCGCCGCCGCAGCCAAAGCAACCATCGACAACGCCGCCGCTGCCAAGGC
5a 478 CACCGCCGCAGCCGCCGCCGTAGCCAAAACAACCATCGACAAGGCCGCCGCCGCCGTGGC
HPLC6 166 TGCCGCAGCCGCCCC----------AGACACCGCCTCTGACGCCGCCGCTGCAGCCGC
AFP9 148 TGCCGCAGCCGCCCC----------AGACACCGCCTCTGATGCCGCCGCAGCAGCCGC
F2-skin 121 ----------GCCGC
Y-Tail 147 TGCCGCAGCCGCTGCCGCAGCCGTTGCAGACACCGCCTCTGACGCCGCCGCAGCAGCCGC
**
A A V A T A V S D A A A T A A T A A A V
Hyp-I 525 AGCCGCAGTCGCCACCGCCGTTTCAGATGCCGCAGCCACCGCCGCCACTGCCGCCGCCGT
5a 538 AGTTGCGGCCGCCACCGCCGTGGCAGCTGCCGCAGCCACCGCCGCCACTGCCGCCGCCAC
HPLC6 214 CCTTACCGCCGCCAACGCCAAAGCCGCTGCCGAA----------
AFP9 196 CGCCACCGCAGCCACCGCCGCCGCAGCAGCCGCCGCCACCGCAGTCACCGCTGCCAAAGC
F2-skin 126 AGCCACCGCCGCCGCCGCCAAGGCCGCCGCAGAA----------
Y-Tail 207 CGCCACCGCCGCTGCCGCCGCTAAAGCCGCAGCCGACACCGCCGCTGCCGCCGCTAAAGC
A A A T L E A A A A K A A A T A V S A A
Hyp-I 585 TGCCGCTGCAACCCTCGAAGCTGCCGCCGCAAAAGCCGCAGCCACTGCAGTCTCTGCCGC
5a 598 CGCTGCCGCAACCCTCGGGGCTGCCACAGTAAAAGCCGCAGCCACCGCCGTCAATGCCGC
HPLC6 248 -------CTCACTGCCGCCAACGCCGCCGCCGCCGCAGCAGCCACC----------GC
AFP9 256 CGCAGCCCTCACCGCCGCCAACGCCGCCGCCGCCGCAGCAGCCACCGCCGCCGCAGCCGC
F2-skin 160 -------GCCACCGCCGCCGCAGCTGCCAAAGCAGCAGCCGCCACCAAAGCCGGAGCAGC
Y-Tail 267 CGCAGCCGACACCGCCGCTGCCGCCGCTGAAGCCGCCGCAGCCACC----------GC

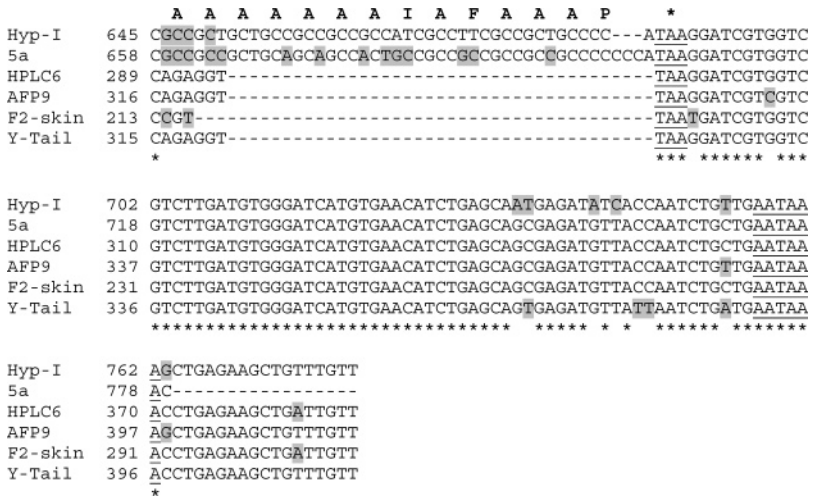


FIGURE 5: Alignment of the cDNA sequences of the AFPs shown in Figure 4. Positions that are identical in all sequences are marked with an asterisk, and differences are highlighted in gray. Dashes represent gaps, bold italics represent known transcription start sites, start and stop codons as well as polyadenylation signals are underlined and intron sites (arrow) are flanked by lowercase letters. The amino-acid sequence of hyp-type I is indicated above the cDNA sequence. Accession numbers are given in Table 3. Features such as transcription start sites, polyadenylation sites and intron positions were either previously determined (*21*, *24*) or predicted based on sequence similarity (*34*).

The MS/MS spectrum of the C-terminal fragment (data not shown) was dominated by b-ion and internal-ion series. However, y1 and y2 ions corresponding in mass to P and AP were seen, consistent with a lack of C-terminal modification. Taken together, this indicates that the only post-translational modification to which hyp-type I is subjected is cleavage of the signal peptide.

An additional fragment with a weaker MS signal and mass inconsistent with the cDNA was also partially sequenced by MS/MS (Table 1). The mass discrepancy corresponds to a single substitution of an Ala for a Thr residue at residue 85. The MALDI-MS spectrum (not shown) of protein purified from pooled sera from a number of individual winter flounders suggests the presence of small amounts of both higher and lower molecular weight species from which variant fragments could have arisen.

