# Amino acid sequence of a new type of antifreeze protein, from the longhorn sculpin *Myoxocephalus octodecimspinosis*

Gejing Deng[a], David W. Andrews[b], Richard A. Laursen[a,*]

[a]*Department of Chemistry, Boston University, Boston, MA 02215, USA*
[b]*Immulogic Pharmaceutical Corporation, Waltham, MA 02154, USA*

Received 9 December 1996

**Abstract A new type of fish antifreeze protein, designated here type IV, has been isolated from the longhorn sculpin, *Myoxocephalus octodecimspinosis*. Sequence analysis of the protein (LS-12) reveals that it contains 108 amino acids, is blocked at the N-terminus by a pyroglutamyl group and has a high (17%) content of glutamine; it is thus completely unrelated to the earlier described types I, II and III fish antifreeze proteins. Circular dichroism spectra and conformational analysis based on the sequence data indicate that LS-12 has a high helix content and probably folds as a four-helix bundle. LS-12 shows sequence similarity to certain plasma apolipoproteins known to have helix bundle structures, suggesting the possibility that LS-12 may have arisen by recruitment and mutation of a plasma apolipoprotein.**

*Key words:* Antifreeze protein; Longhorn sculpin; Amino acid sequence; Helix bundle protein; Apolipoprotein

## 1. Introduction

Many marine fishes inhabiting polar and northern coastal waters avoid freezing by secreting antifreeze polypeptides (AFPs) during winter months. Three distinct types of AFP have been described: type I, alanine-rich amphipathic α-helices with molecular weights of about 3500; type II ($M_r$ ca. 14000), cystine-rich proteins related to certain lectins; and type III ($M_r$ ca. 6500), having a β-sheet sandwich structure [1,2]. The fish AFPs act by binding to the surface of ice crystals and inhibiting or preventing growth, in the process of which the ice crystals assume a characteristic hexagonal bipyramidal shape. Type I AFPs have been isolated from various species of flounder and sculpin [2]. Recently, we examined the blood serum of longhorn sculpin, *Myoxocephalus octodecimspinosis,* for which no AFP had been reported previously, expecting to find a type I AFP. A protein with antifreeze activity was indeed found. However, characterization of this protein revealed that although it is highly α-helical, it has a much higher molecular weight (12299) than the other sculpin AFPs. Sequence and secondary structure analysis, described here, suggest that this protein is the first member of a new class of antifreeze protein, which we call type IV.

*Corresponding author. Fax: (1) (617) 353-6466.
E-mail: laursen@bu.edu

*Abbreviations:* AFP, antifreeze polypeptide; MALDI-TOF, matrix assisted laser desorption time-of-flight

## 2. Materials and methods

### 2.1. Protein isolation

The isolation and physical characterization of longhorn sculpin AFP will be described in detail elsewhere. Briefly, blood plasma from fish collected off the coasts of Massachusetts and New Hampshire in late winter months was chromatographed on a Sephadex G-75 column in 0.1 M ammonium bicarbonate. Fractions which, after concentration, exhibited antifreeze activity [3] were further purified by reversed phase HPLC on a Vydac $C_{18}$ semipreparative column using 0.05% trifluoroacetic acid/water and 0.05% trifluoroacetic acid/acetonitrile gradients. Antifreeze activity was found in a late eluting peak designated LS-12, based on an apparent molecular weight of about 12000.

### 2.2. Measurement of antifreeze activity

Antifreeze activity, defined as thermal hysteresis (the difference between the freezing and equilibrium melting temperatures) was measured as described previously [3] in 0.1 ammonium bicarbonate (pH 7.9) using a nanoliter osmometer mounted on a microscope stage. Crystal morphology and growth rates were recorded using a video camera mounted on the microscope and a video cassette recorder. Crystal dimensions were measured directly on a video monitor screen.

### 2.3. Protein fragmentation

*Cyanogen bromide.* Cyanogen bromide in 70% formic acid (50 μl; 50 mg/ml) was added to about 10 nmol of protein, and the resulting solution was kept in the dark at room temperature for 18 h. The solution was diluted to 0.5 ml with water and lyophilized.