Biased Base Composition of the Coding Sequence. The GC content of the hyp-type I clone is high (64.4% for the entire clone and 72.4% for the coding sequence of the mature protein), explaining why PCR reactions were only successful when an additive (Q-solution) designed to amplify GC-rich templates was employed. This is consistent with the high GC content of the type I AFPs and is due to the high Ala content (62% in the mature protein, Table 2), exacerbated by the preferential use of the GCC codon (63%) for this amino acid (Table 3). Conversely, the GCG codon is almost never used. A comparison was made to eight randomly selected, mutually nonhomologous sequences encoding longer (>300 amino acids) non-antifreeze proteins from winter flounder (Table 3). A similar bias was observed in these sequences, though it was less extreme. The GCC codon was favored in six of the eight sequences, and the GCG codon was used least in the same number of sequences.

Biased codon usage in the AFPs is not restricted to the Ala codons. Interestingly, there is a global bias against G in the third codon position of the AFP genes, which is evident even when Ala is excluded from the analysis (Met and Trp, which are each encoded by a single codon were also excluded). This bias was not observed in the non-antifreeze genes. A slight bias is found toward C in the third position of the non-Ala codons of the AFPs, but it is not unusual when compared with the non-AFPs.

Polymorphisms. The amplified sequence reported represents a single cDNA clone by the criteria described above (Figure 3, clone 3), and it exactly matches the sequences of two partial clones, amplified in separate reactions (Figure 3, clones 1 and 2), that together span the entire coding sequence. Two additional clones, obtained in independent amplifications, shared 6 polymorphisms within the 3′ UTR (data not shown). This is consistent with cDNAs arising from two alleles since the cDNA library was made to RNA from one liver.

The *5a* sequence was not found in any of the clones sequenced. To further investigate whether *5a* was present in the library, three primers were designed to selectively amplify its cDNA. One primer had a single mismatch relative to hyp-type I, and the other two had three and five mismatches, with one at the second base from the 3′ end of each. A DNA product was only obtained with one pair of primers (those with one and three mismatches), and its sequence matched the hyp-type I cDNA, strongly suggesting that *5a* transcripts were absent from the liver cDNA library.

Hyp-type I Is a Homologue of the Type I AFPs and 5a. The deduced signal peptides of hyp-type I and *5a* AFPs are absolutely identical whereas the mature sequences are only 73.5% identical (Figure 4). Differences include the gain or loss of a single residue at either terminus, as well as numerous substitutions distributed throughout the length of the mature protein. The deduced transcripts are 83.7% identical over their length (Figure 5), with a 2.1-fold higher rate of synonymous substitutions over nonsynonymous substitutions (*38*), consistent with selection against changes to the amino acid sequence. One interesting discrepancy is that the coding sequence (585 bases) of the mature protein is 80.5% identical, whereas the remaining 178 bases (5′ UTR,

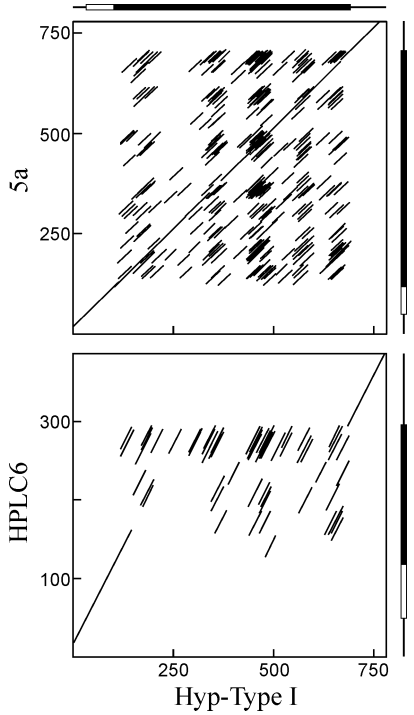


FIGURE 6: Dot matrix comparison of cDNA sequences of selected AFPs. Matches of at least 16 out of 20 nucleotides are indicated by diagonal lines. A schematic of each cDNA is on each axis with lines indicating untranslated regions, white rectangles indicating the signal-peptide encoding region and black rectangles indicating the rest of the coding sequence.

signal peptide-encoding region and 3′ UTR) are 94% identical.

Similar comparisons between the DNA sequences of hyp-type I and the smaller type I AFPs (Figures 4−6) reveal that they also share >90% identity throughout the 3′ UTR. In addition, all but the skin isoforms are highly similar throughout the 5′ UTR and the region encoding the signal peptide. Although the skin isoforms lack a signal peptide, 18 out of 19 bases extending upstream from the transcription start site match a region of the 5′ UTR of hyp-type I. A second match occurs between 15 bases of the 5′ UTR of the skin isoform and the signal peptide of hyp-type I. The position of the intron, near the end of the signal peptide-encoding region (Figure 4), is conserved between the *5a* gene and the liver isoforms.

Alignment of the mature coding sequences of the short and long AFPs (Figure 4) is difficult for the following reasons. First, the 11-amino-acid repeats of hyp-type I are not as clearly defined as those of the shorter isoforms. Second, the sequences are so Ala-rich that the scores for many possible alignments, shifted by three base intervals, are comparable. This can be easily seen in dot matrix comparisons (Figure 6) where numerous matches of 80% or higher over 20 bases are observed, in contrast to the flanking regions which produce single distinct lines. Modification of the dot matrix parameters (data not shown) does not improve the alignment of the mature coding sequences.

A Theoretical Model for the Structure of Hyp-type I AFP. There are two aspects to the model that will be separately presented. One is the nature of the ice-binding site, which we think can be confidently predicted from the knowledge that a monomer of hyp-type I is a continuous alpha-helix. The psi-phi angles of −62° and −42° are such that the undecad repeats make exactly three helical turns placing the ice-binding site(s) on the same face of the helix. The other is the nature of the dimerization interaction/interface, which is more speculative and will be presented in the Discussion section.

The Monomer Presents a Series of Ala-Rich Ice-Binding Sites on one Face. The circular dichroism (CD) spectrum of hyp-type I AFP is consistent with an almost completely (>95%) alpha-helical protein (*32*). The two residues with a high propensity for disrupting α-helical structure are either absent (Gly) or found at, or near, the termini of the protein (Pro). The Pro found near the N terminus (position 4), as well as the charged residues at positions 3 and 7, may be involved in the formation of a "capping" structure, analogous to that modeled in the shorthorn sculpin (*15*). The other Pro is the C-terminal residue, which would not disrupt the structure of a long helix. Not surprisingly, most secondary structure prediction programs, such as Jnet (*39*) and nnpredict (*40*), predict a remarkably long continuous helix, with a few nonhelical residues at each end, in complete agreement with the CD results.

Analytical ultracentrifugation indicated that native hyp-type I AFP formed a highly asymmetrical (18:1 axial ratio) dimer, with dimensions roughly consistent with two extended 195-amino-acid-long helices aligned side-by-side (*32*). Helices frequently dimerize along a hydrophobic interface to form coiled coils. Programs such as Marcoil and PCOILS, which have been shown to efficiently detect left-handed coiled coils which typically have a seven residue (heptad) periodicity (*41*), did not detect this motif in hyp-type I AFP. Right-handed coiled coils typically display an 11-residue (undecad) periodicity (*42*). Undecad periodicity is evident in hyp-type I with Thr and/or Ile appearing at 11-residue intervals (Figures 4 and 7). Moreover, it is consistent with our model of a single alpha-helix that has 11 residues every three turns, similar to that observed with type I AFP (*23*). The only inconsistency in the undecad repeat occurs about two-thirds of the way through the sequence (indicated with slashes in Figure 4) where an additional helix loop turn has been modeled (indicated with an arrow in Figure 7a). The first four residues (NIDP) as well as the last five (FAAAP), which may mediate helix capping structures, are not considered further.

The model of the hyp-type I monomer does not have one long continuous Ala-rich surface, but rather it contains several regions (boxed in Figure 7a) that strongly resemble the ice-binding sites of the small type I AFPs in length and structure, some of which are shown alongside for comparison. The ice-binding sites of sculpin (*15*) and flounder type I AFPs (*14*, *19*) are faces of the helix composed exclusively of Ala residues that abut a rank of regularly spaced i, $i + 11$ residues with larger side chains. In the case of flounder type I AFP, these i, $i + 11$ residues are all Thr and they form a linear array. In the shorthorn sculpin AFP, the i, $i + 11$ positions are occupied by Thr, Ile or Leu. Even a single non-Ala residue projecting from the ice-binding surface can eliminate antifreeze activity (*14*, *15*). In hyp-type I, residues 6 to 43 form helical turns 2 to 12, which would have a face composed entirely of Ala that abuts a rank of Thr and Ile residues. This region resembles the Ala-rich ice-binding surface of the sculpin AFPs, and is two turns longer than