*Trypsin.* The protein was dissolved in 0.1 M $NH_4HCO_3$ (pH 7.9) and a total of 4% w/w of trypsin was added in two aliquots; the digestion was continued overnight at room temperature.

*Endopoteinase Glu-C.* About 7 nmol of protein was digested with 4% (w/w) of endoproteinase Glu-C in 115 μl of 50 mM $NH_4HCO_3$ (pH 7.9) at 37°C for 4.5 h.

*Thermolysin.* About 5 nmol of the N-terminal tryptic peptide was dissolved in 50 μl of 50–mM $NH_4HCO_3$ (pH 7.9) buffer. To this solution was added 5 μl of 0.1 M $CaCl_2$ and 5 μl of thermolysin (0.02 mg/ml in the same buffer), after which the mixture was incubated at 37°C for 4 h.

*Endoproteinase Asp-N.* About 6 nmol of the N-terminal tryptic peptide was digested with 3% (w/w) of endoproteinase Asp-N in 100 μl of 10 mM Tris buffer (pH 7.5) at 37°C for 1 h.

### 2.4. Purification of peptides

All peptides obtained from digestions were separated and purified by reversed-phase HPLC on Vydac $C_{18}$ columns (4.6 mm×250 mm; 5 μM particle side, with a pore diameter of 300 Å) using linear gradients of 0.05% trifluoroacetic acid/water and 0.05% trifluoroacetic acid/acetonitrile on a Millipore/Waters model 600E HPLC system, with detection at 214 nm.

### 2.5. Mass measurements

Molecular mass measurements of peptides and the protein were determined by matrix assisted laser desorption time-of-flight (MALDI-TOF) mass spectroscopy on Finnigan LaserMat and VG TofSpec instruments.

### 2.6. Amino acid analysis

Protein or peptide samples were hydrolyzed in 6 M HCl vapors under vacuum for 24 h at 110°C. The hydrolysates were derivatized with phenyl isothiocyanate and analyzed using the Waters PicoTag


*PII* S0014-5793(96)01466-4

protocol on a Waters NovaPak $C_{18}$ column (3.9 mm×300 mm; 5 μm particle, 60 Å pore size).

*2.7. Sequence analysis*

Peptides were sequenced by solid phase Edman degradation on a MilliGen model 6600 ProSequencer and a Beckman LF3000 protein sequencer. Peptides were immobilized according to standard Millipore protocols by covalent attachment of carboxyl groups to Millipore Sequelon-AA arylamine disks or side chain amino groups to Sequelon-DITC membrane disks, or by adsorptive immobilization of Immobilon-CD membrane disks.

Electrospray MS-MS sequencing of the N-terminal tryptic and cyanogen bromide peptides was done on a Fisons-VG Quattro instrument.

Sequencing of the pyroglutamyl-blocked N-terminus of LS-12 was accomplished by attaching the intact protein (ca. 1 nmol) to a Sequelon-DITC membrane disk, deblocking by treating the membrane-bound protein with pyroglutamate aminopeptidase (200 μl; 60 μg/ml in 50 mM phosphate, 10 mM EDTA, 5 mM dithiothreitol buffer, pH 8.0; from Boehringer) for 18 h at 4°C, and sequencing on the ProSequencer [4].

## 3. Results and discussion

The two distinguishing characteristics of all of the fish antifreeze polypeptides investigated so far are that they exhibit thermal hysteresis – depression of the freezing point of water below the equilibrium melting point of ice – and that they cause ice crystals to grow as hexagonal bipyramids, rather than the hexagonal plates seen in the absence of AFP, presumably by binding to and inhibiting growth on the hexagonal bipyramidal surfaces. In both these respects, LS-12 behaves as a typical antifreeze polypeptide: it shows concentration dependent thermal hysteresis comparable in magnitude to that of other species of AFP (Fig. 1) and it causes ice crystals to assume a hexagonal bipyramidal habit (Fig. 2), with a *c* to *a*-axis ratio of 3.4, which is nearly identical to that seen for the type I AFP from winter flounder [3]. In the case of LS-12, however, the halves of the bipyramid are slightly rotated with respect to each other, suggesting that the binding surface may be different than that for the winter flounder AFP. At the freezing point, ice crystals grown in