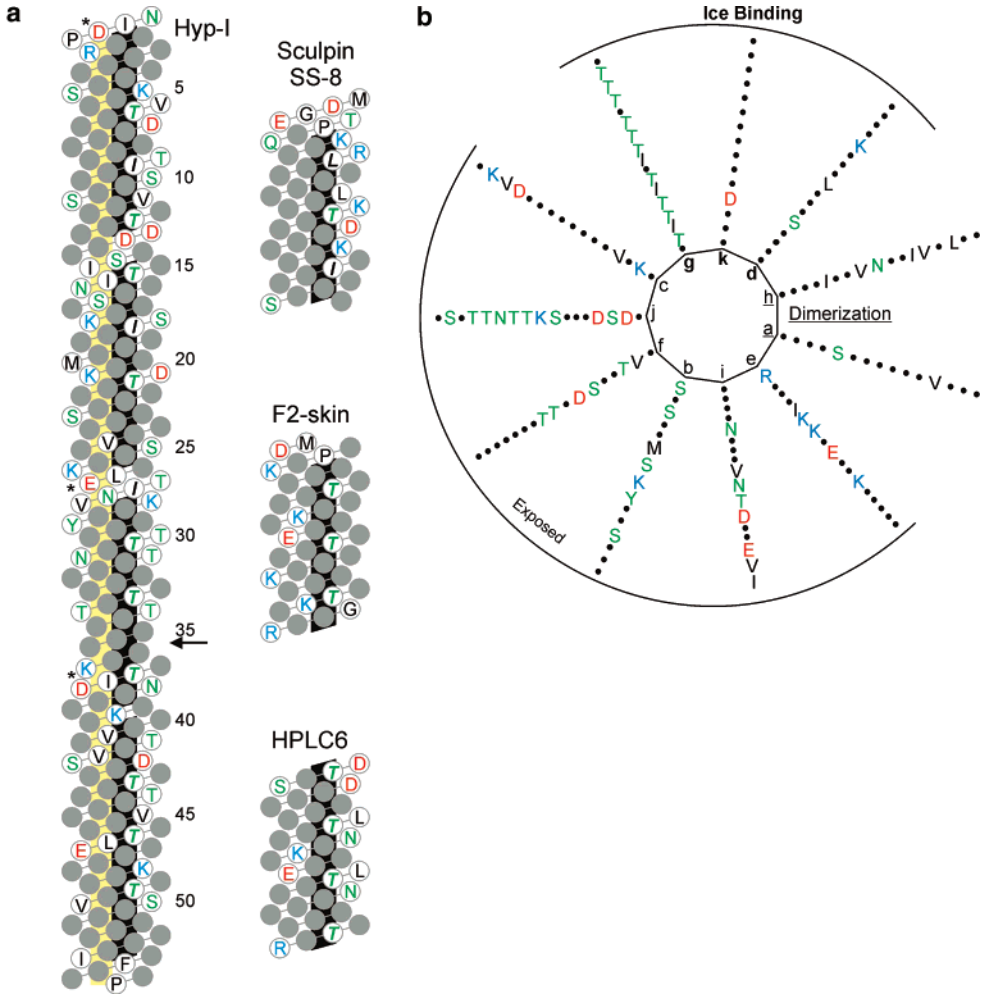


FIGURE 7: Schematic comparisons of hyp-type I and type I AFPs. (a) Helical net with Ala residues indicated by gray circles and other residues indicated by single-letter amino-acid code. The letters are colored by property with black = nonpolar, green = polar, red = acidic, blue = basic. The 11th residue of each repeat (as indicated in Figure 4) is in bold italics and lies in the first row of 4 residues in each pattern of 4:4:3. The black background bars represent known or proposed ice-binding surfaces, and the yellow bar represents the proposed dimerization surface. An arrow indicates the position of the extra helical turn which disrupts the undecad repeat pattern, and asterisks denote potential salt bridges. Helical turns are numbered from the N-terminus. Accession numbers for all but sculpin SS8 (*1*, *15*) are given below Table 3. (b) Helical wheel representation with 11 amino acids for every 3 turns. Ala residues are indicated by dots and other residues indicated as above. Several residues at each terminus (NIDP and FAAAP) as well as the 6 Ala residues indicated by slashes in Figure 4 were omitted.

SS-8, the longest known sculpin isoform with nine turns. The model contains another 11-turn Ala-rich face (turns 15 to 25) which, like the sculpin AFP, is adjacent to a rank of Thr and Ile residues. In the C-terminal portion of hyp-type I, the 12-turn segment composed of turns 27 to 38 forms an Ala-rich face bounded mainly by Thr and another Ala-rich segment of 13 turns (from 40 to 52) is bounded exclusively by Thr. These putative ice-binding faces in the C-terminal half of hyp-type I look very much like those of the type I AFPs produced by the righteye flounders. Winter flounder and shorthorn sculpin AFPs bind to different planes of ice: winter flounder type I AFP binds to the pyramidal {20−21} planes, and the sculpin AFPs bind to the secondary prism planes {2−1−10} (*20*). The presence of distinct ice-binding patches with affinity for different planes of ice may be the basis for the unusual crystal morphology and hyperactivity of hyp-type I AFP. Indeed, the homodimeric quaternary structure of hyp-type I suggests the possibility of presenting compound ice-binding sites assembled with components from the two adjacent helices.

In our model of the hyp-type I AFP dimer, the two helices have an antiparallel arrangement without any stagger. A section of the dimer is illustrated in Figure 8A showing the interdigitation of the Ala residues in the top down view and in the cross-sectional end-on view below (Figure 8B). In the latter view the Ala-rich ice-binding face is flanked by Thr on either side. Charged hydrophilic residues are present on the other surfaces exposed to the solvent.

DISCUSSION

There are presently only three hyperactive AFPs that have been structurally characterized, all of which are from arthropods (*43*−*45*). A fourth hyperactive AFP has been found in an Antarctic bacterium but has not yet been characterized (*46*). This is the first report of a sequence of a hyperactive AFP from a fish, and based on the sequencing described here we recognize the flounder hyp-type I protein as an extreme isoform of the small type I AFPs.

There are many examples where large isoforms of an AFP have more activity than small isoforms of the same type.

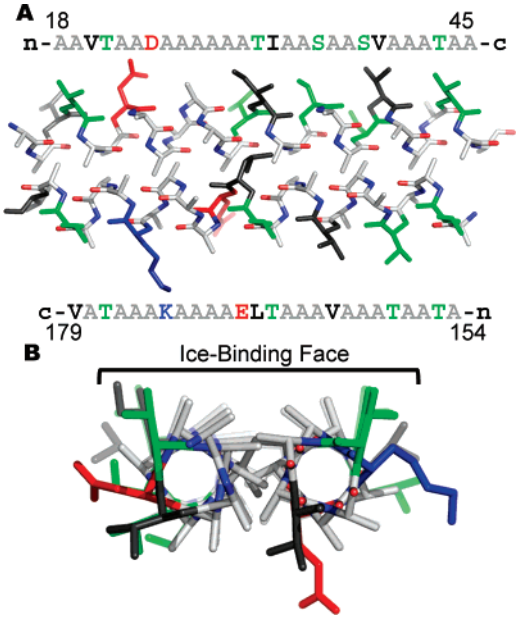


FIGURE 8: (A) Model of the dimerization interface. A segment of the antiparallel dimer (from residues 18 to 45 and residues 154 to 179) is represented in stick format showing the interdigitated alanines that form the dimerization interface. The predicted ice-binding site is facing upward in the plane of the page. N and O atoms are colored blue and red, respectively. Specific residue types are colored: Thr and Ser (green); Lys (blue); Asp (red); Val, Leu and Ile (black). (B) Cross-sectional end-on view of the dimer showing the two rows of Ala that form the dimerization contact and the regularly spaced Thr and Ala on the ice-binding site (marked by a bar).

This is true for the AFGPs, where AFGPs 1−5 are much more active than AFGPs 6−8 (*47*) for type I AFPs where the four-repeat AFP9 is more active than HPLC-6 (*48*); and for type III AFP where the fusion of antifreeze proteins greatly increases activity provided all domains can interact with ice (*49*, *50*). The same trend has been observed for hyperactive AFPs. A slightly longer isoform of the beta-helical spruce budworm AFP is appreciably more active than one with two fewer turns of helix (*51*), and addition of beta-helical coils to *Tenebrio molitor* AFP similarly increases its activity (*52*). However, type I is the only AFP known for which one particular isoform is hyperactive whereas the others have moderate activity. Since it is not clear what structural features make an AFP hyperactive, a detailed comparison of these isoforms might reveal attributes that allow an AFP to produce extremely high thermal hysteresis activity at low concentration.

Ancestry and Evolution. Determination of the cDNA sequence of hyp-type I AFP provided the key insight into the evolutionary relationship between this AFP and the smaller type I AFPs. Homology could not be conclusively demonstrated using the mature protein sequences as they have radically different sizes (195 vs 39−52 amino acids) and quaternary structures, and since they could not be reliably aligned due to the divergence and repetitive nature of the sequences. This is despite their shared Ala−rich amino acid compositions, codon bias, undecad repeat motifs (less conserved in hyp-type I) and ice-binding function. However, conservation of the 5′ and 3′ UTRs, as well as the signal peptide-encoding region, clearly indicates that these genes are homologous. The 5′ ends of the skin isoforms differ from hyp-type I as they do not encode a signal peptide and have a transcription start site distinct from the liver isoforms, HPLC-6 and AFP9 (*21*, *24*). Despite this, it is evident that hyp-type I is also related to the skin isoforms as their 3′ UTRs are more than 90% identical. Another interesting difference between hyp-type I and the shorter liver isoforms is the lack of processing of hyp-type I AFP after removal of the signal peptide. A series of dipeptidyl peptidase consensus sequences (X-Ala/Pro) are found downstream of the signal peptide sequence in type I plasma AFPs derived from the liver. Consequently, the primary transcript encodes a pre-pro-sequence (*21*). In hyp-type I, the second residue following the cleavage site is Ile, which blocks processing by this peptidase. As well, the final Gly residue of type I AFP is removed during C-terminal amidation, which helps cap and stabilize that end of the helix (*23*). Hyp-type I terminates with an unmodified Pro residue.

The disparity between the conservation of the mature coding and noncoding regions of the different type I AFPs is reminiscent of the differences observed between the AFGPs of Antarctic notothenioids and the precursor gene, trypsinogen (*53*). In this case, expansion of a 9-nucleotide sequence within the precursor gave rise to the repetitive AFGP. A similar process is likely to have occurred with flounder AFPs. The presence of undecad repeats as well as the Ala richness with preferential use of a single codon may have facilitated rapid alterations of repeat number, similar to that observed for repeats in the human genome (*54*). It remains to be determined however which AFP is the ancestral protein and whether (a) the 17-kDa AFP is an expanded version of type I AFP or (b) type I AFP represents a truncated version of the long AFP. In scenario (a), dimerization may have been acquired to support the extended helices, whereas in (b), type I AFPs may have lost the requirement for dimerization as they evolved by truncation into shorter, more stable monomers. The finding that the American plaice, a closely related righteye flounder, produces only a large hyperactive AFP and not the smaller type I AFPs in its blood plasma (*31*) has been used to argue that the hyperactive variant might be the ancestral AFP.

A genomic DNA sequence (*5a*) was obtained previously (*34*) with a deduced coding sequence over 80% identical to hyp-type I cDNA. At that time we considered *5a* to be a pseudogene. The recent discovery of the hyp-type I isoform warranted a reexamination of this matter. We have not detected any evidence of expression of the *5a* gene, either as a plasma protein or as a liver mRNA. This supports the notion that *5a* is indeed a pseudogene, however it is possible that it is expressed in a tissue other than the liver (the source of our library), or that the gene is not present in all winter flounders due to population polymorphisms. It is not likely that *5a* is a skin isoform because the gene contains a signal sequence for export from the cell. If it is produced and exported, it should have been be detected in the blood, although the putative protein encoded by *5a* could conceivably be even more thermolabile and susceptible to degradation than hyp-type I AFP due to the presence of two internal helix-breaking Gly residues.

Ice-Binding Sites of Hyp-type I AFP and Hyperactivity. Hyp-type I AFP is 195 residues long, and its amino acid composition is >60% Ala. The secondary structure of the sequence is predicted to be almost entirely helical, which is in agreement with the circular dichroism analysis (*32*). This

protein has TH activity comparable to that of the hyperactive insect AFPs (*30*, *55*), yet its structure is clearly distinct from the beta-helical folds of the beetle and spruce budworm AFPs (*43*, *44*) and the proposed polyproline type II alternating coil fold of the snow flea AFP (*45*). It is becoming apparent that more than one protein scaffold can confer hyperactive antifreeze activity.

The analytical ultracentrifugation analyses of the hyp-type I AFP demonstrated that it is dimeric and highly extended (*32*). In our model of the monomer there are several type I AFP ice-binding sites in series down the length of the helix. It is likely that the two helices are oriented in the dimer in such a way that they both present their binding sites to the ice. This would double the number of ice-binding sites and the surface area in contact with ice. The large size of the dimer (32k) and its potentially large ice-binding face is consistent with the relationship between increased AFP size and increased activity described above. In each of these instances the area of the ice-binding site is the critical determinant of activity rather than size of the protein. Even small (30 − 40%) increases in the area of the ice-binding site can result in several-fold increases in thermal hysteresis as a function of AFP concentration. However, the tremendous activity of hyp-type I AFP relative to other fish AFPs is probably a function of more than just increased size. The complex spindle-shaped morphology of ice crystals produced by this AFP suggests that the dimer might bind to more than one plane of ice. Some of its linked ice-binding sites resemble those of the shorthorn sculpin type I (*15*), and others are more like those of the flounder type I (*14*, *19*). Whatever combination of ice surfaces are protected, hyp-type I's hyperactivity probably stems from its ability to stop growth of the crystal along the *c*-axis (*29*).