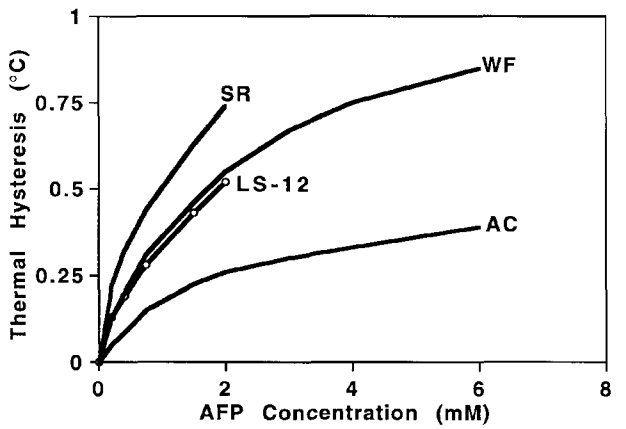


Fig. 1. Antifreeze activity of LS-12 and of other representative antifreeze polypeptides. The concentration dependent thermal hysteresis of LS-12 was measured as described in the text and compared with that of other AFPs: SR, sea raven ($M_r$ 14000); WF, winter flounder ($M_r$ 3300) and atlantic cod ($M_r$ 2600) (data for SR, WF and AC are from [1]). Measurement of LS-12 antifreeze activity could not be obtained at higher concentrations because of the tendency of LS-12 to aggregate and precipitate.

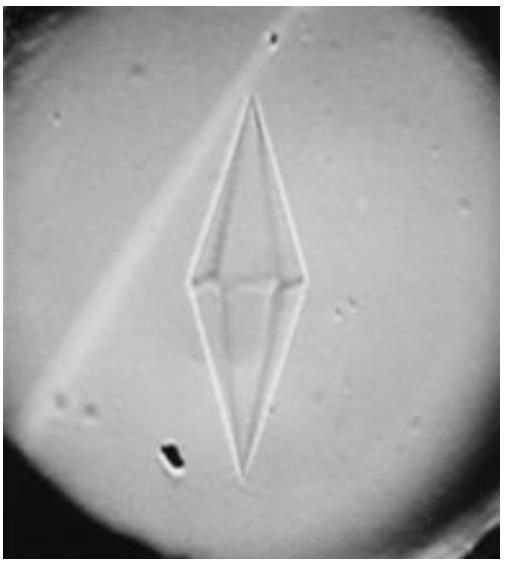

Fig. 2. Effect of LS-12 on ice crystal morphology. Hexagonal bipyramidal shape for ice in the presence of LS-12 in 0.1 M ammonium bicarbonate (pH 7.9) in water undercooled 0.1°C. At the freezing point, the crystal elongates rapidly along the *c*-axis (vertical axis) to give a needle-like shape.

the presence of LS-12 elongate rapidly along the *c*-axis to form needles, as do other AFPs.

MALDI-TOF mass spectrometry showed a molecular mass for LS-12 of 12296.5 Da, which was in good agreement with the average mass (12999 Da) calculated from the sequence. On SDS polyacrylamide gel electrophoresis (data not shown) protein treated with mercaptoethanol migrated to the same position as untreated protein, suggesting that the protein did not contain disulfide bridges. The lack of cystine/cysteine was confirmed by amino acid analysis, which also showed the presence of all other amino acids normally found in hydrolysates, including a particularly high (26%) content of Glu/Gln.

Initial attempts to sequence intact LS-12 were unsuccessful, indicating that the N-terminus was blocked. Subsequently the protein was digested with trypsin, endoproteinase Glu-C and cyanogen bromide, to give overlapping fragments which, after purification and sequencing by solid phase Edman degradation, allowed most of the sequence to be deduced (Fig. 3). For the most part, the proteinases cleaved LS-12 at sites expected based on the known specificity of the enzymes. There were some exceptions, however, the most interesting being cleavage on the amino side of Glu-14 by endopeptidase Asp-N, as well as cleavage on the carboxyl side of Asp-23, Asp-32, Asp-86 and Asp-101 by endoproteinase Glu-C.

The only real difficulty encountered was sequencing the N-terminal tryptic peptide T24. Electrospray MS-MS sequencing of T24 (and also of CB-7) allowed most of the remaining sequence to be deduced, but misinterpretation caused us to think that T24 contained an N-terminal acetyl group. Subsequent fragmentation of T24 with thermolysin and endoproteinase Asp-N, followed by sequencing, allowed for identification of residues 5–17. Eventually a small, N-terminal blocked Asp-N proteinase fragment (T24DN1) was isolated and found to have the composition $Glu_1$, $Gly_1$, $Ala_2$, suggesting the possibility of an N-terminal pyroglutamyl group.

The N-terminal region of the protein was subsequently de-

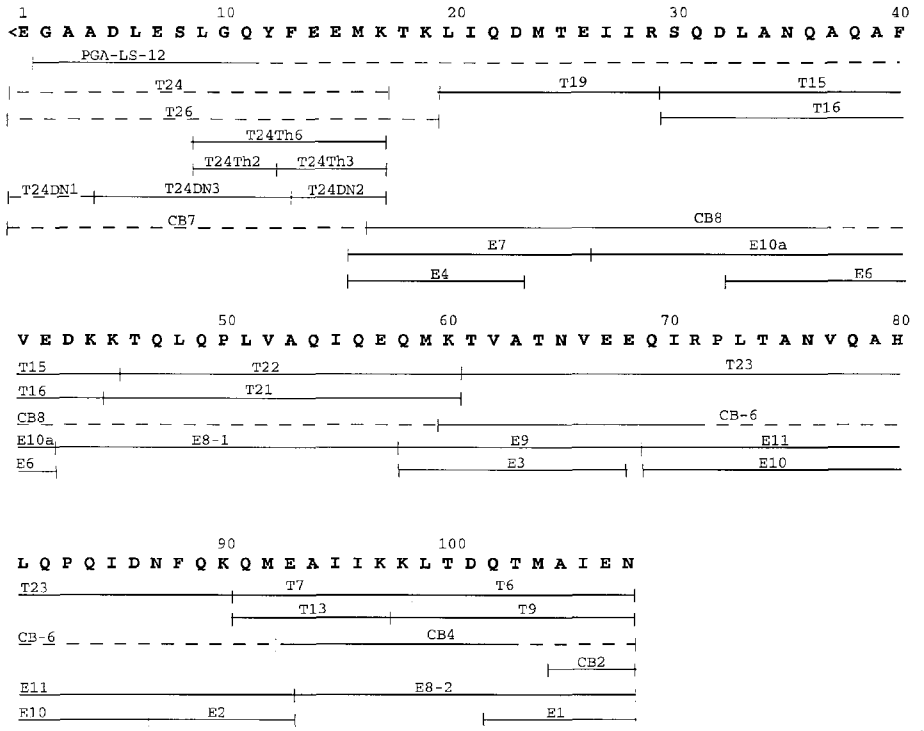


Fig. 3. Amino acid sequence of LS-12 with alignment of fragments. Cleavage was carried out using the enzymes trypsin (T), endoproteinase Glu-C (E), endoproteinase Asp-N (DN), thermolysin (Th) and pyroglutamate aminopeptidase (PGA), as well as cyanogen bromide (CB).

duced after deblocking by prolonged digestion of the intact protein with a two-fold (w/w) excess of pyroglutamate aminopeptidase (an early attempt to deblock with this enzyme under less vigorous conditions failed). To prevent contamination of the sample by the peptidase, which could interfere with interpretation of the sequencing results, the LS-12 sample was first immobilized on a Sequelon-DITC membrane disk and then treated with pyroglutamate aminopeptidase, which was then washed off. Sequencing of the deblocked protein gave residues 2–11, establishing the overlap with the thermolysin and endoproteinase Asp-N fragments (Fig. 3). Residue 1 was deduced to be pyroglutamic acid. MALDI-TOF mass spectra of the intact protein and of most of the peptide fragments were obtained (data not shown). These not only confirmed the