Model for the Dimerization of the Two Helices. The most likely arrangement for the dimer would have the two helices associate along their length as an antiparallel Ala-zipper (Figure 8A). We reject the possibility that the two helices form a coiled coil for two reasons. (1) There is no sign of a heptad repeat of Leu or Ile. Dimerization of helices in a coiled coil typically occurs on an apolar surface enriched for aliphatic amino acids such as Leu or Ile. However, these larger hydrophobic residues are not present in sufficient quantity or with the correct spacing in hyp-type I AFP. (2) Coiled coils typically have a twist (supercoil) as one coil passes over the other. A twist to the dimer would interfere with the uniform presentation of the series of flat ice-binding sites that run the length of the 195-residue helix. These sites would become obscured by the other helix and/or be sterically removed from proximity to the ice.

Instead, the hyp-type I AFP has been modeled as an antiparallel dimer composed of two straight helices for the following reasons. This arrangement is consistent with other undecad repeat helices, which have just under 3.7 residues per turn as exemplified by the tetramerization domain of tetrabrachion (*35*) and type I AFP (*23*). These helices are slightly undercoiled such that every eleventh residue, rather than every eighteenth, lies in the same position along the helical axis. The subunits of the tetramerization domain of tetrabrachion run parallel for 10 turns of the alpha-helix without any rotation about the axis (supercoiling).

The helices of hyp-type I AFP were assumed to have opposite polarity (antiparallel) because this is the only arrangement that allows both subunits to simultaneously present their ice-binding faces to a single ice surface while permitting dimerization to occur in a "symmetrical" manner through the same face of each subunit. On the basis of these assumptions, there are two distinct surfaces available for dimerization, a hydrophilic surface, and a second Ala-rich surface flanking the putative ice-binding site (Figure 7b). The hydrophilic surface most likely faces the solvent to improve the solubility of this otherwise relatively hydrophobic protein. This leaves the Ala-rich surface adjacent to the ice-binding site as the dimerization interface.

We envisage the dimer interaction to be that of an Alacoil or Ala zipper (*56*) with the Ala side chain methyl groups interdigitating along the length of the dimer. The tighter packing facilitated by this small side chain allows the two helices to approach each other more closely (∼8 Å between helix axes) than they would in a coiled coil. Thus the dimer is 275 Å in length and approximately 16 Å along the major cross-sectional axis, giving an axial ratio of 17, which is consistent with the experimentally determined value (*32*). The axial ratio suggests that the two helices assemble with no appreciable horizontal translation that would result in unpaired overhangs. In our model the helices are annealed along the entire length to maximize the area of the dimerization interface and hence the strength of binding. The lability of the dimer to temperature is consistent with the relatively weak interactions between the Ala side chains (*57*).

CONCLUSIONS

We have cloned and sequenced the cDNA of the first hyperactive AFP found in a fish. We have designated it hyp-type I, as the sequence clearly shows that it is homologous to the type I AFPs. The 5′ and 3′ UTRs and signal peptides are highly conserved; however alignment of the mature proteins is complicated by the repetitive, Ala-rich sequence.

The sequence of hyp-type I indicates that it is distinct from the putative pseudogene, *5a*. On the basis of our sequence we propose a model of hyp-type I in which the polypeptide forms an extraordinarily long (∼190 aa) helix that presents many Ala-rich patches, each of which resembles the ice-binding faces of smaller type I AFPs. These surfaces could be independent ice-binding motifs and recognize distinct planes of ice, contributing to hyperactivity. Finally, we propose that this AFP dimerizes through an Ala-zipper motif in an antiparallel orientation such that both subunits can simultaneously engage an ice surface.

ACKNOWLEDGMENT

We are grateful to Dr. David Hyndman and David McLeod of the Protein Function Discovery Facility at Queen's University for mass spectrometry analyses, to Sherry Gauthier for technical assistance and to Garth Fletcher of the Ocean Sciences Centre, Memorial University of Newfoundland for the gift of winter flounder plasma.

REFERENCES

1. Cheng, C. C., and DeVries, A. L. (1991) The role of antifreeze glycopeptides and peptides in the freezing avoidance of cold-water fish, in *Life Under Extreme Conditions* (de Prisco, D., Ed.) pp 1−14, Springer-Verlag, Berlin.
2. Ewart, K. V., Lin, Q., and Hew, C. L. (1999) Structure, function and evolution of antifreeze proteins, *Cell. Mol. Life Sci. 55*, 271−283.