Fig. 4. Alignment of LS-12 with representative high scoring proteins from BLAST and BLITZ searches [4]: APA2_MACFA, apolipoprotein A-II (crab-eating macque); APA1_CHICK, apolipoprotein A-I (chicken); APE_CAVPO, apolipoprotein E (guinea pig); 1AEP, apolipophorin III (African locust); 1LPE, apolipoprotein-E3 LDL receptor binding domain (human).

composition of the peptides and of the protein, deduced by sequencing, but demonstrated the absence of posttranslational modifications. In all, about 110 nmol (1.4 mg) of LS-12 was used for sequence analysis, about 60% of which was expended on the N-terminal region.

Comparison of the LS-12 sequence with other proteins using BLITZ and BLAST database search tools [5] revealed up to about 22% sequence similarity with several apolipoproteins (Fig. 4). These proteins are characterized by containing a high proportion of amphipathic α-helix [6]. Circular dichroism spectra (to be reported elsewhere) indicate that LS-12 also is highly helical, and protein structure prediction algorithms suggest that LS-12 is comprised of four amphipathic helixes which could fold into an antiparallel bundle, with the hydrophobic faces on the interior and the polar sides facing solvent water [7]. Such a structure has been determined for the low density lipoprotein receptor binding domain of human apolipoprotein E3 (Fig. 4) [8]. It is not clear whether LS-12 is truly homologous with the apolipoproteins or whether these proteins merely have analogous functions. For example, the apolipoproteins form discoidal complexes with phospholipids, with the non-polar surfaces of the amphipathic helixes aligned with the hydrophobic phospholipid chains, and the polar surfaces facing water [6]. In the absence of phospholipid, the hydrophobic surfaces of the helixes could interact with each other forming helical bundle structures, such as seen in apolipoprotein E3 [8]. Thus in view of these structural and functional similarities it is conceivable that LS-12 could have evolved from a plasma apolipoprotein, with the outer, polar surfaces becoming specialized for binding to ice crystal surfaces.

Identification of the ice-binding surface(s) of LS-12 will require elucidation of the three-dimensional structure by crystallographic or NMR techniques and analysis of the activity of mutants, for example, as has been done with type III AFPs [9]. To this end we have initiated cloning studies with the aim of being able to produce sufficient quantities of LS-12 and mutants for structural analysis. Sequencing of cDNA fragments encoding the LS-12 gene completely confirms the sequence reported here.

*Acknowledgements:* This work was supported by a grant from the Army Research Office. We would like to thank Dr. LeRoy Martin of Fisons Instruments for MS-MS analysis of some peptides.

## References

[1] Davies, P.L. and Hew, C.L. (1990) FASEB J. 4, 2460–2468.
[2] Ananthanaryanan, V.S. (1989) Life Chem. Rep. 7, 1–32.
[3] Wen, D. and Laursen, R.A. (1992) J. Biol. Chem. 267, 14102–14108.
[4] Andrews, D.W., Girardi, M. and Mark, J. (1990) in: Techniques in Protein Chemistry II (Villafranca, J., Ed.), Academic Press, Orlando, FL, pp. 75–80.
[5] BLAST and BLITZ search forms are accessible on the World-WideWeb at <http://expasy.hcuge.ch/cgi-bin/BLAST.pl/BLASTEPFL.pl> and <http://www.ebi.ac.uk/searches/blitz_input.html>, respectively.
[6] Brasseur, R., Lins, L., Vanloo, B., Ruysschaert, J.-M. and Rosseneu, M. (1992) Proteins Struct. Funct. Genet. 13, 246–257.
[7] Branden, C. and Tooze, J. (1991) Introduction to Protein Structure, Garland Publishing, New York, pp. 33–41.
[8] Wilson, C., Wardell, M.R., Weisgraber, K.H., Mahley, R.W. and Agard, D.A. (1991) Science 252, 1817–1822.
[9] Chao, H., Sönnichsen, F.D., DeLuca, C.I., Sykes, B.D. and Davies, P.L. (1994) Protein Sci. 3, 1760–1769.