3. 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.
4. Fletcher, G. L., Hew, C. L., and Davies, P. L. (2001) Antifreeze proteins of teleost fishes, *Annu. Rev. Physiol. 63*, 359−390.
5. Davies, P. L., and Hew, C. L. (1990) Biochemistry of fish antifreeze proteins, *FASEB J. 4*, 2460−2468.
6. Harding, M. M., Ward, L. G., and Haymet, A. D. (1999) Type I 'antifreeze' proteins. Structure-activity studies and mechanisms of ice growth inhibition, *Eur. J. Biochem. 264*, 653−665.
7. Duman, J. G., and DeVries, A. L. (1974) Freezing resistance in winter flounder, *Nature 274*, 237−238.
8. Scott, G. K., Davies, P. L., Shears, M. A., and Fletcher, G. L. (1987) Structural variations in the alanine-rich antifreeze proteins of the pleuronectinae, *Eur. J. Biochem. 168*, 629−633.
9. Hew, C. L., Joshi, S., Wang, N. C., Kao, M. H., and Ananthanarayanan, V. S. (1985) Structures of shorthorn sculpin antifreeze polypeptides, *Eur. J. Biochem. 151*, 167−172.
10. Low, W. K., Miao, M., Ewart, K. V., Yang, D. S., Fletcher, G. L., and Hew, C. L. (1998) Skin-type antifreeze protein from the shorthorn sculpin, *Myoxocephalus scorpius*. Expression and characterization of a Mr 9,700 recombinant protein, *J. Biol. Chem. 273*, 23098−23103.
11. Low, W. K., Lin, Q., Stathakis, C., Miao, M., Fletcher, G. L., and Hew, C. L. (2001) Isolation and characterization of skin-type, type I antifreeze polypeptides from the longhorn sculpin, *Myoxocephalus octodecemspinosus*, *J. Biol. Chem. 276*, 11582−11589.
12. Evans, R. P., and Fletcher, G. L. (2001) Isolation and characterization of type I antifreeze proteins from Atlantic snailfish (*Liparis atlanticus*) and dusky snailfish (*Liparis gibbus*), *Biochim. Biophys. Acta 1547*, 235−244.
13. Evans, R. P., and Fletcher, G. L. (2004) Isolation and purification of antifreeze proteins from skin tissues of snailfish, cunner and sea raven, *Biochim. Biophys. Acta 1700*, 209−217.
14. Baardsnes, J., Kondejewski, L. H., Hodges, R. S., Chao, H., Kay, C., and Davies, P. L. (1999) New ice-binding face for type I antifreeze protein, *FEBS Lett. 463*, 87−91.
15. Baardsnes, J., Jelokhani-Niaraki, M., Kondejewski, L. H., Kuiper, M. J., Kay, C. M., Hodges, R. S., and Davies, P. L. (2001) Antifreeze protein from shorthorn sculpin: Identification of the ice-binding surface, *Protein Sci. 10*, 2566−2576.
16. Mao, Y., and Ba, Y. (2006) Insight into the binding of antifreeze proteins to ice surfaces via ^{13}C spin lattice relaxation solid-state NMR, *Biophys. J. 91*, 1059−1068.
17. 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.
18. Zhang, W., and Laursen, R. A. (1998) Structure-function relationships in a type I antifreeze polypeptide. The role of threonine methyl and hydroxyl groups in antifreeze activity, *J. Biol. Chem. 273*, 34806−34812.
19. Wierzbicki, A., Dalal, P., Cheatham, T. E., III, Knickelbein, J. E., Haymet, A. D., and Madura, J. D. (2007) Antifreeze proteins at the ice/water interface: Three calculated discriminating properties for orientation of type I proteins, *Biophys. J. 93*, 1442−1451.
20. Knight, C. A., Cheng, C. C., and DeVries, A. L. (1991) Adsorption of alpha-helical antifreeze peptides on specific ice crystal surface planes, *Biophys. J. 59*, 409−418.
21. Pickett, M., Scott, G., Davies, P., Wang, N., Joshi, S., and Hew, C. (1984) Sequence of an antifreeze protein precursor, *Eur. J. Biochem. 143*, 35−38.
22. Hew, C., Liunardo, N., and Fletcher, G. L. (1978) *In vivo* biosynthesis of the antifreeze protein in the winter flounder: Evidence for a larger precursor, *Biochem. Biophys. Res. Commun. 85*, 421−427.
23. Sicheri, F., and Yang, D. S. (1995) Ice-binding structure and mechanism of an antifreeze protein from winter flounder, *Nature 375*, 427−431.
24. Gong, Z., Ewart, K. V., Hu, Z., Fletcher, G. L., and Hew, C. L. (1996) Skin antifreeze protein genes of the winter flounder, *Pleuronectes americanus*, encode distinct and active polypeptides without the secretory signal and prosequences, *J. Biol. Chem. 271*, 4106−4112.
25. Murray, H. M., Hew, C., Kao, K. R., and Fletcher, G. L. (2002) Localization of cells from the winter flounder gill expressing a skin-type antifreeze protein gene, *Can. J. Zool. 80*, 110−119.
26. Murray, H. M., Hew, C. L., and Fletcher, G. L. (2003) Spatial expression patterns of skin-type antifreeze protein in winter flounder (*Pseudopleuronectes americanus*) epidermis following metamorphosis, *J. Morphol. 257*, 78−86.
27. Kristiansen, E., and Zachariassen, K. E. (2005) The mechanism by which fish antifreeze proteins cause thermal hysteresis, *Cryobiology 51*, 262−280.
28. Walker, V. K., Kuiper, M. J., Tyshenko, M. G., Doucet, D., Graether, S. P., Liou, Y. C., Sykes, B. D., Jia, Z., Davies, P. L., and Graham, L. A. (2001) Surviving winter with antifreeze proteins: Studies on budworms and beetles, in *Insect Timing: Circadian Rhythmicity to Seasonality* (Denlinger, D. L., Giebultowicz, J., and Saunders, D. S., Eds.) pp 199−212, Elsevier Science, Amsterdam.
29. 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.
30. Marshall, C. B., Fletcher, G. L., and Davies, P. L. (2004) Hyperactive antifreeze protein in a fish, *Nature 429*, 153.
31. Gauthier, S. Y., Marshall, C. B., Fletcher, G. L., and Davies, P. L. (2005) Hyperactive antifreeze protein in flounder species. The sole freeze protectant in American plaice, *FEBS J. 272*, 4439−4449.
32. Marshall, C. B., Chakrabartty, A., and Davies, P. L. (2005) Hyperactive antifreeze protein from winter flounder is a very long rod-like dimer of alpha-helices, *J. Biol. Chem. 280*, 17920−17929.
33. Lin, Q., Ewart, K. V., Yang, D. S., and Hew, C. L. (1999) Studies of a putative ice-binding motif in winter flounder skin-type antifreeze polypeptide, *FEBS Lett. 453*, 331−334.
34. Davies, P. L., and Gauthier, S. Y. (1992) Antifreeze protein pseudogenes, *Gene 112*, 171−178.
35. Stetefeld, J., Jenny, M., Schulthess, T., Landwehr, R., Engel, J., and Kammerer, R. A. (2000) Crystal structure of a naturally occurring parallel right-handed coiled coil tetramer, *Nat. Struct. Biol. 7*, 772−776.
36. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA.
37. Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0, *J. Mol. Biol. 340*, 783−795.
38. Korber, B. (2000) HIV sequence signatures and similarities, in *Computational and Evolutionary Analysis of HIV Molecular Sequences* (Rodrigo, A. G., and Learn, G. H., Eds.) pp 55−72, Kluwer Academic Publishers, Dordrecht, The Netherlands.
39. Cuff, J. A., and Barton, G. J. (2000) Application of multiple sequence alignment profiles to improve protein secondary structure prediction, *Proteins 40*, 502−511.
40. Kneller, D. G., Cohen, F. E., and Langridge, R. (1990) Improvements in protein secondary structure prediction by an enhanced neural network, *J. Mol. Biol. 214*, 171−182.
41. Gruber, M., Söding, J., and Lupas, A. N. (2006) Comparative analysis of coiled-coil prediction methods, *J. Struct. Biol. 155*, 140−145.
42. Burkhard, P., Stetefeld, J., and Strelkov, S. V. (2001) Coiled coils: A highly versatile protein folding motif, *Trends Cell Biol. 11*, 82−88.
43. Graether, S. P., Kuiper, M. J., Gagne, S. M., Walker, V. K., Jia, Z., Sykes, B. D., and Davies, P. L. (2000) Beta-helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect, *Nature 406*, 325−328.
44. Liou, Y. C., Tocilj, A., Davies, P. L., and Jia, Z. (2000) Mimicry of ice structure by surface hydroxyls and water of a beta-helix antifreeze protein, *Nature 406*, 322−324.
45. 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.
46. Gilbert, J. A., Davies, P. L., and Laybourn-Parry, J. (2005) A hyperactive, Ca^{2+}-dependent antifreeze protein in an Antarctic bacterium, *FEMS Microbiol. Lett. 245*, 67−72.
47. Schrag, J. D., O'Grady, S. M., and DeVries, A. L. (1982) Relationship of amino acid composition and molecular weight of antifreeze glycopeptides to non-colligative freezing point depression, *Biochim. Biophys. Acta 717*, 322−326.
48. Chao, H., Hodges, R. S., Kay, C. M., Gauthier, S. Y., and Davies, P. L. (1996) A natural variant of type I antifreeze protein with four ice-binding repeats is a particularly potent antifreeze, *Protein Sci. 5*, 1150−1156.
49. Baardsnes, J., Kuiper, M. J., and Davies, P. L. (2003) Antifreeze protein dimer: When two ice-binding faces are better than one, *J. Biol. Chem. 278*, 38942−38947.

50. Nishimiya, Y., Ohgiya, S., and Tsuda, S. (2003) Artificial multimers of the type III antifreeze protein. Effects on thermal hysteresis and ice crystal morphology, *J. Biol. Chem. 278*, 32307−32312.
51. Leinala, E. K., Davies, P. L., Doucet, D., Tyshenko, M. G., Walker, V. K., and Jia, Z. (2002) A beta-helical antifreeze protein isoform with increased activity. Structural and functional insights, *J. Biol. Chem. 277*, 33349−33352.
52. Marshall, C. B., Daley, M. E., Sykes, B. D., and Davies, P. L. (2004) Enhancing the activity of a beta-helical antifreeze protein by the engineered addition of coils, *Biochemistry 43*, 11637−11646.
53. 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.
54. Jeffreys, A. J., Wilson, V., and Thein, S. L. (1985) Hypervariable 'minisatellite' regions in human DNA, *Nature 314*, 67−73.
55. 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.
56. Gernert, K. M., Surles, M. C., Labean, T. H., Richardson, J. S., and Richardson, D. C. (1995) The Alacoil: A very tight, antiparallel coiled-coil of helices, *Protein Sci. 4*, 2252−2260.
57. Zhou, N. E., Kay, C. M., and Hodges, R. S. (1992) Synthetic model proteins: The relative contribution of leucine residues at the nonequivalent positions of the 3−4 hydrophobic repeat to the stability of the two-stranded alpha-helical coiled-coil, *Biochemistry 31*, 5739−5746.
58. Wilkins, M. R., Lindskog, I., Gasteiger, E., Bairoch, A., Sanchez, J. C., Hochstrasser, D. F., and Appel, R. D. (1997) Detailed peptide characterization using PEPTIDEMASS: A World-Wide-Web-accessible tool, *Electrophoresis 18*, 403−408.

BI7020316